Molecular dynamics simulations of aqueous ion solutions

by

Mohomed Nawavi Mohomed Naleem

B.Sc. (Hons), University of Colombo, Sri Lanka, 2008

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Abstract

The activity and function of many macromolecules in cellular environments are coupled with the binding of ions such as alkaline earth metal ions and poly oxo anions. These ions are involved in the regulation of important processes such as protein crystallization, nucleic acid and protein stability, enzyme activity, and many others. The exact mechanism of ion specificity is still elusive. In principle, computer simulations can be used to help provide a molecular level understanding of the dynamics of hydrated ions and their interactions with the biomolecules. However, most of the force fields available today often fail to accurately reproduce the properties of ions in aqueous environments.

Here we develop a classical non polarizable force field for aqueous alkaline earth metal halides (MX_2) where M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^-, and for some biologically important oxo anions which are NO_3^-, ClO_4^-, $H_2PO_4^-$ and SO_4^{2-}, for use in biomolecular simulations. The new force field parameters are developed to reproduce the experimental Kirkwood-Buff integrals. The Kirkwood-Buff integrals can be used to quantify the affinity between molecular species in solution. This helps to capture the fine balance between the interactions of ions and water. Since this new force field can reproduce the experimental Kirkwood-Buff integrals for most concentrations of the respective salts, they are capable of reproduce the experimental activity derivatives, partial molar volumes, and excess coordination numbers. Use of these new models in MD simulations also leads to reasonable diffusion constants and dielectric decrements.

Attempts to develop force field parameters for CO_3^{2-}, HPO_4^{2-} and PO_4^{3-} ions were unsuccessful due to an excessive aggregation behavior in the simulations. Therefore, in an effort to overcome this aggregation behavior in the simulations, we have investigated scaling the anion to water interaction strength, and also the possibility of using a high frequency permittivity in the

simulations. The strategy of increasing relative permittivity of the system to mimic electronic screening effects are particularly promising for decreasing the excessive ion clustering observed in the MD simulations.

Molecular dynamics simulations of aqueous ion solutions

by

Mohomed Nawavi Mohomed Naleem

B.Sc. (Hons), University of Colombo, Sri Lanka, 2008

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2017

Approved by:

Major Professor
Dr. Paul Edward Smith

Copyright

© Mohomed Nawavi Mohomed Naleem 2017.

Abstract

The activity and function of many macromolecules in cellular environments are coupled with the binding of ions such as alkaline earth metal ions and poly oxo anions. These ions are involved in the regulation of important processes such as protein crystallization, nucleic acid and protein stability, enzyme activity, and many others. The exact mechanism of ion specificity is still elusive. In principle, computer simulations can be used to help provide a molecular level understanding of the dynamics of hydrated ions and their interactions with the biomolecules. However, most of the force fields available today often fail to accurately reproduce the properties of ions in aqueous environments.

Here we develop a classical non polarizable force field for aqueous alkaline earth metal halides (MX_2) where M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^-, and for some biologically important oxo anions which are NO_3^-, ClO_4^-, $H_2PO_4^-$ and SO_4^{2-}, for use in biomolecular simulations. The new force field parameters are developed to reproduce the experimental Kirkwood-Buff integrals. The Kirkwood-Buff integrals can be used to quantify the affinity between molecular species in solution. This helps to capture the fine balance between the interactions of ions and water. Since this new force field can reproduce the experimental Kirkwood-Buff integrals for most concentrations of the respective salts, they are capable of reproduce the experimental activity derivatives, partial molar volumes, and excess coordination numbers. Use of these new models in MD simulations also leads to reasonable diffusion constants and dielectric decrements.

Attempts to develop force field parameters for CO_3^{2-}, HPO_4^{2-} and PO_4^{3-} ions were unsuccessful due to an excessive aggregation behavior in the simulations. Therefore, in an effort to overcome this aggregation behavior in the simulations, we have investigated scaling the anion to water interaction strength, and also the possibility of using a high frequency permittivity in the

simulations. The strategy of increasing relative permittivity of the system to mimic electronic screening effects are particularly promising for decreasing the excessive ion clustering observed in the MD simulations.

Table of Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

Acknowledgements

First and foremost, I would like to express my sincere gratitude for my supervisor Prof. Paul Smith for all his support and guidance throughout my graduate studies. I feel very lucky to have the opportunity to work with such a nice and brilliant computational chemist. I want to thank him very much for being so patient with me and encouraging and inspiring me to achieve my goals. Your advices and guidance have a profound impact in my life and made me a competent to undertake and tackle challenges in my life.

I would like to express my thank to Prof. Ryszard Jankowiak, Prof. Stefan Bossmann, Prof. Jeremy Schmit and Prof. Mary Rezac for their valuable time for working as the members of my Ph. D. defense committee. I want to thank Prof. Eric Maatta and Prof. Ramaswamy Krishnamoorthi who were in my committee before they were retired from Kansas State University. I appreciate Prof. Stefan Bossmann and Prof. Jeremy Schmit who joined my Ph. D. defense committee without any hesitation to replace the retired professors.

I would like to thank Prof. Samantha Weerasinghe from university of Colombo, Sri Lanka for all his guidance throughout my undergraduate studies. He encouraged and helped me to peruse my graduate studies at Kansas State University. Also, I want to thank all past and present group members in the Smith research group (Yuanfang Jiao, Shu Dai, Elizabeth Ploetz, Sadish Karunaweera, Gayani Pallawella and Nilusha Kariyawasam) for being so helpful. You guys treated me like a own family member and made my stay at Kansas State University memorable.

Also I would to thank Dr. Nikolaos Bentenitis, Jin Zou, Prof. Christine Aikens and Amendra Fernando for their help in my research projects. I want to thank all the academic and nonacademic staff in the department of chemistry of Kansas State University for all their support given to me though out my graduate studies.

Last but not least, I would like to thank my wonderful parents, family members and my wife for being understanding and encouraging me throughout my life to be what I am today. Thank you very much for all of you for all your sacrifices to make my life successful.

Dedication

To my wonderful parents for their endless love, encouragement and support

Chapter 1 - Introduction

Computational chemistry is a branch of chemistry that gains increasing attention from the thriving scientific community day by day. As an example of this is the fact that the Nobel Prize in the chemistry for 2013 was awarded to three scientists for their contribution to the advancement of the computational chemistry. In computational chemistry, a model of a real chemical system is designed and studied using computers.[1] These models include a set of equations, and parameters associated with those equations, which can represent either quantum mechanical or classical mechanical levels of theory. Then, both measurable and unmeasurable properties are computed for a particular model. A comparison of those measurable properties with experimental properties are then used to validate or invalidate the model. Once a valid model is achieved additional properties can be computed to help solve real world chemical problems, complement experimental results, or to predict unknown or unmeasurable quantities. Although there is a tendency towards using computational chemistry in research when collaborating with experimentalist to solve chemical problems, there are limitations regarding what is currently possible in this field. Hence, steady improvements in computer hardware and software, coupled with more refined representations of the currently available chemical models, are required in order to improve many appreciations in this field.

1.1. The impact of computer hardware on computer simulations

At the atomistic level, even a very small size chemical system may consist of a very large number of atoms. Therefore, it requires significant parallel computing capabilities to simulate many chemical systems. The advancement of parallel computing has enabled the simulation of larger and larger chemical systems and/or smaller systems for longer times. The unprecedented

growth in computational capabilities of high performance computing in the last decade has led to many new disciplines using computer simulations as an important tool to compliment theory and experiment.

In 1965, one of the co-founders of the Intel cooperation, Gordon Moore, introduced "Moore's law" which predicts that the transistor density of semiconductor chips would roughly double every 18 months. But, more transistors mean a processer requires higher power densities. Therefore, it was forecasted around the year 2000 that computer microprocessors would produce enough heat to destroy the microprocessor itself. In response to this problem, the computer industry started to shift towards multi core designs adapting parallel computing instead of the conventional serial approach.[2] This strategy enabled an increase in the transistor density in microprocessors by increasing the number of cores without increasing power consumption. Nowadays, almost all computers are equipped with multi core processors.

1.2. Computer simulations in chemistry

In general, computer simulations of chemical systems fall in to two major divisions. Quantum mechanical (QM) simulations and classical mechanical simulations. From an accuracy point of view QM is usually the best approach to analyze a given chemical system. However, QM cannot be used to simulate a system with even just a few thousand atoms due to limitations in theory and applications.[3-4] Therefore, classical mechanics is the most widely used approach to simulate larger chemical systems.

In classical computer simulations the time trajectories for a system of interacting atoms, molecules or ions are generated via numerical integration of the classical equations of the motion. Ideally, a full description of the system includes the dynamics of both the electrons and nuclei of

the atoms. But, in the classical simulations one invokes the Born-Oppenheimer (BO) approximation,[5] due to the large separation between the times scales for nuclei and electron motions. Therefore, the dynamical evolution of the nuclei can be examined without considering any explicit contributions from the electrons. Since electrons are not treated explicitly in classical simulations an inherent disadvantage is that these simulations cannot model QM effects such as proton tunneling, and bond making or breaking. Here atoms and bonds are treated as simple "ball and spring" models.[6] Another major limitation of classical simulations are the simulation time for a given system size (Figure 1.1). Fast chemical bond vibrations limit the time step used in numerical integration to the femtosecond time scale.[7] This means millions of steps are required to reach the nanosecond time scale, and billions are required to reach the microsecond time scale. Hence, it can be very difficult to study important biologically relevant processes, which can involve processes occurring many orders of magnitude slower than this, using currently available computers.

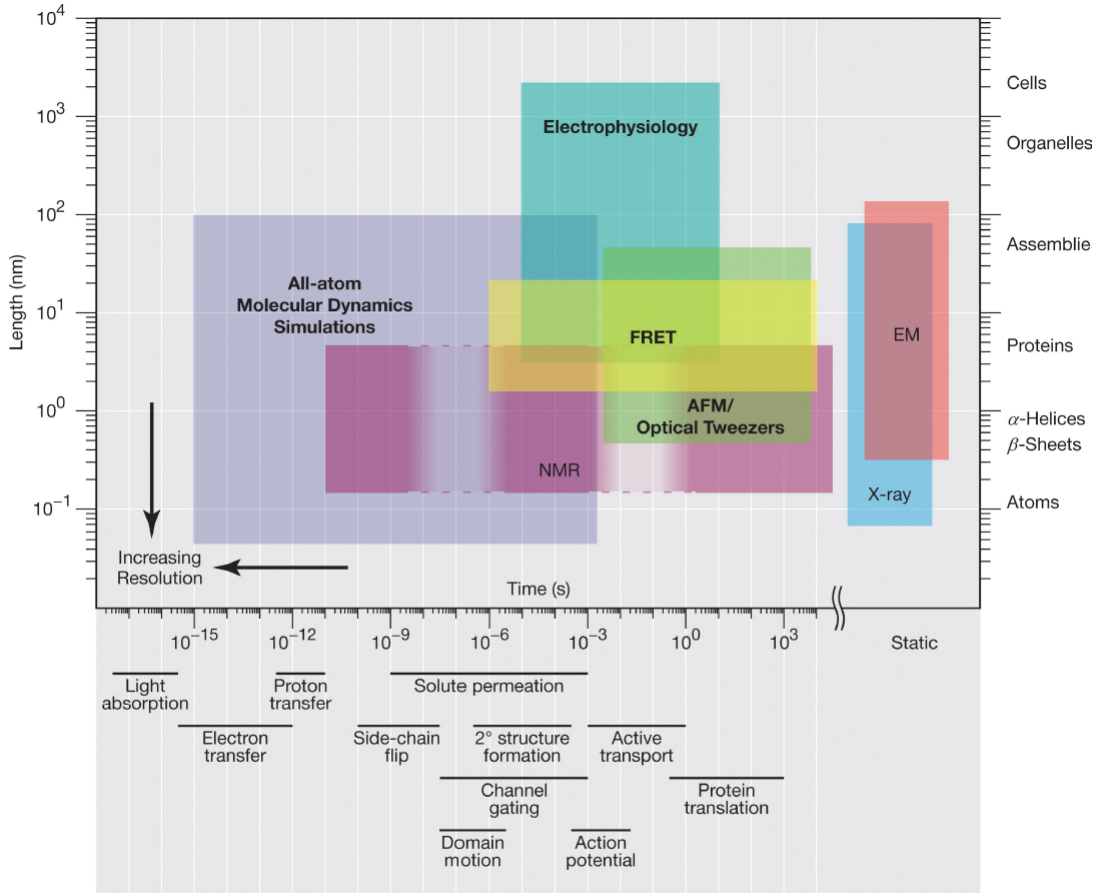

Figure 1.1. Spatial and temporal resolution of various experimental and simulation techniques (taken from Dror et al[7])

Classical simulations can also be subdivided in to several categories. These are Monte Carlo (MC) simulations, Langevin dynamics (LD) simulations, Brownian dynamics (BD) simulations, and molecular dynamics (MD) simulations.[8] In MC one uses randomly generated probabilities to decide if the movement of a molecule is allowed. The earliest applications of MC used highly idealized representations of molecules such as hard spheres and disks.[9] LD and BD simulations also use simplified description of the chemical system where the solvent molecules are not treated explicitly. Hence, many important equilibrium and dynamic properties, which are affected by the solvent, cannot be studied in these systems. In most MD simulations of biological

systems the solvent is treated explicitly. MD simulations can be computationally costly for electrolyte solutions due to the multiplicity of the components, the presence of long range interactions, and the need to perform longer simulations to obtain converged ion -water properties.[10] Finally, there are hybrid variants of MD simulations where a small portion of the simulation box is treated using a QM level of the theory. However, these simulations are much slower than classical MD simulations, and they are only used when necessary.

1.2.1. Molecular dynamics

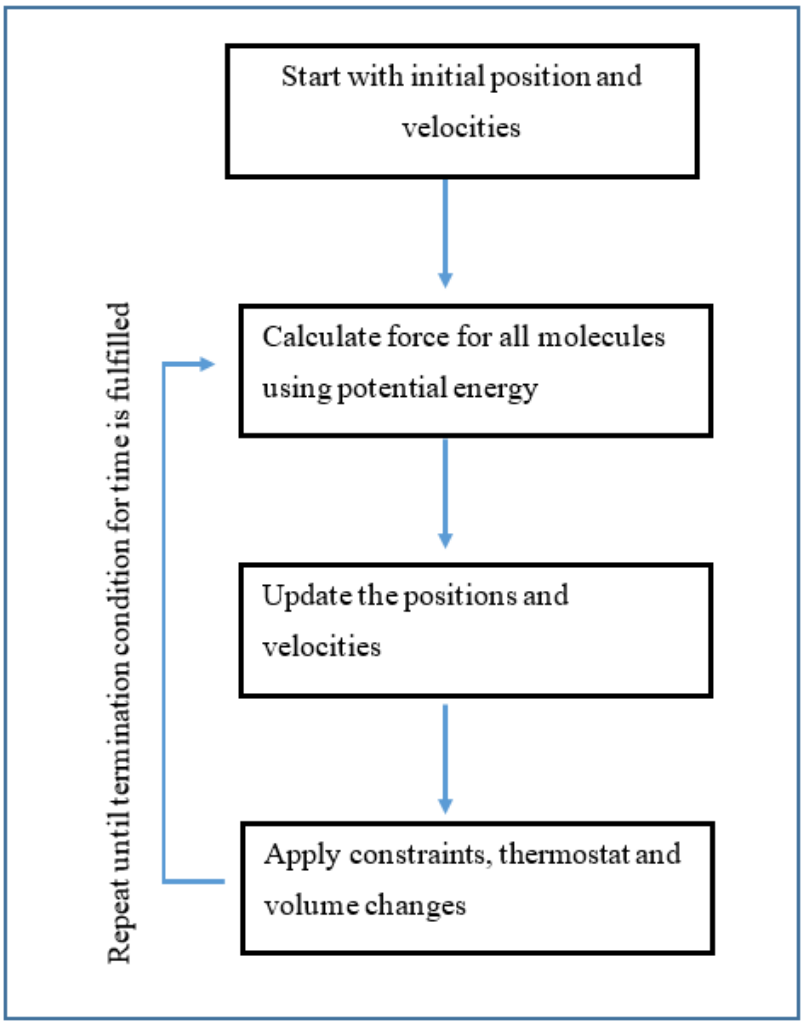

Figure 1.2. A simple molecular dynamics scheme

Since MD simulations use classical mechanics to calculate the total energy of the system the Hamiltonian for the system can be written as shown in Eq. 1.1. Here, H is the total energy (Hamiltonian) of the system, p is the momentum, r is the position, m is the mass, and U is the potential energy.

$$H(\mathrm{p},r) = \sum_i \frac{p_i^2}{2m_i} + U \qquad \textbf{Eq. 1.1}$$

Derivatives of the Hamiltonian then provide the equations of motion. These involve the forces, $-\nabla_r U$, on the atoms as displayed in the Eq. 1.2 and Eq. 1.3. The force field, which is used to calculate the potential energy of system, is discuss later in this thesis. The classical equations of motion are then,

$$\dot{r}_i = \frac{\partial H}{\partial p_i} = \frac{p_i}{m_i} \qquad \textbf{Eq. 1.2}$$

$$\dot{p}_i = -\frac{\partial H}{\partial r_i} = -\nabla_r U \qquad \textbf{Eq. 1.3}$$

As mentioned previously, the time step depends on the fastest degrees of freedom in the system. Usually, in classical MD simulations the bond vibrations are the fastest motions. Typically, the values range between 10-20 femtoseconds. Therefore, most of the MD simulations use a value close to 0.5 femtoseconds for a time step. However, for MD simulations systems with no bond vibrations the time step can be increased to 5 femtoseconds.

Alder and Wainwright in the late 1950's performed the first molecular dynamic simulations to study the interactions of hard spheres and to obtain insights about the behavior of simple liquid systems.[11-12] In 1964, Rahman performed the first molecular dynamics simulation with a realistic potential to study liquid argon system.[13] The next major advance was in 1974 and involved the

simulation of the complex liquid water.[14] Computer simulations of biological systems have evolved greatly since the first 10 picosecond molecular dynamics simulation of the BPTI protein in a vacuum by McCammon, Gelin and Karplus in 1977.[15] Indeed, a recent protein folding study of BPTI reached the millisecond time scale in 2009 with the help of a computer specially designed for simulation named, Anton.[16-17]

1.3. What is a force field?

There are several challenges for performing a successful biomolecular simulation.[18] The force field must realistically represent the atomic and molecular interactions, the conformational space must be sampled in a manner that is both fast and efficient and appropriate experimental data must be available against which the simulations can be validated. The quality of the force field is perhaps the most important factor in these considerations as it is ultimately responsible for the accuracy of the MD results.[18]

The force field is the set of equations and associated parameters used to describe the interactions between all the atoms in a system. The force field has to be sufficiently accurate to reproduce the properties and mechanisms underlying the processes of interest.[17] Force fields are usually parametrized by using experimental data obtained from crystallographic, spectroscopic, and thermodynamic studies (densities, free energies), together with some properties obtained via fitting to QM data for the molecule of interest.[18-19]

According to classical mechanics, the kinetic energy and the potential energy lead to the total energy of a given chemical system. To calculate the kinetic energy, we only need to know the masses (m) and the velocities (v) of the respective atoms according to,

$$E_{Kinetic} = \sum_i \frac{1}{2} m v^2 \qquad \textbf{Eq. 1.4}$$

The potential energy is usually expressed in terms of a set of bonded and non-bonded interactions. The bonded interactions include contributions to potential energy from any bonds, angles and dihedrals within the molecules as displayed in Eq. 1.5.

$$\begin{aligned} U = {} & E_{Bonds} + E_{Angles} + E_{Proper\ Dihedrals} + E_{Improper\ Dihedrals} + \\ & E_{Electrostatics} + E_{Van\ der\ Waals} \end{aligned}$$

$$E_{Bonds} = \sum_{Bonds} \frac{1}{2} k_b \left(r - r_0\right)^2$$

$$E_{Angles} = \sum_{Angles} \frac{1}{2} k_\theta \left(\theta - \theta_0\right)^2 \qquad \textbf{Eq. 1.5}$$

$$E_{Proper\ Dihedrals} = \sum_{Torsions} k_\phi \left[1 + \cos\left(n\phi - \delta\right)\right]$$

$$E_{Improper\ Dihedrals} = \sum_{\text{Improper}} \frac{1}{2} k_\xi \left(\xi - \xi_0\right)^2$$

Here, k_b is the force constant of the bond and r_0 is the equilibrium bond distance in the bonds term displayed in Eq. 1.5. Similarly, k_θ is the angle constant and θ_0 is the equilibrium angle for the angle term in Eq. 1.5. For the proper dihedral term displayed in Eq. 1.5 the k_ϕ, n, δ parameters are the force constant, multiplicity, and phase shift, respectively. The multiplicity determines how many maxima and minima there are in the 360 degrees of rotation, and the phase shift is use to shift the maxima and minima to appear at the correct values of the dihedral angle. The improper dihedrals are used to model conjugation and planarity of the molecule, where k_ξ is the improper dihedral constant and ξ_0 is the equilibrium improper dihedral angle. The non-bonded contributions to the potential energy includes both van der Waals and electrostatics according to,

$$E_{Electrostatics} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \textbf{Eq. 1.6}$$

$$E_{Van\ der\ Waals} = \sum_{i,j} 4\varepsilon_{ij} \left[\left[\frac{\sigma_{ij}}{r_{ij}} \right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}} \right]^{6} \right]$$

Here, q is the effective atomic charge and ε_0 is the permittivity of the medium. The van der Waals term in the Eq. 1.6 includes the parameter σ_{ij}, corresponding to the LJ diameter which is a measure of the size or contact distance between atoms, and the parameter ε_{ij} which determines the LJ interaction strength between i and j atoms. There are x-ray crystallographic experimental data available to determine the values of r_0, θ_0, and ξ_0, and spectroscopic data that can be used to determine the values of k_b, k_θ, and k_ξ. Furthermore, equation of state data can be used to determine ε_{ij} and σ_{ij}, or they can be obtained by using QM calculations on small molecules. Unfortunately, there are no experimental data available to determine parameters for the dihedral potentials (k_ϕ, n, δ), or for the partial atomic charges (q). Hence, these parameters are usually obtained from QM calculations. Unfortunately, the charges on atoms are mostly based on gas phase QM calculations, which do not include the polarization and screening effects present in the solvent medium. Therefore, most of the commonly used force field parameters differ in terms of the parameters that are not available experimentally, such as the partial atomic charges and dihedral parameters.

AMBER[20], CHARMM[21],GROMOS[22],OPLS[23], KBFF[24] are some of the most common non polarizable force fields used in biomolecular simulations nowadays. Non polarizable force fields have been very successful in modeling many complex biological systems despite their simplicity.[25]

1.4. Polarization

Incorporating polarization and electronic screening effects in to computer simulations is a challenging task. The permittivity ε is used to describe how a dielectric medium interacts with an external electric field. Even a low intensity electric field in a molecular liquid can establish an effective charge separation by aligning the dipole of the molecules in the liquid. This phenomenon is known as polarization ($\vec{P}$). The polarization is defined as macroscopic dipole moment ($\vec{M}$) per unit volume (V).[26]

$$\vec{P} = \vec{M}/V \qquad \textbf{Eq. 1.7}$$

The polarization can be divided into two microscopic contributions. Those two are the orientational polarization ($\vec{P}_\mu$) and the induced polarization ($\vec{P}_\alpha$).[26]

$$\vec{P} = \vec{P}_\mu + \vec{P}_\alpha \qquad \textbf{Eq. 1.8}$$

The orientational polarization ($\vec{P}_\mu$) originates from the alignment of the permanent molecular dipoles along the applied external field. The induced polarization ($\vec{P}_\alpha$) accounts for the generation of an additional molecular electric dipole moment due to the displacement of electron cloud by the local field. This molecular polarization is a combination of atomic polarization and electronic polarization.

All polarizable models attempt to incorporate the effect of the deformation of the electron cloud due to any surrounding electric field.[27-28] This has been achieved for many liquids by adding extra explicitly polarizable sites, or allowing atomic charges to fluctuate.[29-30] However, these models are computationally expensive. On the other hand, simpler non-polarizable models fix the molecular charge distribution and "effective charges", which implicitly include polarization effects, are used instead.

1.5. Modeling water in MD simulations

Water is the most abundant solvent and it plays very important role in the function of biological systems. Although the macroscopic properties of liquid water are well known, the microscopic forces that define the properties of water are not so well understood.[31] A water molecule can act as both hydrogen bond donor as well as a hydrogen bond accepter. A network of three dimensional hydrogen bonding interactions between water and other surrounding molecules gives many unique characteristics to water, in contrast to other solvents. These special and unusual properties have led to continued efforts to develop accurate models of water for use in computer simulations.

The development of water models, which can mimic the properties of real water, is an active area of research and has been for decades.[32-33] However, a universal water model capable of reproducing the correct experimental properties of the different water phases remains elusive. Inevitably, classical molecular dynamics simulations, or Monte Carlo simulations, are required to understand water at the molecular level because even small systems contain larger number of water molecules, prohibiting the use of QM calculations. However, the accuracy of the computer simulations depends on the quality of the interaction potentials used to describe the inter and intra molecular interactions. There are many water models that have been developed since the publication of first water model for molecular dynamics in 1971.[34] These water models can vary from simple rigid three site models to more complex and polarizable models with many sites and with flexible bonds. The QM calculations interpret the hydrogen bond in a water dimer as a result of the competition between an attractive interaction potential energy, which can be approximated to classical electrostatic interaction, and a repulsive electronic kinetic energy.[35] Different water models use different empirical potential energy functions to describe the interactions in water.

These empirical potentials can have different levels of sophistication and may include electronic polarization, quantum effects, etc. However, complex potential energy functions often lead to slower, more memory intensive, computer simulations. Hence, a balance between accuracy and computational efficiency is required.

Empirical non polarizable models assume the molecular polarizability is unaffected by the environment. However, some non-polarizable water models attempt to include polarization effects by modeling the charge distribution by multipoles.[36] The monopole, dipole, quadrupole, octupole and hexadecapole moments are the zeroth, first, second, third and fourth moments of the charge distribution. The significance of the multipole moments are inversely related to their order. The Bratko-Blum-Luzar (BBL) water model includes point dipoles and a square well tetrahedral octupole potential to model hydrogen bonding.[37] The SSDQO potential model describes a water molecule as a Lennard-Jones sphere with point dipole, quadrupole, and octupole moments.[38] These water models include multipole moments to help capture effects that are hard to model in simple non polarizable water models. As an example, quadrupole moments are claimed to be more sensitive than dipoles and influence the orientation of molecules leading to a dense packing of the system.[39] It is also claimed that including octupoles is necessary to distinguish anion and cation solvation dynamics.[40] Although including higher order multipole moments can increase the accuracy of the water model, this can increase the computational cost as well. Consequently, most of the water models neglect quadrupole or higher moments as many of the liquid properties, such as density or energy, used to fit the potential are not sensitive to the higher moments.

Most water models simply increase the dipole moment in an effort to model the effective contributions by all multipole interactions. The average effect of electronic polarization and screening of a water molecule by its neighboring water molecules is accounted for in most non

polarizable force fields by simply increasing the dipole moment from gas phase value.[41-42] Here, fixed charged water models with a permanent dipole of 2.2 - 2.4 Debye are used. However, this dipole moment is still lower than experimental liquid water value which is closer to 3.0 Debye.[41, 43].

1.5.1. Incorporation of electronic screening in MD simulations

Including electronic screening effects in non-polarizable force fields is crucial. A common way of doing this in simple force fields is by adjusting the partial charges of a molecule to represent screening and polarization effect of the solvent. This strategy is straight forward for neutral molecules such as water. But, unfortunately, not so simple for ions which are restricted to possess fixed integer charges. As mentioned before, most common force fields use fixed gas phase integer charges for ions irrespective of any effects of the solvent. This makes the ion charges and standard water models incompatible and modeling screening effects of water as a solvent in MD simulations is a challenging task.

The gas phase experimental electrostatic dipole moment of water is about 1.85 Debye.[43] For the water dimer the electrostatic dipole moment is about 2.1 Debye and this value increases in larger water clusters and is closer to 3.0 Debye in the liquid phase.[41, 43] But most of the water models used in MD simulations use effective charges that provide a dipole moment (μ^{eff}) value around 2.3 Debye.[41] This value for the dipole moment of water in non polarizable models arises after fitting to a variety of different chemical properties. It is somewhat surprising that most of the common water models display an effective dipole moment with a value close to 2.3 Debye. Leontyev et al. justify this value as an effective dipole moment that corresponds to the ratio between of real dipole moment of a water molecule in the liquid and square root of the electronic

dielectric constant of water (ε_{el}) which is 1.78. [44] This is called the MDEC (Molecular dynamics in electronic continuum) model.[44] In this model, one considers that the point charges are moving in a homogenous electronic continuum with a known dielectric constant of ε_{el} corresponding to the screening proved by the motion of the electrons in the system. If this is ignored, then the charges in the system need to be scaled down by a factor of $1/\sqrt{\varepsilon_{el}}$ (approx. 0.7 in an aqueous solution), leading to the relatively low dipole moments of many water models. Alternatively, the background dielectric constant in the MD simulation can be increased to ε_{el}, and the partial atomic charges increased by $\sqrt{\varepsilon_{el}}$, to generate a model with the "correct" liquid state dipole moment but the same properties as the original model.

As discussed before for ionized groups, integer charges are used in most of the non-polarizable force fields. By ignoring electronic screening effects, this leads to exaggerated electrostatic interactions by almost a factor of two. This is the main driving force behind the MDEC model. Unfortunately, using this approach for ions leads to non integer charges. For example, when modeling divalent metal ion using the MDEC model one would use a cation charge of +1.4 instead of +2 charge at the atom center. This is not totally satisfactory.

1.5.2. Selecting a water model for a MD simulation

TIP3P[45], SPC[46] and SPC/E[47] are a few of the popular water models used in biological simulations. These are simple water models with three interacting sites. TIP4P[48] involves four sites, and the TIP5P[49] water model includes five sites. Remarkably, despite its simplistic nature these non-polarizable rigid water models, which do not include explicit electronic polarization or screening effects, have been capable of reproducing many properties of liquid water. As an example, hydration free energies can be computed with reasonable accuracy using non polarizable

water models.[50-51] Moreover, the addition of flexibility or adding electronic polarization effects to the water model have not shown major improvements in the liquid water properties compared to the existing simple rigid, non-polarizable, water molecules.[32]

The best water model is a matter of debate. The SPC/E water model appears to reproduce the experimental structural and dynamical properties of pure water better than TIP3P. However, TIP3P can reproduce better experimental dielectric properties than SPC/E.[52] These three site water models are widely used in biological simulations due to their computational efficiency compared to the more recent four and five site water models. A review published in 2011 to evaluate seventeen chemical properties of different rigid non polarizable water models concluded that SPC/E is a reasonable water model.[53]

There are simulations studies that have shown the necessity to use compatible water models with a given force field in biomolecular simulations.[54] While which water model is the best fit for a given force field in computing a given chemical property is still a bone of contention in the research community, the widely accepted opinion is as GROMACS[55] manual suggests; that the TIP3P water model is suitable for the AMBER and CHARMM force fields, the SPC water model is compatible with GROMOS, and the TIP4P water model is compatible with OPLS force field.[54] Most of the time a force field is most compatible with the water model that was used in the parameterization process.

1.6. Modeling ions in MD simulations

Aqueous electrolyte solutions are very important in biological, geological and industrial systems.[56] In biological systems ions stabilize proteins and nucleic acids.[57-58] Ions dissolved in sea water vapor can participate in the process of corrosion, and ions in atmospheric aerosols are

involved in many environmental processes.[59] Most inorganic salt crystals dissolve easily in water. This is because ion to water interactions are more favorable than the cation and anion interactions in the salt crystals. However, capturing the fine balance between the coulomb attraction of ions and the solvation process using non-polarizable force fields is a challenging task.

The strong electrostatic interactions between ions and solvent molecules clearly contribute to the thermodynamic properties of these solutions. The hydration properties of water can be either hydrophilic or hydrophobic in nature. Polar molecules and ions participate in hydrophilic chemical interactions with water molecules. The strength relative to the thermal energy (k_BT), and long range nature of the hydrophilic interactions, makes modeling these interactions in the simulations a tedious task.

Undoubtedly, water is an important solvent. It is also a highly polarizable solvent, which makes modeling water difficult. Despite the importance, there is still no consistent molecular picture concerning the impact of ions on water dynamics.[60] One of the major unresolved questions is whether a particular ion accelerates or slows down water dynamics. Most of the recent time resolved experiments suggest water dynamics in salt solutions are slower than in pure water.[60-61] But NMR, dielectric and viscosity experiments suggest that, for some salts, the reverse is true.[60] This effect is concentration dependent with all ions slowing down water dynamics at high salt concentrations, but some ions accelerating water dynamics under dilute conditions.[62]

Another essential unresolved question is the range of an ion's influence on the properties of water molecules. Many theoretical and experimental studies indicate a ion with low charge density typically only interacts with the first solvation shell water molecules.[62-63] But ions with high charge densities can influence waters in the second or third solvation shells.[62-63] Furthermore, there are studies of high charge density ions, such as Mg^{2+}, where the counter ion can influence

the surrounding water molecules even at dilute concentrations due to some long range cooperative effects.[64]

In MD simulations the water molecules tend to point the OH bond toward the anion while its dipole points towards the cation. Therefore, it is argued that in simple water model simulations the anion would affect the water OH bond dynamics and the cation would act on the water dipole. The experimental bulk water reorientation is considered to be only slightly anisotropic.[65] However, water molecules next to a ion with high charge density, such as Mg^{2+} or SO_4^{2-} , could display a significant anisotropy in their water reorientation dynamics, as suggested by some experiments and MD simulations.[60] These are difficult features to include in simple non-polarizable force fields.

Other than the strong influence ions can have on structure and dynamics of nearby water molecules, there are ion specific effects as well. A good example indicating how different ions can display distinctive effects is the Hofmeister ion series.[66] This series ranks ions according to their ability to precipitate proteins from aqueous solutions.

There are several advantages of using MD simulations over experimental studies. The MD simulations can be used to simulate physiologically relevant low ion concentration. Furthermore, in a computer simulation, ion and water interaction can be studied without the impact of the counter ions. This is not feasible in experiments and the counter ion can influence the structure and dynamics of the solvent. Also, MD simulations can be used to study the effect of competitive ion interactions in the presence of multiple ions to investigate the interactions with biomolecules. This is very difficult to study under experimental conditions due to other interferences inside a living cell.

All the above advantages are futile if the ions modeled in the MD simulations do not mimic real aqueous ionic solution. One of the major challenges in force field development is to decide on

the partial charges on atoms for a molecule. These partial charges cannot be validated by experimental studies and do not correspond to any observable property of the system. Partial atomic charges can be used as adjustable parameters in the force field, and most of the time QM calculations are used to elucidate the values. Even modeling electrostatic interactions for ions is a challenging task for non-polarizable force fields. Most force fields model metal ions using a Lennard-Jones sphere with a constant integer charge. However, this disregards the effects of electronic dielectric screening by the solvent and other ions. This is thought to provide a reasonable representation of solute and solvent interactions, and this is probably sufficient for monovalent ions, but for the divalent and multivalent ions this simple description does not appear to be so accurate. As a example, for divalent ions the non additive effects for the first solvation shell are about 70 kcal mol^{-1}, most of which is neglected by the force field. This is approximately 14 times larger than the gas phase binding interaction energy of a water dimer.[67] A strategy to overcome this issue in MD simulations is still not clear.

The anisotropy of the water molecules in the vicinity of a solute influence the solute hydration thermodynamics in aqueous solution. In a MD simulation of water and ions it has been shown that changing the diameter of monovalent ion from 0.4 nm to 1.7 nm can result in a distinct asymmetry in the structure and thermodynamic properties of the hydration of ions.[68] This is consistent with previous theoretical and experimental studies. This study use monovalent cations, anions and some large hypothetical ions. The same study has also shown that the free energy of hydration is more favorable for negatively charged ions than for positively charged ions of the same size.[68] Therefore, the differences in size and charge of ions can impact solvation properties.

When a salt crystal dissolves in water at finite concentrations ion paring still exists. Stronger ion paring in solutions can be achieved by reducing the dielectric constant of water by

using super critical water,[69] or by changing to polar solvents,[70] or by moving from monovalent ions to multivalent salt ions.[71] Also, there are experimental and simulation studies that suggest ion pairing is a possible clue for discriminating ions with different sizes, but which have the same charge. The sodium and potassium ions have the same charge but only differ slightly in size. However, biological systems can often distinguish between the two ions with very high accuracy. There are MD studies that suggest the observed free energy change upon replacing a potassium with a sodium in a contact ion pair can be an important factor to help discriminate between these ions in a biological system.[72-73] This concept of the law of matching water affinities, which relates hydration strengths of ions to the inclination to form contact ion pairs, has been supported by some experimental studies.[74]

To model metal ions in classical force fields one only needs to define three parameters, since a metal ion only exhibits non bonded energy interactions. The first one is the point charge (q) required for electrostatic interactions. Then, for the dispersion interactions, most of the force fields use a simple LJ potential where the parameters are the interaction length (σ) and the interaction strength (ε). However, if it is an ion with multiple atoms then one need to define bonds, angles, dihedrals and the partial charge distribution for each atom too. To calculate cross interaction parameters in a force field it is possible to use the Lorentz-Berthelot combination rule.

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii} \times \varepsilon_{jj}} \qquad \textbf{Eq. 1.9}$$

$$\sigma_{ij} = 0.5\left(\sigma_{ii} + \sigma_{jj}\right)$$

To parameterize the LJ interaction length (σ) and interaction strength (ε) one typically use experimentally accessible parameters. Occasionally, using experimental data, such as solvation free energies, one can develop the LJ interaction length (σ) and interaction strength (ε). However, these infinite dilution properties can cause force fields to fail in attempts to reproduce

thermodynamic or realistic structural properties at finite concentrations. Indeed, it is known that most of the popular force fields fail to reproduce thermodynamic properties such as activity coefficients, osmotic coefficients, or compressibilities of salt solutions at finite concentrations.[75-79] Therefore, ion force field development remains an active field of research.

1.7. Kirkwood-Buff Theory

The Kirkwood-Buff (KB) theory of solutions was published in 1951 by John G. Kirkwood and Frank P. Buff to interpret solution mixtures.[80] In this paper, the solution thermodynamic properties - such as derivatives of the chemical potentials, osmotic pressure, partial molar volumes and compressibility - are related to molecular distribution functions for a multicomponent system in a grand canonical ensemble where the chemical potential (μ), volume (V) and temperature (T) are constant. At the heart of KB theory are the KB integrals (KBI) defined by,

$$G_{ij} = \int_0^\infty \left[g_{ij}(r) - 1 \right] 4\pi r^2 \, dr \qquad \textbf{Eq. 1.10}$$

where $g_{ij}(r)$ is the pair correlation function in a system between species *i* and *j*, and *r* is the intermolecular separation. The pair correlation function is sometimes called the radial distribution function (rdf). The KBI provides a measure of the affinity between all components in a solution. A positive value for the KBI implies a stronger net attraction between species *i* and *j* in a solution mixture, and vice versa.

KB theory has been reviewed extensively and can be widely applied compared to other theories such as McMillan-Mayer (MM) theory, which is typically used to understand dilute solution mixtures.[81] MM and KB theory are the same at infinite dilution. At higher solute concentrations, however, MM theory becomes difficult to use in practice, while KB theory remains

equally valid. This is because MM requires higher integrals over n-body correlation functions to be successful at high solute concentrations, whereas KB theory still only requires the pair distribution function, albeit after averaging over all other molecules in the system.

At the time KB theory was introduced it was impossible to determine the required pair distribution function for a solution. Even today, most of the time we use computer simulations to obtain the pair distribution function. This limitation led to very few potential applications of KB theory for almost 20 years.[82] A breakthrough came in 1978 when Ben-Naim introduced the KB inversion technique, which relates the experimental microscopic chemical properties to the experimental KBI values such that the KBIs could be obtained from experimental data.[83] A publication in 2008 by Smith provides more detailed information concerning the KB inversion procedure and the extension of the theory to three or four component systems.[84] The ability to compare simulated and experimental KBIs led to the use of KBIs to develop improved all atom parameters for a urea and water force field.[75] This publication was a very important milestone in force field development. More importantly, it resolved the unphysical urea self-aggregation displayed by existing models at higher concentrations. This was a major problem with many other urea force fields including OPLS. Consequently, the KB theory parameters for urea have since been used to study the reversible folding/unfolding equilibrium of the Trp-cage mini protein over a broad range of urea concentrations reproducing the experimental linear dependence of the folding free energy.[85] In this study, the correct balance of van der Waals and electrostatic interactions provided by computational parameters was used to explain the experimental agreement. Some other important applications of the KB theory to biological systems have been discussed by Pierce et al.[86] The interested reader can find more information about applications of KB theory to improve current force fields in chapter 5 of a monograph published by Smith et al.[87]

Some advantages of KB theory include that it is an exact theory. This means, other than standard statistical mechanical assumptions, no approximations or simplifications are required to describe solution behavior. Also, the theory can be applied to solutions of any number of components, any type of components, at any (stable) composition. Another advantage of KB theory is that one does not have to assume a pair wise additive potential. However, the most important practical aspect of KB theory is that it can be easily connected to computer simulations. In particular, the ability to reproduce the experimental KB integrals in a simulation can be used to benchmark whether a simulation captured the real solute and solvent interactions in a system.

1.7.1. Extracting KBIs from experimental data

The KB integrals can be used to validate MD simulations. The experimental KBIs can be extracted by using the experimental compressibility, partial molar volumes (density), and composition dependent chemical potential (activity) data. A MD simulation can be analyzed to provide the pair distribution function and therefore the corresponding simulated KBI values. Comparison of the experimental and simulated KBI values can then be used to benchmark whether the simulation captured the correct balance of solute and solvent interactions.

1.7.1.1. Relating KBI values and chemical properties

KB theory is derived for a grand canonical ensemble where the chemical potential, volume and temperature are constant. However, there are technical difficulties for the computer simulation of open system, where particles are added or deleted in order to maintain a constant chemical

potential. Therefore, most computer simulations that use KB theory are performed in closed systems where the total number of particles, temperature and pressure are maintained constant. For a system in a Gibbs ensemble, where number of particles (N), pressure (P) and temperature (T) are constant, the equation for the KBIs can be modified to give,[80]

$$G_{ij} = \int_0^R \left[g_{ij}(r) - 1 \right] 4\pi r^2 \, dr \qquad \textbf{Eq. 1.11}$$

Here, $g_{ij}(r)$ is the pair correlation function in a system with two different species *i*, *j* and *r* is the intermolecular separation. The R is a cutoff distance from the particle of interest.

$$n_{ij}(R) = \rho_j \int_0^R g_{ij}(r)\, 4\pi r^2 dr \qquad \textbf{Eq. 1.12}$$

The coordination number (n_{ij}) can be calculated by integrating the respective rdf distributions using the equation displayed in Eq. 1.12. Here, ρ_j is the number density of j particles. The coordination number quantifies the number of molecules associated with a given central molecule. Similarly, the excess coordination number ($N_{ij} = \rho_j G_{ij}$) can be used to interpret the coordination number over the whole range of distances in a solution, compared to that observed in bulk solution. Furthermore, a positive value for N_{ij} implies an excess of species j in the vicinity of species I, over the random distribution, and a negative value indicates a depletion of species j around species i.

Typically, three types of KB integrals are found in the aqueous salt solutions. These are G_{cc} for ion-ion interactions, G_{cw} for ion-water interactions and G_{ww} for the water interactions. The cations and anions are considered indistinguishable and therefore treated as a general co-solute (c) in the KB analysis.[76, 88] Furthermore, the excess coordination numbers (N_{ij}) for a binary solution,

and the experimental compressibility, partial molar volume (density), and composition dependent chemical potential derivatives are related by the equations displayed in Eq. 1.13.

$$1+N_{cc}=\rho_c RT\kappa_T+\rho_w^2\frac{\bar{V}_w^2}{\mu_{cc}}$$

$$1+N_{ww}=\rho_w RT\kappa_T+\rho_w\rho_c\frac{\bar{V}_c^2}{\mu_{cc}} \qquad \textbf{Eq. 1.13}$$

$$N_{wc}=\rho_c RT\kappa_T+\rho_w\rho_c\frac{\bar{V}_w\bar{V}_c}{\mu_{cc}}$$

where R is the gas constant, T is the temperature, κ_T is the isothermal compressibility, $\bar{V}_i$ is the partial molar volumes, and $\mu_{cc}$is the chemical potential derivatives given by,

$$\mu_{cc}=\beta\left(\frac{\partial\rho_c}{\partial\ln \mathrm{m}_c}\right)_{T,P}=1+\left(\frac{\partial\ln\gamma_c}{\partial\ln \mathrm{m}_c}\right)_{T,P} \qquad \textbf{Eq. 1.14}$$

where γ_c is the molal activity coefficient and $\mathrm{m_c}$ is the molality of the salt. The solution compressibility has a negligible effect on the calculated KBI values at moderate temperature and pressure.[89] The compressibility can therefore be assumed to follow,

$$\kappa_T=\varphi_w\kappa_{Tw}^0+\varphi_c\kappa_{Tc}^0 \qquad \textbf{Eq. 1.15}$$

where κ_{Tw}^0 is the compressibility of pure water, κ_{Tc}^0 is the compressibility of the pure salt, φ_w is the volume fraction of the water, and φ_s is the volume fraction of the salt. The standard approach using the fitted experimental density data for the salt solution can be used to obtain the partial molar volumes.[80] Therefore, Eq. 1.13 and the experimental data can be used to calculate the three experimental G_{ij} values for a given aqueous salt solution.

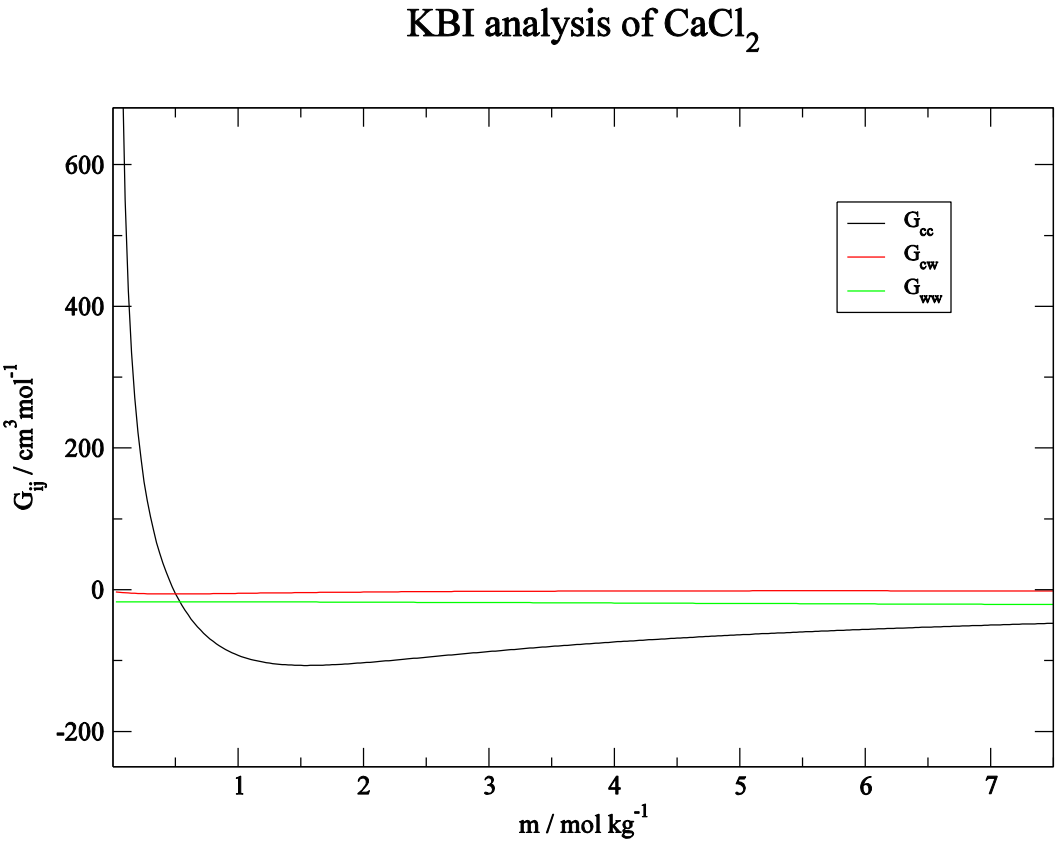

Figure 1.3. An experimental analysis of the G_{cc}, G_{cw} and G_{ww} integrals for aqueous $CaCl_2$ salt solutions at 298 K

The experimental KBI values for G_{cc}, G_{cw} and G_{ww} corresponding to an aqueous $CaCl_2$ salt solution are displayed in Figure 1.3. This indicates a typical set of KBIs for aqueous salt solutions, although all salts are slightly different. At lower concentrations, the ion - ion KB integral (G_{cc}) diverges to infinity for all salt solutions (Debye-Huckel limit). At higher salt concentrations the G_{cc} integrals adopt finite values. However, at high concentrations the KBI value for ion-ion interactions (G_{cc}) displays higher negative numbers compared to the G_{cw} for G_{ww} integrals. That implies, in the aqueous $CaCl_2$ salt solutions, that the ion-ion interactions are more unfavorable compared to ion-water interactions and water-water interactions. The ion-water interactions (G_{cw}) and the water-water interactions (G_{ww}) are similar. However, the G_{cw} values are less negative than

the G_{ww} values reflecting that ion-water interactions are preferred compare to water-water interactions.

Once the KBI values are known they can be used to determine different thermodynamic properties. The following expressions are used to calculate the activity derivatives with respect to molality (a_{cc}), and the partial molar volumes ($\bar{V}_c, \bar{V}_w$) in a binary mixture. Here w denotes water and c denotes the ions,

$$\bar{V}_c = \frac{1+\rho_w\left(G_{ww}-G_{cw}\right)}{\eta}$$

$$\bar{V}_w = \frac{1+\rho_c\left(G_{cc}-G_{cw}\right)}{\eta} \qquad \textbf{Eq. 1.16}$$

$$\eta = \rho_w + \rho_c + \rho_w\rho_c\left(G_{ww}+G_{cc}-2G_{cw}\right)$$

$$a_{cc} = \left(\frac{\partial \ln a_c}{\partial \ln \rho_c}\right)_{p,T} = \frac{1}{1+\rho_c\left(G_{cc}-G_{cw}\right)}$$

The activity coefficients (a_c) are related to the chemical potentials and the free energy of solvation of the solute. In particular, $-k_BT\ln a_c$, is the free energy of solvation of the solute, where k_B is the Boltzmann constant. As shown in Eq. 1.16 for a_{cc} it is noteworthy that G_{ww} is not included as a variable. This implies the change in the activity of a solute with changing solute concentration does not depend explicitly on the solvent-solvent distribution. Hence, the interplay between the ion-ion interactions (G_{cc}) and the ion-water interactions (G_{cw}) is important to determine the a_{cc} value.

1.7.2. Extracting KBIs from MD simulations

The atom distribution from a MD simulation is typically characterized by using radial distribution functions (rdf). Once the rdf has been obtained for $g_{cc}(r)$, $g_{cw}(r)$, and $g_{ww}(r)$ - which reflect the distributions of the ion-ion, ion-water and water-water - they can be used to calculate the corresponding KBI values. Once the corresponding KBI values are obtained they can be used to calculate different thermodynamic properties.

Figure 1.4 displays a typical ion-ion rdf obtained from a MD simulation, and used to determine the corresponding simulated KBIs for an aqueous salt solution. The value of the KBI should be determined at a distance where the rdf value converges to unity, i.e. a random distribution.

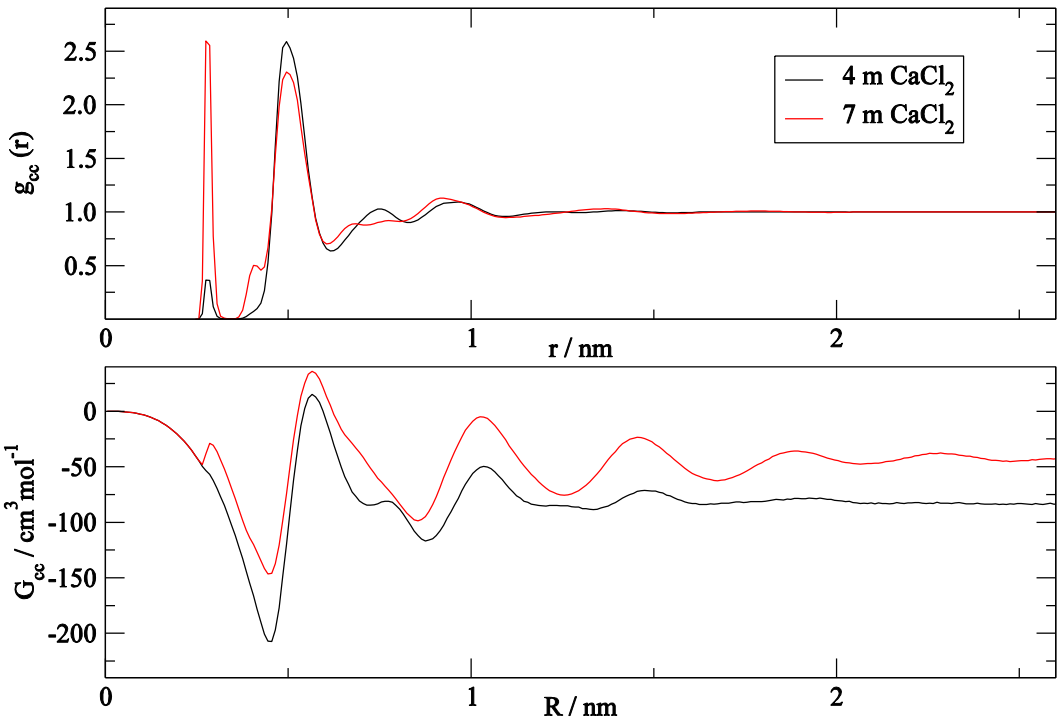

Figure 1.4. The center of mass distance (r) based radial distribution functions (rdf) and corresponding KB integrals as function of distance (R) for the ion-ion interactions obtained from a simulation of 4 mol kg^{-1} and 7 mol kg^{-1} aqueous $CaCl_2$ salt solutions.

As displayed in Figure 1.4, even a small fluctuation in the rdf distribution reflects a sizeable change in the KBI calculation at longer distances. Furthermore, increasing the salt concentration results in an increase in the net ion-ion association. This is reflected by the increasing height of the contact ion pair peak in the radial distribution function and a lower negative G_{cc} value for the 7 mol kg^{-1} salt solutions.

1.7.3. The effect of experimental errors on the KBIs

As discussed elsewhere in this thesis, in order to calculate the experimental KB integrals three types of experimental data are required. These are the isothermal compressibility, the partial molar volumes, and derivatives of the chemical potentials with respect to concentration, all as a function of composition. There are experimental errors associated with any type of experimental data. Those errors can propagate to the calculated experimental KBI values too. The general opinion is that uncertainties in the compressibility data do not impact the quality of the experimental KBI values under ambient conditions.[89] Experimental errors in the partial molar volumes contribute slightly to the calculated experimental KBI values. Indeed, it is often safe to assume a zero excess volume of mixing, i.e. composition independent partial molar volumes. The calculated KBIs are, however, very sensitive to the composition dependent chemical potential derivatives. The experimental activity data is typically obtained by measuring the experimental vapor pressures of chemical species in the mixture. These vapor pressures are fitted using different nonlinear fitting methods. These fitting methods can also introduce uncertainties which can then propagate to the calculated activity data and their derivatives, and thereby the corresponding experimental KBIs.

If there are multiple experimental datasets available for a given chemical property for a given system, then a simple comparison can provide information concerning possible errors in the KB analysis. However, in many cases, multiple experimental activity datasets for aqueous salt solutions do not exist in the literature. Therefore, it is difficult to identify possible errors. There are sizable differences in the extracted KBIs observed for lower concentration regions when using multiple data sets of activity data.[87] This coincides with systems for which the partial vapor pressure is low. Hence, one should be wary of over analyzing experimental and simulation data for compositions where the concentration of either component is very low. A more detailed discussion of the problems associated with experimental data, and the impact it can make on the extracted experimental KBIs, can be found in the literature.[89-92]

1.7.4. Development of the Kirkwood Buff derived force field (KBFF) models

The KBFF force field models were developed by Smith and coworkers at Kansas State University, see Table 1.1. In the KBFF models most of the bonded parameters are taken from the GROMOS96 force field. However, most of the non-bonded parameters, such as the Lennard-Jones parameters and the partial atomic charges, are adjusted to reproduce the experimental KBI values in MD simulations of small molecule solute-solvent systems.[93]

The KB theory of solutions can be used to quantify molecular distributions and to describe the net interactions between the solute and solvent. The sensitivity of the simulated KBI values to the partial charge distribution of a molecule is a behavior that can be used to develop more realistic effective condensed phase charge distributions, than provided by gas phase QM calculations.[94] The SPC/E model was chosen for use with KBFF models. The better diffusion coefficients and the

polarization corrections incorporated in to this water model were the main reasons to use the SPC/E water model in the KBFF implementation process.

Table 1.1. Completed KBFF models

Solute	Solvent	Reference
Acetone	water	Weerasinghe and Smith[95]
Urea	water	Weerasinghe and Smith[75]
NaCl	water	Weerasinghe and Smith[76]
GdmCl	water	Weerasinghe and Smith[96]
Methanol	water	Weerasinghe and Smith[97]
Amides	water	Kang and Smith[98]
Thiols and sulfides	methanol	Bentenitis et al[99]
Aromatics, Heterocycles	methanol, water	Ploetz and Smith[100]
Alkali halides	water	Gee et al[101]

1.7.5. Technical aspects to consider when calculating KBIs from simulation

In a MD simulation, the box size should be larger than the local correlation length between the particles. Specifically, the isothermal compressibility calculations are affected if a smaller simulation box is used.[87] Theoretically, the radical distribution function should be converged to unity at large distances in infinite (thermodynamic limit) systems. But, in all computer simulations the simulation box is actually finite. Therefore, using a smaller simulation box can introduce finite size effects which can impact the radial distribution function and calculated chemical properties.[102-103] Hence, small simulation boxes should be avoided.

The radial distribution function can show poor convergence values due to statistical noise. This can lead to very large and unrealistic fluctuations in the KBI values due to the r^2 term included in Eq. 1.11. Therefore, relatively long simulations times (many ns) may be required to achieve sufficient convergence. The integration distance R should be chosen carefully. Theoretically, the same experimental KBI value can be obtained on integrating a diverse set of radial distribution functions. However, the experimental first solvation shell values can usually be used to eliminate many, unrealistic, radial distribution functions.

In the present study ions are treated as indistinguishable in the KBI analysis. Therefore, no preference is given to either the cation or anion in the aqueous solution. However, there are some studies which use KB theory and treat the cation and anion as distinguishable species.[104] In those studies, the system is then a ternary system and one can calculate ternary KB coefficients. An excellent book is available that discusses KB theory for binary and ternary systems, and how to extract thermodynamic and the preferential solvation properties, and was published in 2013.[93]

1.7.6. KBFF models for ionic solutions

Most of the MD simulations including biomolecules need ions to neutralize the system or to set the ionic strength to mimic biological fluid environments. Hence, the parametrization of ions is crucial and one of the major challenges faced by all force field developers. A publication in 2013 claims that more complex mathematical forms should be incorporated in to the force field development of ions.[56] Here, deviations from the Lorentz-Berthelot rules, fine tuning of the water

and ion interactions, and incorporation of an exponential term instead of the r^{12} term in the non-bonded interactions, have been proposed for improved force field development.

The KBFF force field parameters for alkali halides can reproduce the experimental activity coefficient derivatives and also do not show any unphysical cation and anion aggregations in MD simulations.[101] A subsequent MD simulation performed to demonstrate the ion specific salting out behavior of benzene in aqueous solutions shows that the KBFF alkali halide parameters are effective.[105]

The modeling of the alkali halide ions using KBFF approach seems to reproduce the correct experimental structural effects in the MD simulations.[101] In this work the ion and the water interactions are scaled to model polarization effects. This strategy is discussed in detail in the next chapter. Another review published in 2016 discusses the recent progress in molecular simulations of aqueous electrolytes and claims the special treatment of cation and water oxygen interactions in KBFF is a promising strategy in the empirical force field development.[79]

Nowadays, most biomolecular force fields produce higher melting temperatures and smaller denatured volumes in protein MD simulations compared to the experimental structures.[106] However, the KBFF force field can reproduce the lower melting temperatures and comparably bigger denatured volumes due to the fact of incorporating higher solvation effects into the force field.[106] A MD study of an intrinsically disordered protein structure which compared the AMBER, CHARMM, OPLS and KBFF force fields reported that the KBFF force field produced less compact structures which are close to the experimental results.[106] Therefore, using KB theory in the force field development appears to be a promising strategy for biomolecular simulations.

1.8. Summary

Correctly modeling ions and solvation processes in classical MD simulations is crucial to achieve realistic biomolecular simulations. Since most classical MD simulations do not explicitly include electrons, polarization and screening effects that play an important role in a real system are challenging to incorporate into a force field. KB theory is a powerful strategy to capture and quantify ion and water interactions in a solution. Therefore, the focus of this thesis is to parameterize and include some biologically important ions in to the Kirkwood Buff derived force field (KBFF) to use them in classical simulations.

Chapter 2 of this thesis is the parameterization of important divalent ions. In this chapter, the aqueous alkaline earth metal halides (MX_2) where M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^- are parameterized to include in the KBFF models. Chapter 3 describes the parameterization of some biologically important oxo anions such as NO_3^-, ClO_4^-, and SO_4^{2-}. Chapter 4 investigates the parameterization of some phosphate ions such as $H_2PO_4^-$, and $(CH_3)_2PO_4^-$. Some of the high charge density ions attempted during the parametrization, such as CO_3^{2-}, HPO_4^{2-}, and PO_4^{3-}, produced unrealistic aggregation behavior in the MD simulations. Therefore, as a way to overcome this problem we will explore the possibility of including electronic screening, without changing the charge of the ion, by simply changing the water model dipole moment to be closer to the experimental value. Hence, chapter 5 explores the possibility of using a modified water model to solve the excessive ion clustering in the MD simulations. The appendix describes a separate project describing a theoretical analysis of Gaussian and Non-Gaussian fluctuations in pure classical fluids that was recently published.

References

1. National Research Council . Committee on Challenges for the Chemical Sciences in the 21st, C., *Beyond the molecular frontier : challenges for chemistry and chemical engineering*. Washington, D.C. : National Academies Press: 2003.

2. Mattson, T.; Wrinn, M., Parallel programming. In *Proceedings of the 45th annual conference on Design automation - DAC '08*, ACM Press: New York, New York, USA, 2008; p 7.

3. Barnett, R. N.; Landman, U., *Physical Review B* **1993,** *48*, 2081-2097.

4. Ponder, J. W.; Case, D. A., *Advances in protein chemistry* **2003,** *66*, 27-85.

5. Born, M.; Oppenheimer, J., *Ann. Phys. (Leipzig)* **1924,** *74*, 1-31.

6. Meller, J., Molecular Dynamics. In *Encyclopedia of Life Sciences*, John Wiley & Sons, Ltd: Chichester, 2001.

7. Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E., *Annual Review of Biophysics* **2012,** *41* (1), 429-452.

8. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: Oxford [England] :, 1989; p 385.

9. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E., *The Journal of Chemical Physics* **1953,** *21* (6), 1087-1092.

10. Chowdhuri, S.; Chandra, A., *The Journal of Chemical Physics* **2001,** *115*, 3732.

11. Alder, B. J.; Wainwright, T. E., *The Journal of Chemical Physics* **1957,** *27*, 1208.

12. Alder, B. J.; Wainwright, T. E., *The Journal of Chemical Physics* **1959,** *31* (2), 459-466.

13. Rahman, A., *Physical Review* **1964,** *136*, A405-A411.

14. Stillinger, F. H.; Rahman, A., *The Journal of Chemical Physics* **1974,** *60*, 1545-1557

15. McCammon, J. A.; Gelin, B. R.; Karplus, M., *Nature* **1977,** *267*, 585-590.

16. Shaw, D. E.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Dror, R. O.; Piana, S.; Shan, Y.; Towles, B.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B. In *Millisecond-scale molecular dynamics simulations on Anton*, Proceedings of the Conference on High Performance Computing Networking, Storage and Analysis - SC '09, New York, USA, ACM Press: New York, USA, 2009; p 1.

17. Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S., *Current Opinion in Structural Biology* **2013,** *23*, 58-65.

18. Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F., *European Biophysics Journal* **2011,** *40*, 843-856.

19. Jin, Z.; Yang, C.; Cao, F.; Li, F.; Jing, Z.; Chen, L.; Shen, Z.; Xin, L.; Tong, S.; Sun, H., *Journal of Computational Chemistry* **2016,** *37*, 653-664.

20. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. SOC* **1995,** *117*, 5179-5197.

21. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *Journal of Computational Chemistry* **1983,** *4*, 187-217.

22. Van Gunsteren, W. F.; Berendsen, H. J. C., *Library manual (Nijenborgh, Groningen)* **1987**, 1-221.

23. Jorgensen, W. L.; Tirado-Rives, J., *Journal of the American Chemical Society* **1988,** *110*, 1657-1666.

24. Ploetz, E. A.; Bentenitis, N.; Smith, P. E., *Fluid Phase Equilibria* **2010,** *290*, 43-47.

25. Karplus, M., *Accounts of Chemical Research* **2002,** *35*, 321-323.

26. Stillinger, F. H., *The Journal of Chemical Physics* **1979,** *71* (4), 1647-1651.

27. Wang, J.; Cieplak, P.; Cai, Q.; Hsieh, M.-J.; Wang, J.; Duan, Y.; Luo, R., *The Journal of Physical Chemistry B* **2012,** *116* (28), 7999-8008.

28. Goyal, P.; Qian, H.-J.; Irle, S.; Lu, X.; Roston, D.; Mori, T.; Elstner, M.; Cui, Q., *The Journal of Physical Chemistry B* **2014,** *118* (38), 11007-11027.

29. Rick, S. W.; Stuart, S. J.; Berne, B. J., *The Journal of Chemical Physics* **1994,** *101* (7), 6141-6156.

30. Tröster, P.; Lorenzen, K.; Schwörer, M.; Tavan, P., *The Journal of Physical Chemistry B* **2013,** *117* (32), 9486-9500.

31. Neilson, G. W.; Enderby, J. E. In *Water and aqueous solutions* proceedings of the Thirty-seventh Symposium of the Colston Research Society, University of Bristol A. Hilger: University of Bristol 1985; p 349.

32. Guillot, B., *Journal of Molecular Liquids* **2002,** *101*, 219-260.

33. Fuentes-Azcatl, R.; Barbosa, M. C., *Physica A: Statistical Mechanics and its Applications* **2016,** *444*, 86-94.

34. Rahman, A.; Stillinger, F. H. J., *The Journal of Chemical Physics* **1971,** *55*, 3336-3359.

35. Gordon, M. S.; Jensen, J. H., *Accounts of Chemical Research* **1996,** *29* (11), 536-543.

36. Spackman, M. A., *Chemical Reviews* **1992,** *92*, 1769-1797.

37. Bratko, D.; Blum, L.; Luzar, A., *The Journal of Chemical Physics* **1985,** *83*, 6367.

38. Ichiye, T.; Tan, M.-L., *The Journal of chemical physics* **2006,** *124*, 134504.

39. Vega, C.; Monson, P. A., *Molecular Physics* **1995,** *85*, 413-421.

40. Kusalik, P. G.; Patey, G. N., *The Journal of Chemical Physics* **1988,** *88*, 7715.

41. Leontyev, I.; Stuchebrukhov, A., *Physical Chemistry Chemical Physics* **2011,** *13* (7), 2613.

42. Ichiye, T., Water in the Liquid State: A Computational Viewpoint. In *Advances in Chemical Physics*, John Wiley & Sons, Inc.: 2014; pp 161-200.

43. Schropp, B.; Tavan, P., *Journal of Physical Chemistry B* **2008,** *112*, 6233-6240.

44. Leontyev, I. V.; Stuchebrukhov, A. A., *The Journal of Chemical Physics* **2009,** *130*, 085102.

45. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *The Journal of Chemical Physics* **1983,** *79*, 926.

46. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J., Interaction Models for Water in Relation to Protein Hydration. Springer Netherlands: 1981; pp 331-342.

47. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

48. Jorgensen, W. L.; Madura, J. D., *Molecular Physics* **1985,** *56*, 1381-1392.

49. Mahoney, M. W.; Jorgensen, W. L., *The Journal of Chemical Physics* **2000,** *112*, 8910.

50. Hummer, G.; Pratt, L. R.; García, A. E., *The Journal of Physical Chemistry* **1996,** *100*, 1206-1215.

51. Vener, M. V.; Leontyev, I. V.; Dyakov, Y. A.; Basilevsky, M. V.; Newton, M. D., *The Journal of Physical Chemistry B* **2002,** *106*, 13078-13088.

52. van der Spoel, D.; van Maaren, P. J.; Berendsen, H. J. C., *The Journal of Chemical Physics* **1998,** *108*, 10220.

53. Vega, C.; Abascal, J. L. F., *Physical Chemistry Chemical Physics* **2011,** *13*, 19663.

54. Nguyen, T. T.; Viet, M. H.; Li, M. S.; Nguyen, T. T.; Viet, M. H.; Li, M. S., *TheScientificWorldJournal* **2014,** *2014*, 536084.

55. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Computer Physics Communications* **1995,** *91*, 43-56.

56. Moučka, F.; Nezbeda, I.; Smith, W. R., *Journal of Chemical Theory and Computation* **2013,** *9*, 5076-5085.

57. Record, M. T.; Zhang, W.; Anderson, C. F., *Advances in protein chemistry* **1998,** *51*, 281-353.

58. Lo Nostro, P.; Ninham, B. W., *Chemical reviews* **2012,** *112*, 2286-322.

59. Knipping; Lakin; Foster; Jungwirth; Tobias; Gerber; Dabdub; Finlayson-Pitts, *Science (New York, N.Y.)* **2000,** *288*, 301-6.

60. Stirnemann, G.; Wernersson, E.; Jungwirth, P.; Laage, D., *Journal of the American Chemical Society* **2013,** *135*, 11824-11831.

61. Fayer, M. D.; Moilanen, D. E.; Wong, D.; Rosenfeld, D. E.; Fenn, E. E.; Park, S., *Accounts of Chemical Research* **2009,** *42* (9), 1210-1219.

62. Marcus, Y., *Chemical reviews* **2009,** *109*, 1346-70.

63. Ohtaki, H.; Radnai, T., *Chemical Reviews* **1993,** *93*, 1157-1204.

64. Tielrooij, K. J.; Garcia-Araez, N.; Bonn, M.; Bakker, H. J., *Science (New York, N.Y.)* **2010,** *328*, 1006-9.

65. Ropp, J.; Lawrence, C.; Farrar, T. C.; Skinner, J. L., *Journal of the American Chemical Society* **2001,** *123*, 8047-8052.

66. Hofmeister, F., *Archiv für Experimentelle Pathologie und Pharmakologie* **1888,** *24*, 247-260.

67. J. Šponer; M. Otyepka; P. Banáš; Réblová, K.; Walter, N. G., Molecular Dynamics Simulations of RNA Molecules. In *Innovations in biomolecular modeling and simulations*, Royal Society of Chemistry: 2012; Vol. 2, pp 129-155.

68. Rajamani, S.; Ghosh, T.; Garde, S., *Journal of Chemical Physics* **2004,** *120*, 4457-4466.

69. Chialvo, A. A.; Gruszkiewicz, M. S.; Cole, D. R., *Journal of Chemical & Engineering Data* **2010,** *55*, 1828-1836.

70. Marcus, Y.; Hefter, G., Ion pairing. In *Chemical Reviews*, American Chemical Society: 2006; Vol. 106, pp 4585-4621.

71. Kohagen, M.; Pluhařová, E.; Mason, P. E.; Jungwirth, P., *The Journal of Physical Chemistry Letters* **2015,** *6*, 1563-1567.

72. Vrbka, L.; Jungwirth, P.; Bauduin, P.; Touraud, D.; Kunz, W., *J. Phys. Chem. B* **2006,** *110*, 7036-7043.

73. Barbara Jagoda-Cwiklik; Robert Vácha; Mikael Lund; Monika Srebro; Jungwirth, P., *J. Phys. Chem* **2007,** *111* (51), 14077–14079.

74. Uejio, J. S.; Schwartz, C. P.; Duffin, A. M.; Drisdell, W. S.; Cohen, R. C.; Saykally, R. J., *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105* (19), 6809-6812.

75. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2003,** *107*, 3891-3898.

76. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2003,** *119*, 11342.

77. Mamatkulov, S.; Fyta, M.; Netz, R. R., *The Journal of Chemical Physics* **2013,** *138*, 024505.

78. Jiang, H.; Panagiotopoulos, A. Z., *The Journal of Chemical Physics* **2016,** *145*, 046101.

79. Nezbeda, I.; Moučka, F.; Smith, W. R., *Molecular Physics* **2016,** *8976*, 1665-1690.

80. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19*, 774.

81. Vafaei, S.; Tomberli, B.; Gray, C. G., *Journal of Chemical Physics* **2014,** *141* (15).

82. Ben-Naim, A., *Molecular Theory of Solutions*. Oxford University Press: 2009; p 380.

83. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67*, 4884.

84. Smith, P. E., *The Journal of chemical physics* **2008,** *129*, 124509.

85. Canchi, D. R.; Paschek, D.; García, A. E., *Journal of the American Chemical Society* **2010,** *132*, 2338-2344.

86. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E., *Cell biochemistry and biophysics* **2008,** *50*, 1-22.

87. Smith, P. E.; Matteoli, E.; O'Connell, J. P., *Fluctuation theory of solutions : applications in chemistry, chemical engineering, and biophysics*. CRC Press: 2013; p 368.

88. Kusalik, P. G.; Patey, G. N., *The Journal of Chemical Physics* **1987,** *86* (9), 5110-5116.

89. Matteoli, E.; Lepori, L., *The Journal of Chemical Physics* **1984,** *80* (6), 2856-2863.

90. Matteoli, E.; Lepori, L., *Journal of the Chemical Society, Faraday Transactions* **1995,** *91* (3), 431-436.

91. Ploetz, E. A.; Smith, P. E., *The Journal of Chemical Physics* **2015,** *142* (9), 094504.

92. Ploetz, E. A.; Rustenburg, A. S.; Geerke, D. P.; Smith, P. E., *Journal of Chemical Theory and Computation* **2016,** *12* (5), 2373-2387.

93. Smith, P. E.; Matteoli, E.; O' Connell, J. P., *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*. CRC Press: Boca Raton, 2013.

94. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2003,** *118* (13), 5901-5910.

95. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2003,** *118* (23), 10663-10670.

96. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2004,** *121* (5), 2180-2186.

97. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

98. Kang, M.; Smith, P. E., *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

99. Bentenitis, N.; Cox, N. R.; Smith, P. E., *The Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

100. Ploetz, E. A.; Smith, P. E., *Physical Chemistry Chemical Physics* **2011,** *13* (40), 18154-18167.

101. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E., *Journal of Chemical Theory and Computation* **2011,** *7* (5), 1369-1380.

102. Salacuse, J. J.; Denton, A. R.; Egelstaff, P. A., *Physical Review E* **1996,** *53*, 2382-2389.

103. Perera, A.; Zoranić, L.; Sokolić, F.; Mazighi, R., *Journal of Molecular Liquids* **2011,** *159*, 52-59.

104. Schnell, S. K.; Englebienne, P.; Simon, J.-M.; Krüger, P.; Balaji, S. P.; Kjelstrup, S.; Bedeaux, D.; Bardow, A.; Vlugt, T. J. H., *Chemical Physics Letters* **2013,** *582*, 154-157.

105. Ganguly, P. Modeling and Understanding Aqueous Mixtures Using Kirkwood-Buff Theory of Solutions. Technische Universität Darmstadt 2014.

106. Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Gräter, F., *Journal of Physical Chemistry B* **2015,** *119*, 7975-7984.

Chapter 2 - Molecular Dynamics Simulations of Alkaline Earth Halide Solutions

2.1. Abstract

The activity and function of many macromolecules in cellular environments are coupled with the binding of divalent ions such as calcium or magnesium. In principle, computer simulations can be used to understand molecular level aspects of how important macromolecules interact with ions. However, most of the force fields available today often fail to accurately reproduce the properties of divalent ions in aqueous environments. Here we develop a classical non polarizable force field for aqueous alkaline earth metal halides (MX_2) where M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^- for bimolecular simulations which is compatible with the SPC/E water model. These new force field parameters are specifically developed to reproduce the experimental Kirkwood-Buff integrals. Since this new force field can reproduce the experimental Kirkwood-Buff integrals for most concentrations of the respective salts, they are also capable of reproducing the experimental activity derivatives, partial molar volumes, and excess coordination numbers. Use of these new models in MD simulations also leads to reasonable diffusion constants and dielectric decrements.

2.2. Introduction

Aqueous electrolyte solutions play a very important role in biological systems and also in terrestrial and marine environments. Under physiological conditions divalent ions are involved in many important processes such as nucleic acids and protein folding, the activation of enzymes, and cellular signal transduction, etc.[1-2] As an example, although it is well established that many

enzymatic reactions involving nucleic acids are known to rely on the cooperative behavior of divalent cation such as magnesium, the functional role and the reaction mechanism are still a matter of debate.[3] Nevertheless, nucleophilic attack through a water molecule or a hydroxide ion is widely accepted to be the reason for the cleavage of the phosphodiester linkage in the presence of metal ions. Divalent ions can bind strongly with biomolecules, which leads to fundamental structural and functional activity of proteins and nucleic acids. $Mg^{2+}$and Ca^{2+} ions can act very differently in the context of blood clotting where only Ca^{2+} ions can act as a cofactor.[4] However, the correlation between the size of the ions, their electronegativity, or their coordination number and the regulation of enzyme activity in the presence of the ions are still not fully understood.[5] Magnesium ion channels and transporters transport Mg^{2+} ions against a high background concentration of its major competitor, Ca^{2+} ions. The major determinants of Mg^{2+}/Ca^{2+} selectivity in Mg^{2+} ion channels are still unknown[6], because the experimental structures of Mg^{2+} ion channels in an open conformation with bound Mg^{2+} have not yet been resolved. The degree of metal ion hydration, which correlates with the pore size and rigidity, is argued to be a major factor.[6]

We can use computer simulations as an atomistic microscope to assist experimentalist to capture phenomenon, which are not accessible through conventional experimental techniques, with high special and temporal resolutions. However, an accurate and realistic representation of ion solvation in aqueous solution is crucial to achieve this goal. A realistic computational model should accurately capture the relevant microscopic interactions between ions and water molecules. However, it is still not trivial to achieve the required fine balance between the ion-ion, ion-water and water-water interactions in a computational model. In particular, how to include significant contributions from polarization effects, which may be very important in the case of multivalent metal ions, to the solvation using conventional non-polarizable force fields is still unclear.

Unfortunately, standard non polarizable water models used in computer simulations do not explicitly include electronic polarization. To overcome this issue, most water models attempt to include the average effect of electronic polarization by its neighboring water molecules by simply increasing the dipole moment from the gas phase value.[7-8] Hence, to make the molecules compatible with water models need to adjust the partial charges to represent screening and polarization effect of the solvent. This strategy can easily applicable to neutral molecules. However, this is not an effective strategy in term of ions. The intrinsic nature of ions with fixed integer charges leads to incompatibly of the standard water models and ions.

There are many more parameter sets available for monovalent ions[9-10] in the force field literature compared to divalent ions. This may be due to inherent difficulty modeling the higher charge ions and their solvent interactions. Ion interactions with biomolecules can be the result of direct binding or indirect binding. Direct binding occurs in places like a binding pocket of a biomolecule where negatively charged chemical groups interact directly with cations. Indirect binding of metal ions can also occur where cations interact with negatively charged chemical groups through bonds mediated by water molecules. However, the properties of water molecules bound to a divalent cation can be drastically different from the bulk water properties.[11]

On the nanoscale, hydration repulsion dominates the interactions between well solvated polar surfaces and prevent the sticking together of biological matter.[12-13] An experimental and simulations study in 2016 reported that polysaccharides coated nanoparticles show ion specific colloidal stability in the presence of $MgCl_2$ and $CaCl_2$.[14] This study also reported that $CaCl_2$ enhances the colloidal particle stability, while $MgCl_2$ lowers the stability, inducing nanoparticle aggregation. Furthermore, MD simulations were used to explain this phenomenon and suggested that surface attached Ca^{2+} ions promote better hydration and also induce strongly polarized

repulsive water structure beyond at least 3 hydration shells around the carbohydrate. However, all the effect of specific ions on the hydration repulsion are still not fully understood. [14]

One of the earliest MD simulations of Mg^{2+} ion was performed in 1982.[15] Here, a 3.3 picosecond simulation was done in a 18.30 Å cubic box which contained 200 water and 4 $MgCl_2$ molecules. In 1985 a MD simulation was performed to investigate the structural properties of aqueous $CaCl_2$ solution.[16] Both of these studies used ab initio derived data to model the effective pair potential between the ions and water. In 1990 a simulation performed by Åqvist used calculated hydration free energies of the Mg^{2+}, Ca^{2+}, Sr^{2+} and Ba^{2+} chlorides to derive condensed phase ion-solvent interaction parameters.[17] A potential of mean force (PMF) study performed in 1992 on a calcium chloride ion pair is one of the first for divalent alkaline earth metal ions.[18] In 2006, Gavryushov published an effective ion-ion potential for all alkaline earth metal halides with the SPC/E water model,[19] but this study uses approximations for the ion hydration shell polarization and only short range electrostatic interactions are considered.

Although Mg^{2+} ions play a very important role in the activity of nucleic acids, Hashem et al. reported that up to 2008 only 22 out of 113 RNA simulations used Mg^{2+} as one of the ions.[20] This review highlighted the difficulty of modeling ions with high charge densities as one reason Mg^{2+} ions are avoided in RNA simulations. Due to this high charge density, the water molecules bound to Mg^{2+} ions display very long residence times compared to monovalent cations.[21] Therefore, current MD simulations cannot capture the desolvation process of Mg^{2+} ions.[22] A study in 2012 reported that MD simulations of Mg^{2+} ions with a DNA molecule can result in considerable artifacts for the AMBER[23] and CHARMM[24] force fields.[25] This work suggested fine tuning the van der Waals interactions parameter for the ions to reproduce experimental osmotic pressure as a promising strategy.

As reviewed by Auffinger et al.[26], no MD simulations of Ca^{2+}, Sr^{2+} and Ba^{2+} ions with RNA molecules are available in the literature up to 2012, despite those ions being found in the experimental crystallographic structures.[27] Many MD simulations have appeared in the literature where the crystallographic biomolecular structure originally contains Ca^{2+}, Sr^{2+} and Ba^{2+}, but these are replaced by Mg^{2+} ions in simulations. Alternatively, Mg^{2+} ions are ignored or replace by two monovalent Na^{+} ions in some biomolecular MD studies. The review also highlights that, although including divalent ions in MD simulations is a challenging task, it is necessary to perform realistic RNA simulations. A recent study shows that the AMBER and CHARMM force fields for Ca^{2+} ions form artificial clusters in MD simulations also containing chloride, acetate and phosphate ions. These salt solutions also fail to reproduce experimental osmotic pressures.[28] The same work indicated that MD simulations of the Ca^{2+} ions mediated DNA-DNA interactions and leads to strong inter DNA attractions. However experimentally the DNA molecules repel each other.[28] It is the failure to capture the correct solvation of Ca^{2+} ions by standard force fields, leading to an artificial attraction of the Ca^{2+} ions and the phosphate groups in the DNA molecules, that appears to be responsible for this inconsistency.

Another study in 2015 attempted to model the strong solvation shells around Mg^{2+} ions by using a refined dummy atom model.[29] In this model a single Mg^{2+} ion is represented as an octahedron with a metal center with a -1 charge covalently bound by six dummy atoms. Each of the dummy atoms possesses a +0.5 charge. This work attempts to optimize the AMBER ff03 force field to get better performance for the experimental geometries and thermodynamic properties. However, Saxena et al. reported that multi-site models of the Ca^{2+} and Mg^{2+} ions in the CHARMM force field overestimate the attraction of the cation to chloride ions resulting in unusual ion-ion

pairing and clustering in the MD simulations.[30-31] This also leads to osmotic pressures that were inconsistent with experimental results.

Another excellent review in 2015 compared 17 widely used force fields for Mg^{2+} ions to evaluate their ability to reproduce the experimental structural, thermodynamic, kinetic and mass transport properties.[32] In this study, all the models using a simple 12-6 LJ representation failed to reproduce the structural and thermodynamically properties simultaneously with reasonable accuracy. The study suggested that exploring 12-6-4 LJ potential might be the right direction for the next generation of divalent cation force fields.

Most of the commonly used force field ion parameters are typically optimized using single ion properties and the experimental data for the crystalline state. Hence, they cannot necessarily reproduce experimental thermodynamic properties at finite concentrations.[17, 33-35] Mamatkulov et al. reported that widely used divalent ion force field parameters can only reproduce the experimental solvation free energy, and the first peak of water ion radial distribution function with reasonable accuracy at low salt concentrations.[35] Furthermore, this study emphasized the necessity of refining existing divalent ion parameters as none of the widely used force fields, such as AMBER[23], CHARMM[24] or GROMOS,[36] are capable of describing the ion specific effect of divalent ions in ion channels.

Therefore, there is huge a potential to develop a force field for the alkaline earth ions which can reproduce correct activity, osmotic and solvation properties at finite concentrations. The new force field should be able to capture the fine balance between ion and water interactions. The Kirkwood Buff (KB) theory was used successfully in a previous publication to model alkaline halides to overcome limitations of reproducibility of correct thermodynamic and solvation

properties in MD simulations.[9-10] That same approach is attempted to develop parameters for the alkaline earth halides.

2.3. Experimental Analysis and MD simulations

2.3.1. Experimental KB inversion analysis

As discussed in chapter one, the Kirkwood Buff derived force field (KBFF) use KB theory[37] in the parameterization process. At the heart of KB theory are the KB integrals (KBI) defined by,

$$G_{ij} = \int_0^R \left[g_{ij}(r) - 1 \right] 4\pi r^2 \, dr \qquad \textbf{Eq. 2.1}$$

Here, $g_{ij}(r)$ is the pair correlation function in a system with two different species i, j and r is the intermolecular separation. The R is a cutoff distance from the particle of interest. The KB integrals (G_{ij}) quantify solute and solvent interactions in a solution. Three types of KB integrals (G_{cc}, G_{ww}, $G_{cw} = G_{wc}$) are sufficient to describe all the interactions of a binary solution consisting of water (w) and an ionic cosolvent (c). The cations and anions are treated as indistinguishable in this approach, and are referred to as a general cosolvent.[9, 38] These three KB integrals, together with the number densities (ρ_c and ρ_w), can be used to determine different thermodynamic quantities. The KB inversion procedure,[39-40] was used together with the composition dependent experimental data for the isothermal compressibility, the partial molar volumes, and the cosolvent activity, in order to extract the corresponding three experimental KB integrals. The relationships used in this study are displayed as Eq. 2.2.[40]

$$1 + N_{cc} = \rho_c RT\kappa_T + \rho_w^2 \frac{\bar{V}_w^2}{\mu_{cc}}$$

$$1 + N_{ww} = \rho_w RT\kappa_T + \rho_w \rho_c \frac{\bar{V}_c^2}{\mu_{cc}} \qquad \textbf{Eq. 2.2}$$

$$N_{wc} = \rho_c RT\kappa_T + \rho_w \rho_c \frac{\bar{V}_w \bar{V}_c}{\mu_{cc}}$$

Here, N_{ij} is the excess coordination number, R is the gas constant, T is the temperature, κ_T is the isothermal compressibility, $\bar{V}_i$ is the partial molar volume, and μ_{cc} is the chemical potential derivative given by,

$$\mu_{cc} = \beta\left(\frac{\partial \mu_c}{\partial \ln m_c}\right)_{T,P} = 1 + \left(\frac{\partial \ln \gamma_c}{\partial \ln m_c}\right)_{T,P} \qquad \textbf{Eq. 2.3}$$

where γ_c=$\gamma_\pm$ is the molal activity coefficient of the salt and m_i is the molality of species i.

The experimental activity coefficients and densities for alkaline earth metal halides solutions (MX_2), with M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^- , where obtained from the literature.[41-42] The experimental activity derivatives were fitted to the Pitzer equation[43-44] and the densities were fitted to a simple polynomial expression.[41] For a single electrolyte, $M_{\nu+}X_{\nu-}$, the Pitzer equation is written as following.

$$\ln \gamma_{\pm} = \left|z_+ z_-\right| f^{\gamma} + m\left(\frac{2\nu_+\nu_-}{\nu}\right) B^{\gamma} + m^2\left(\frac{2\left(\nu_+\nu_-\right)^{3/2}}{\nu}\right) C^{\gamma}$$

$$f^{\gamma} = -A_{\phi}\left[\frac{\sqrt{I/m^0}}{1+b\sqrt{I/m^0}} + \frac{2}{b}\ln\left(1+b\sqrt{I/m^0}\right)\right] \qquad \textbf{Eq. 2.4}$$

$$B^{\gamma} = 2\beta^{(0)} + \frac{2\beta^{(1)}}{\alpha^2\left(I/m^0\right)}\left[1-\exp\left(-\alpha\sqrt{I/m^0}\right)\left(1+\alpha\sqrt{I/m^0}-(1/2)\alpha^2\left(I/m^0\right)\right)\right]$$

$$C^{\gamma} = (3/2)C^{\phi}$$

$$A_{\phi} = \frac{1}{3}\left(2\pi N_A \rho_A\right)^{1/2}\left[\frac{e^2}{4\pi\varepsilon_0 DkT}\right]^{3/2}$$

Here, I is the ionic strength on a molality scale, Z is the charge number of ions and the $\nu = \nu_+ + \nu_-$ is the number of ions dissociated in one unit of electrolyte formula. The $\beta^{(0)}, \beta^{(1)}$ and C^{ϕ} are parameters of the Pitzer equation. In addition, $\alpha = 2\ kg^{1/2}\ mol^{-1/2}$ and $b = 1.2\ kg^{1/2}\ mol^{-1/2}$ and the Debye-Huckel coefficient for the osmotic coefficient is given by A_{ϕ}. Finally, N_A, ρ_A, D, k and ε_0 are the Avogadro number, the density of the mixed solvent, the dielectric constant, the Boltzmann constant, and the vacuum permittivity, respectively.

Experimental KBI were obtained by using those experimental activity and density data for the solubility range for all salts (MX_2) where M = Mg^{2+}, Ca^{2+}, Sr^{2+}, Ba^{2+} and X = Cl^-, Br^-, I^- solutions using the KB inversion procedure. The experimental partial molar volumes are obtained from the experimental densities using previously established standard approaches.[37, 41] The experimental compressibility of salt solution is obtained by using the same approach described in the chapter one. More details about KB inversion procedure to obtain KBI using experimental data is explained in detail in earlier works.[9-10, 45]

2.3.2. Parameterization of the alkaline metal ions

Aqueous ion solutions comprise only the ions and water molecules. The water models are already established and the compatible water model for KBFF force field is SPC/E.[46] However, the ions are monatomic species. Therefore, only the non-bonded interactions contribute to the potential energy of a given system via the usual Coulomb and LJ potentials,

$$E_{Electrostatics} = \sum_{i,j} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad \textbf{Eq. 2.5}$$

$$E_{Van\ der\ Waals} = \sum_{i,j} 4\varepsilon_{ij} \left[\left[\frac{\sigma_{ij}}{r_{ij}} \right]^{12} - \left[\frac{\sigma_{ij}}{r_{ij}} \right]^{6} \right]$$

Hence, only 3 parameters; the charge of the ion, the Lennard Jones interaction length (σ), and the interaction strength (ε) are required to model an ion in a simple force field. Since the charge of an ion is fixed this results in just two adjustable parameters, σ and ε, to include an ion in the KBFF force field.

The inclusion of missing polarization effects in simple non polarizable force fields, by perturbing ion and water interaction strength by a scaling factor as shown in Eq. 2.6, gave promising results in our parameterization of alkali metal halide ions.[9-10] The same strategy is used in this project to parameterize the alkaline earth halide salts. This strategy not only changes the ion-water interaction well depth, but also modifies the repulsive wall of the LJ potential well, as displayed in the Figure 2.1. This is an implicit way to modify the ion-water interactions at a shorter distances due to account for polarization effects.

$$\varepsilon_{ij} = s\sqrt{\varepsilon_{ii} \times \varepsilon_{jj}} \qquad \textbf{Eq. 2.6}$$

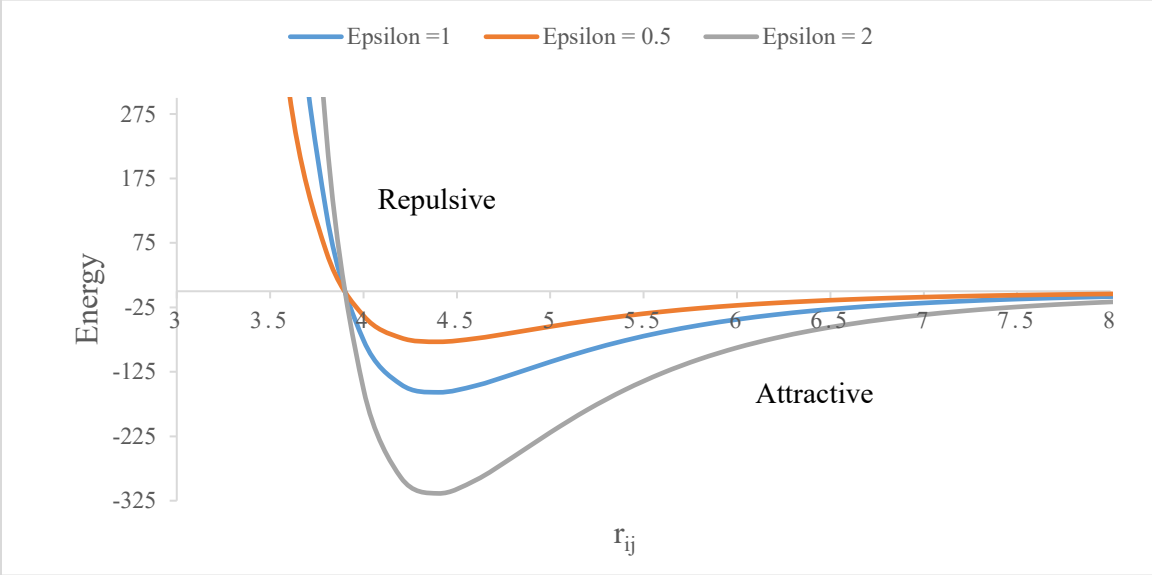

Figure 2.1. Impact on the Lennard Jones potential of the scaling of the epsilon value without changing the sigma value. By scaling the interaction strength the cation and water interactions at the closer distances (repulsive regime) are also modified.

Since Cl^-, Br^-, I^- already have been parameterized in the KBFF force field, we only need to parameterize the metal cations in the present work.[10] Therefore, as the first step of the parameterization process, we use data for the alkaline earth metal chlorides to help determine LJ interaction length (σ), the interaction strength (ε), and the water to metal cation interaction scaling factor (s). Then successful transferability of those scaling factors to model the salt interactions of the bromide and iodide solutions were investigated.

To determine the LJ interaction length (σ) and the interaction strength (ε) of the alkaline earth metal cations two pieces of experimental data were used. Those are the ionic radii consistent with the crystal lattice unit dimensions and the ion to water contact distances.[47-48] Those values are shown in the Table 2.1.

Table 2.1. Experimental data for alkaline earth cations. r, is the ionic radii of the alkaline earth ions, a,b,c are the crystal unit cell dimensions, and d is the cation to water oxygen contact distance.

	MCl_2			
	Mg^{2+}	Ca^{2+}	Sr^{2+}	Ba^{2+}
r (nm)	0.066	0.099	0.112	0.134
a (nm)	0.3596	0.624	0.69767	0.7865
b (nm)	0.3596	0.643	0.69767	0.4731
c (nm)	1.7589	0.42	0.69767	0.9421
d (nm)	0.209	0.242	0.264	0.29

To determine the interaction length (σ) and the interaction strength (ε) for the alkaline earth metal chlorides a set of crystal structure and aqueous salt solution MD simulations were performed. The values of the ε_{++} were varied to obtain the correct experimental lattice dimensions and water-cation contact distances. The obtained results for alkaline earth cations in present work are shown in the Table 2.2. The σ increases as the size of the cation increases. However, the ε decreases as the size of the cation increases. Unfortunately, these values in Table 2.2 for sigma and epsilon alone could not reproduce the experimental KBI values in the simulations.

Table 2.2. The LJ sigma (σ), LJ epsilon (ε), LJ cation-water ε, scaling factor (s), charge (q) of the ions(q). For the combination rules $\varepsilon_{ij} = s\,(\varepsilon_{ii} * \varepsilon_{jj})^{0.5}$ and $\sigma_{ij} = (\sigma_{ii} * \sigma_{jj})^{0.5}$ were used in this work.

Model	Atom	Sigma σ_{ii} (nm)	Epsilon ε_{ii} (kj/mol)	Cation-Water ε_{i-OW} (kj/mol)	Scaling Factor	Charge q (e)	Ref
KBFF	Mg^{2+}	0.2100	0.7500	0.0698	0.1	2	
	Ca^{2+}	0.2900	0.4700	0.3871	0.7	2	
	Sr^{2+}	0.3100	0.5000	0.5703	1.0	2	
	Ba^{2+}	0.3800	0.2000	0.4329	1.2	2	
	Cl^-	0.4400	0.4700			-1	[9]
	Br^-	0.4760	0.3000			-1	[10]
	I^-	0.5350	0.2000			-1	[10]
SPC/E	O	0.3166	0.6506			-0.8476	[46]
	H	0.0000	0.0000			0.4238	

Hence, the metal cation to water interaction strength were then scaled by different factors in an attempt to reproduce the experimental KBIs. The final scaling factors are given in Table 2.2. Once the scaling factors were established for the alkaline earth metal chlorides, a series of simulations were performed to analyze the transferability of those same scaling factors to bromides and iodides.

2.3.3. MD simulation details

All the computer simulations were performed by using GROMACS 4.6 version software.[49] The SPC/E water model[46] used to model water in the simulations. All simulations in this study were maintained in the isothermal isobaric ensemble at 300 K temperature and 1 atm pressure. The Berendsen pressure and temperature coupling was used in this study.[50] Here, temperature and pressure are weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively. Also a 2 fs time step was used to integrate the equation of motion. All the bonds in the simulations were constrained with the LINCS algorithm.[51] The particle mesh Ewald (PME) approach was used to model electrostatic interactions with 1 nm cut off distance for real space calculations,[52] with a 1.5 nm cutoff distance for the Van der Waals interactions. All the simulations were performed in a 10 nm size cubic box. A random initial configuration of known number of ions and water in a cubic box with length of 10 nm was generated to provide a known concentration of aqueous salt solution using a custom written Fortran code. An energy minimization using the steepest descend method followed by 2 ns of equilibration was performed before the production run. The production run of 15 ns was used to calculate ensemble averages. In the production run, configurations were saved every 0.1 ps for analysis.

For the salt crystals, anisotropic MD simulations were performed at 300 K and 1 atm. Here, roughly 5 nm length crystals were generated using a custom written Fortran code. The unit cell dimensions and the symmetry group operations need to generate salt crystals were obtained from the crystallography open database.[48] However, the $MgCl_2$, $MgBr_2$, MgI_2 and CaI_2 , which have non orthogonal crystal structures with angles of 90°, 90°, 120°, in the nature were not stable in the MD simulations. Therefore, orthogonal crystals (90°,90°,90°) were also used for analaysis.

A custom written Fortran program was used to calculate center of mass radial distribution function and to calculate the simulated KB integrals. Here, a center of mass RDF calculation was performed. For some of the systems the accuracy of the KBI values was verified by calculating the integrals using the equivalent particle fluctuation approach.[53] The simulated KBI integrals were use to calculate the activity derivatives, excess coordination numbers, and partial molar volumes. Furthermore, translational self-diffusion coefficients were calculated using the mean square fluctuation approach,[54-55] dielectric coefficients were calculated by the mean dipole fluctuations analysis,[56-57] and the enthalpy of mixing determined from the average potential energies of a molecule in the solution compared with that of the pure salt crystal and pure water.

2.4. Results and Discussion

Most of the widely used force fields fail to reproduce experimental KB integral in computer simulations. As shown in the Figure 2.2 for a $CaCl_2$ solution AMBER, CHARM27 and OPLSA force fields could not reproduce experimental KBI compared to the KBFF model. At low concentrations the deviations are particularly prominent and large deviations for the G_{cc} values implies the ion-ion interactions are not correctly captured by the force fields.

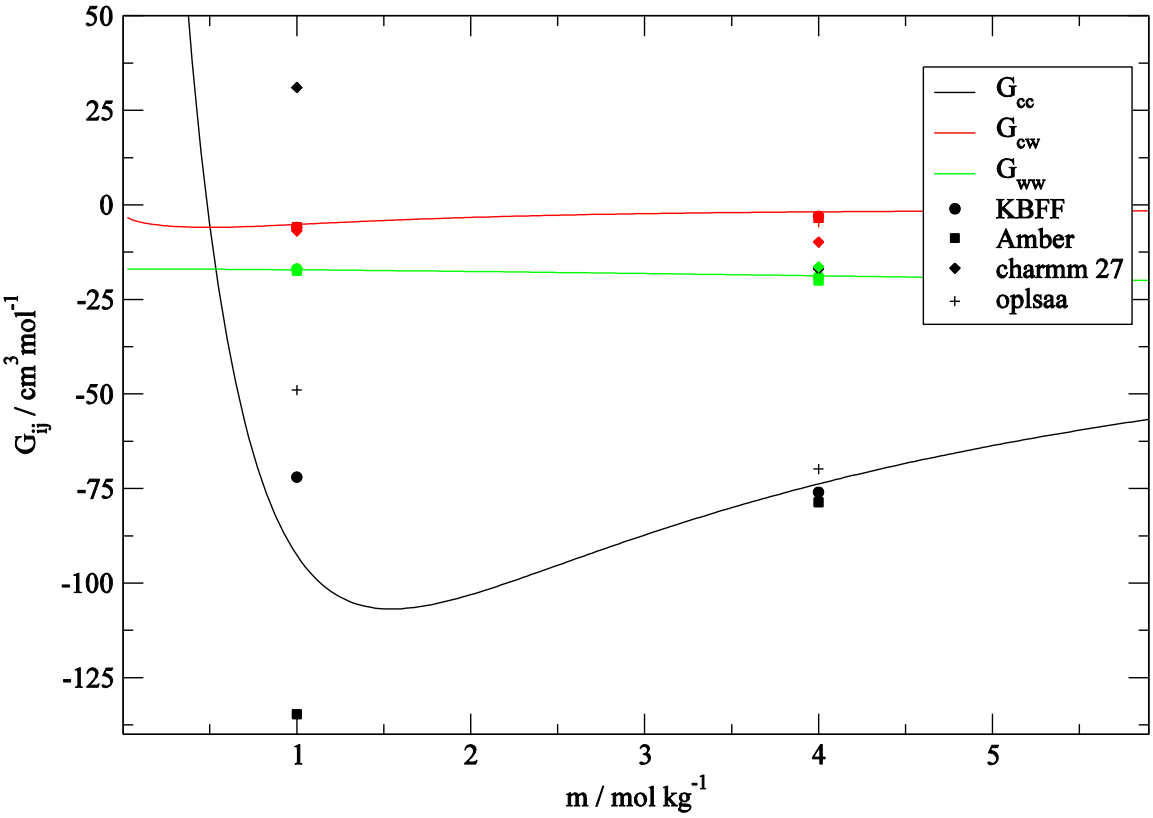

Figure 2.2. Comparison of the reproducibility of the experimental KB integrals in MD simulations using different force fields for the aqueous $CaCl_2$ salt solution. Four different force field types (KBFF, Amber, Charmm 27 and OplsAA) are investigated. Lines are obtained from a KB analysis of the experimental data and symbols are for the MD simulations.

Figure 2.3 shows the experimental excess coordination numbers for the aqueous magnesium halide solutions. This is the primary data used in the parametrization process. The data shows systematic trends between different salts which provide information about the underlying molecular distributions. The positive values of the excess coordination numbers for ion-ion interactions at lower concentrations arises from the domination of the Debye-Huckel behavior. However, the ion-ion excess coordination number goes down when moving to higher concentrations, and the changes depend on the salt.

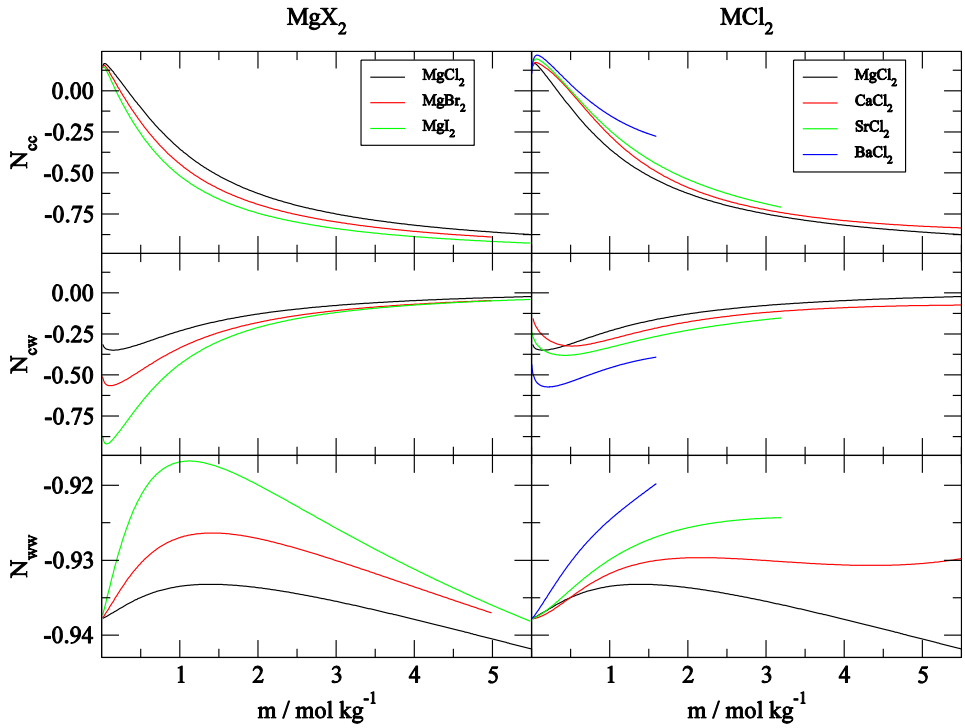

Figure 2.3. Experimentally derived excess coordination numbers for aqueous alkaline earth halide solutions as a funtion of salt molality at 298 K and 1 atm.

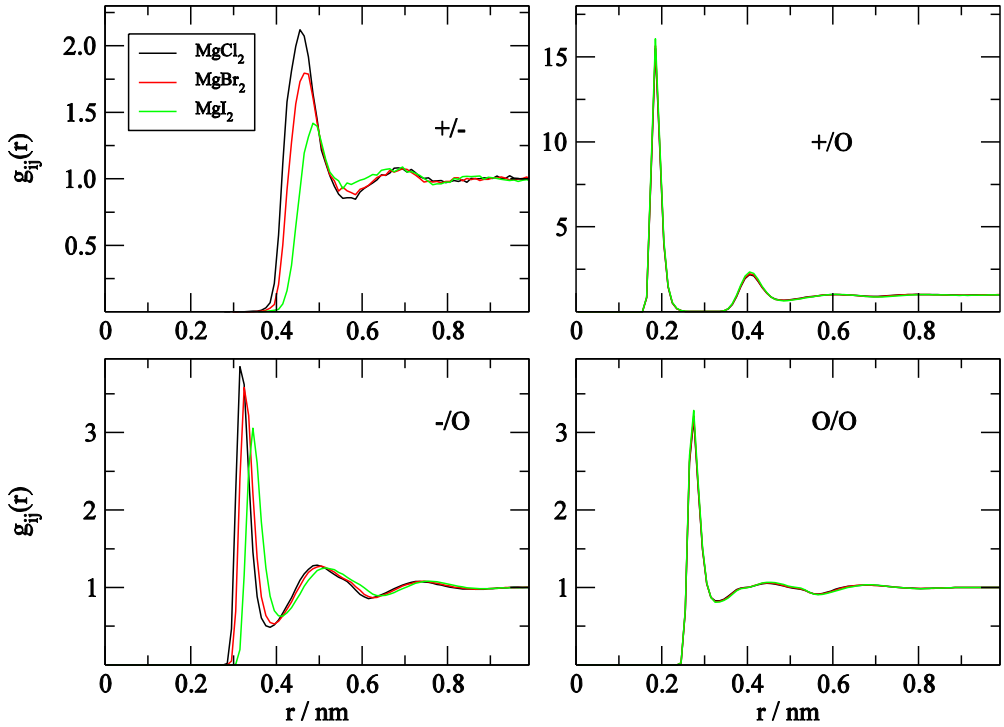

Figure 2.4. Radial distribution functions obtained from simulatiosn of 1m magnesium salt solutions. The cation, anion and the water oxygen is denoted by +, -, O, respectively.

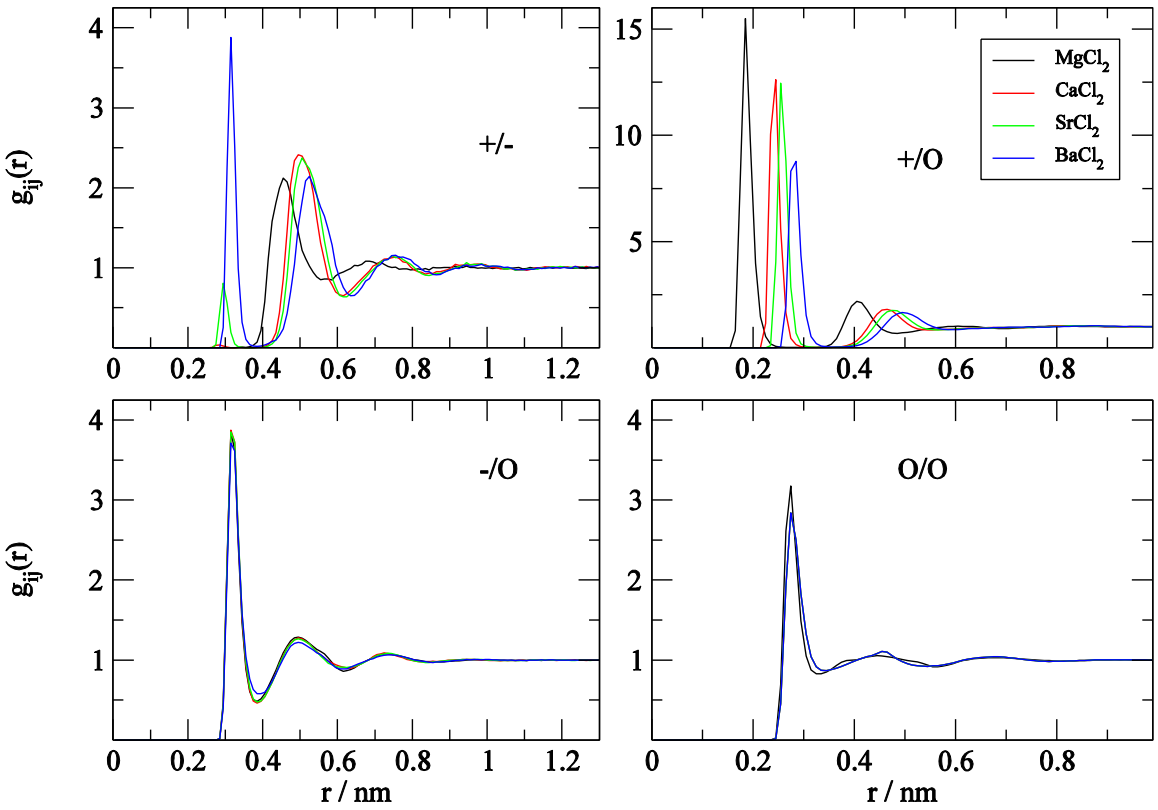

Figure 2.5. Radial distribution functions obtained from simulations of 1m chloride salt solutions. The cation, anion and the water oxygen is denoted by +, -, O, respectively.

The radial distribution functions (RDF) obtained from simulations of 1 m magnesium halide solutions using the KBFF models are displayed in Figure 2.4, with the results for 1m alkali earth chloride solutions shown in Figure 2.5. In Figure 2.4 the cation-anion rfs display a gradual increase as the size of the anion increases. The water-water interactions or the cation-water interactions are not affected by changing the anion from chloride to bromide or iodide. However, the number of water molecules interacting with the anion increases as the size of the anion increases as shown in the anion-water rdf.

In Figure 2.5 the cation-anion rdf displays several peaks. The first peak, corresponding to the contact ion pair, is not prominent for the Mg^{2+} ion. The reason for this could be the high charge density which leads to higher number of strongly bound water molecules surround the Mg^{2+}ion. For Ca^{2+} there is a small peak for the contact ions pair. For Sr^{2+} and Ba^{2+} ions there are prominent peaks for the contact ions pairs. However, all the cations display prominent peaks corresponding to solvent separated ion pairs. The cation-water interactions are affected when the size of the cation increases. The smaller cations, such as Mg^{2+} and Ca^{2+}, are surrounded by more water molecules, probably due to their high charge density. The water-water interactions are not significantly affected when the cation of the system changed.

All the RDF values are converged to unity beyond 1 nm. The first and second shell coordination numbers, the distance to the first rdf maximum (contact distance), and the first rdf minimum (first solvation shell), were calculated form the corresponding rdf distributions as shown in the supporting material, see Table S 2.1 and Table S 2.2. The hydration shell increases as the cation or anion size increases. The coordination numbers are also sensitive to the concentration of the salt solutions. The higher concentrations increase the degree of ion pair formation. However, no strong aggregation or crystallization was observed in the simulation.

Several X-ray diffraction studies have shown that the magnesium to water oxygen distance is about 0.209 nm and the average coordination number is six.[47, 58-60] In the present study, the magnesium to water contact distance is 0.185 nm and the coordination number is 5.9. The coordination number is consistent with many other MD simulations reported in the litrature.[61-62] Furthermore, the experimental ion to water oxygen distance for the second hydration shell is reported as 0.410 nm, compared to 0.445 nm in the present study.[59] The contact ion pair is not visible in Figure 2.5 for magnesium and chloride ions.

The calcium ion to water distance is 0.245 nm and the coordination number is 7.7 in the present study. The neutron x-ray diffraction experimental studies have reported the calcium ion to water oxygen distance is 0.242 nm. However, the coordination number can vary between 7-10.[47, 59-60, 63-64] A X-ray absorption fine structure (XAFS) spectroscopy experimental study of a 6 m calcium chloride solution found that the coordination number is around 7.2 and the cation to water contact distance is about 0.244 nm.[65] Furthermore, this study claims that there is no evidence for the formation of significant number of the ion pairs at high concentrations, and that solvent shared ion pairs, rather than contact ion pairs, are responsible for the unusual thermodynamic behavior of $CaCl_2$ solutions. Another experimental study of $CaCl_2$ solutions reported that 0.26 as the coordination number of contact ion pair at distance of 0.271 nm and the coordination number as 3.4 for solvent separated ion pair at distance of 0.498 nm.[66] The values reported for the $CaCl_2$ ion pair and solvent separated ion pair in the supporting material, see Table S 2.2, suggest a very small peak at 0.284 nm for the ion pair and a prominent peak at 0.505 nm for the solvent separated ion pair, respectively. A coordination number of 7.3 for the Ca^{2+} ions was reported in a MD simulation which uses a polarizable potential.[61-62]. Furthermore, the experimental second hydration shell distance for the calcium ion was reported to be 0.450-0.460 nm[59, 67], compared to 0.466 nm in this study.

A molecular dynamics simulation to study the solvation of Mg^{2+} and Ca^{2+} ions with polarizable atomic multipole potential reported that divalent ions perturb the structure and the dipole moment of the first solvation shell water molecules.[62] This study also claims that the water structure of the first solvation shell is more structured compare to that surrounding monovalent ions. This study reports average coordination numbers for Mg^{2+} and Ca^{2+} ions of 6 and 7.3,

respectively. The present MD study supports this claim, as reflected by the sharp first peak and the prominent separation between the first and second peaks in the RDFs.

In the current MD simulations the strontium ion to water contact distance is 0.256 nm and the coordination numbers are close to 6 for 1 m and 8 for 2 m. The X-ray diffraction, neutron scattering and other spectroscopic studies have reported the strontium ion to water oxygen distance in the range of 0.256-0269 nm, and a coordination number in the range of 6.7-10.3.[47, 59-60, 68-70] An experimental and MD simulation study performed in 2016 using X-ray adsorption near edge structure (XANES) reported that the strontium ion to water oxygen distance is 0.26 nm and the coordination number is 8.[71] The study also reports that the hydration shells of the Sr^{2+} ion seem to have a flexible nature with a fast ligand exchange rate between the first and second hydration shells. The second hydration shell distance of the strontium ion and the water molecule was reported to be 0.494 nm.[59]

For the barium ion the water distance is reported as 0.29 nm with a coordination number of 9.5 in several experimental studies.[47, 59-60] In the present study, the barium ion to water contact distance is 0.285 and the coordination number is close to 8.

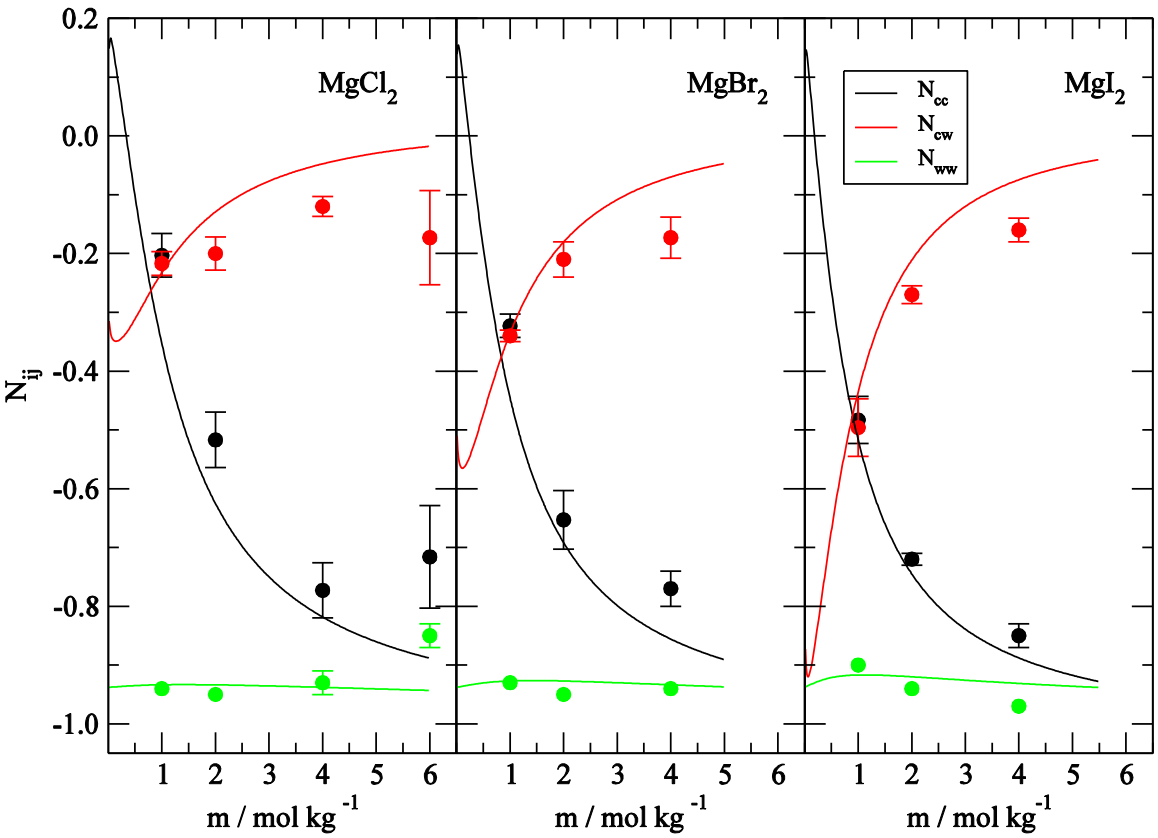

Figure 2.6. Excess coordination number as a function of magnesium salt concentraion. The lines are obtained from the experimental KBI analysis. The symobols are from the simulations performed by using the KBFF models.

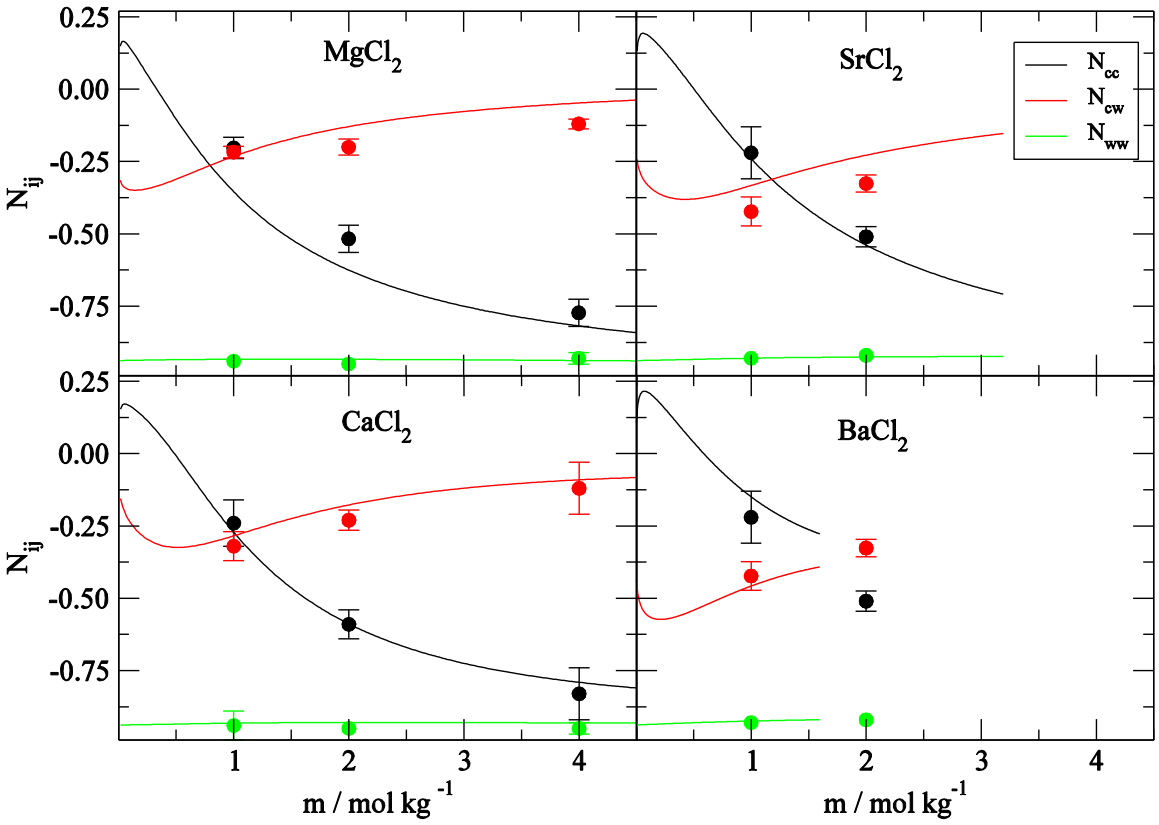

Figure 2.7. Excess coordination number as a function of chloride salt concentraion. The lines are obtained from experimental KBI analysis. The symobols are from the simulations performed by using the KBFF models.

The simulated and experimental excess coordination numbers N_{ij} are displayed in Figure 2.6 for magnesium halide solutions, and in Figure 2.7 for the alkaline earth chloride solutions. The excess coordination number N_{ij} values for calcium halides, strontium halides and barium halides are displayed in the supporting materials, see Figure S 2.1, Figure S 2.2 and Figure S 2.3, respectively. The KBFF force fields quantitatively reproduce the experimental data. However, at higher concentrations the simulated values for the $MgCl_2$ solution starts to deviate from the experimental numbers. The correct trends are reproduced in all the alkaline earth halide salt solutions. As displayed in Figure 2.6, the ion-ion interactions are less affected compare to ion-

water interactions when the halide ion is changed for a given alkaline earth metal ion. This is a clear indication that ion-water interactions determine the solvation behavior in solution, as discussed in our earlier publication of alkali halide solutions.[10]

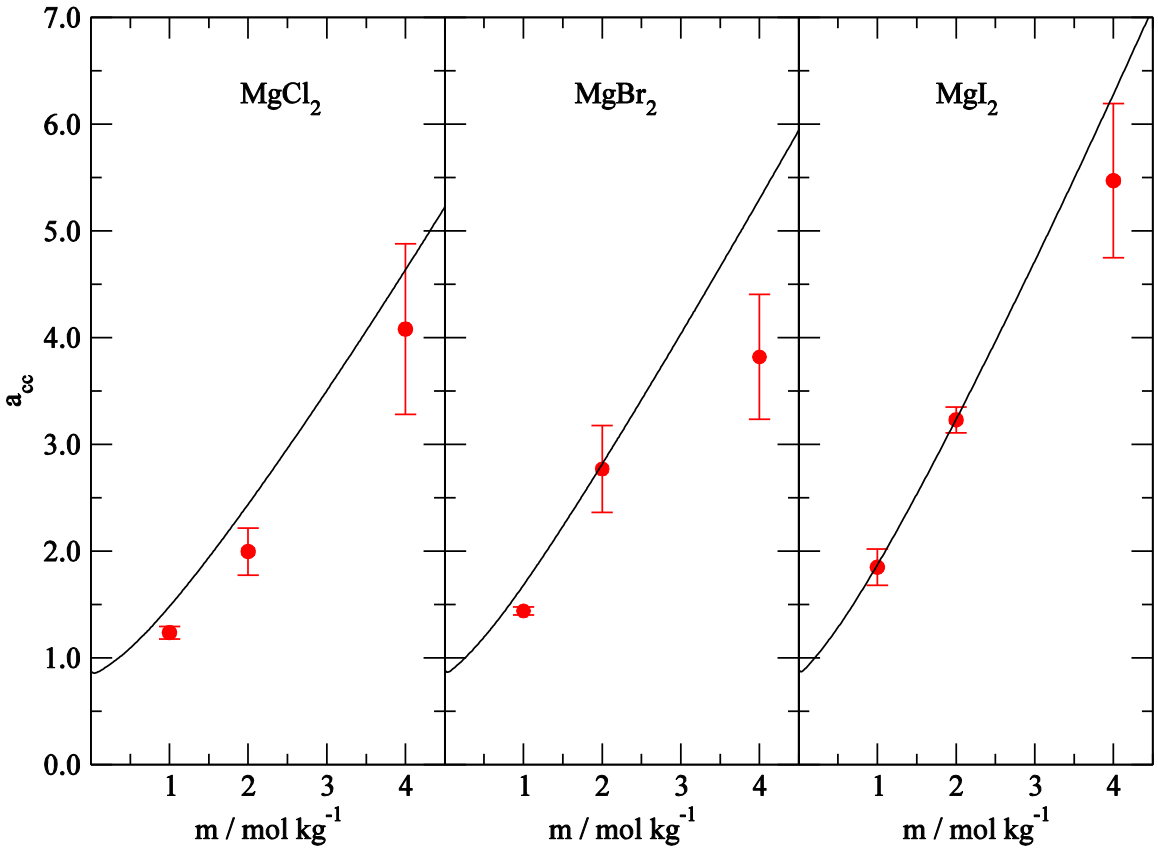

Figure 2.8. Activity derivatives for magnesium salts as a function of the salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

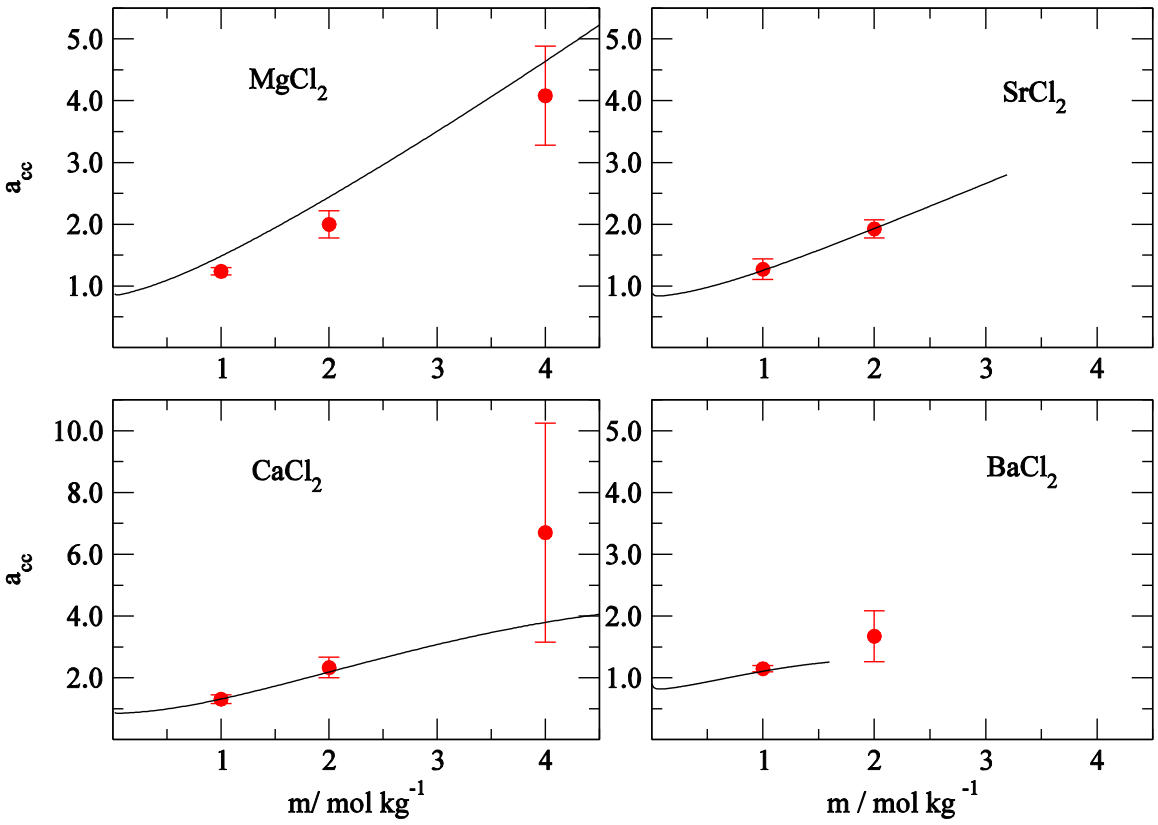

Figure 2.9. Activity derivatives for chloride salts as a salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

A comparison of the simulated and experimental activity coefficient derivatives is displayed in Figure 2.8 and Figure 2.9. The activity coefficient derivatives values for calcium halides, strontium halides, and barium halides are displayed in the supporting materials, see Figure S 2.4, Figure S 2.5 and Figure S 2.6, respectively. The KBFF models can reproduce the correct experimental activity derivatives. However, at higher concentrations the simulated activity derivatives tend to deviate slightly from the experimental values. It should be noted that many force fields for alkaline earth ions fail to reproduce the correct thermodynamic properties.[35]

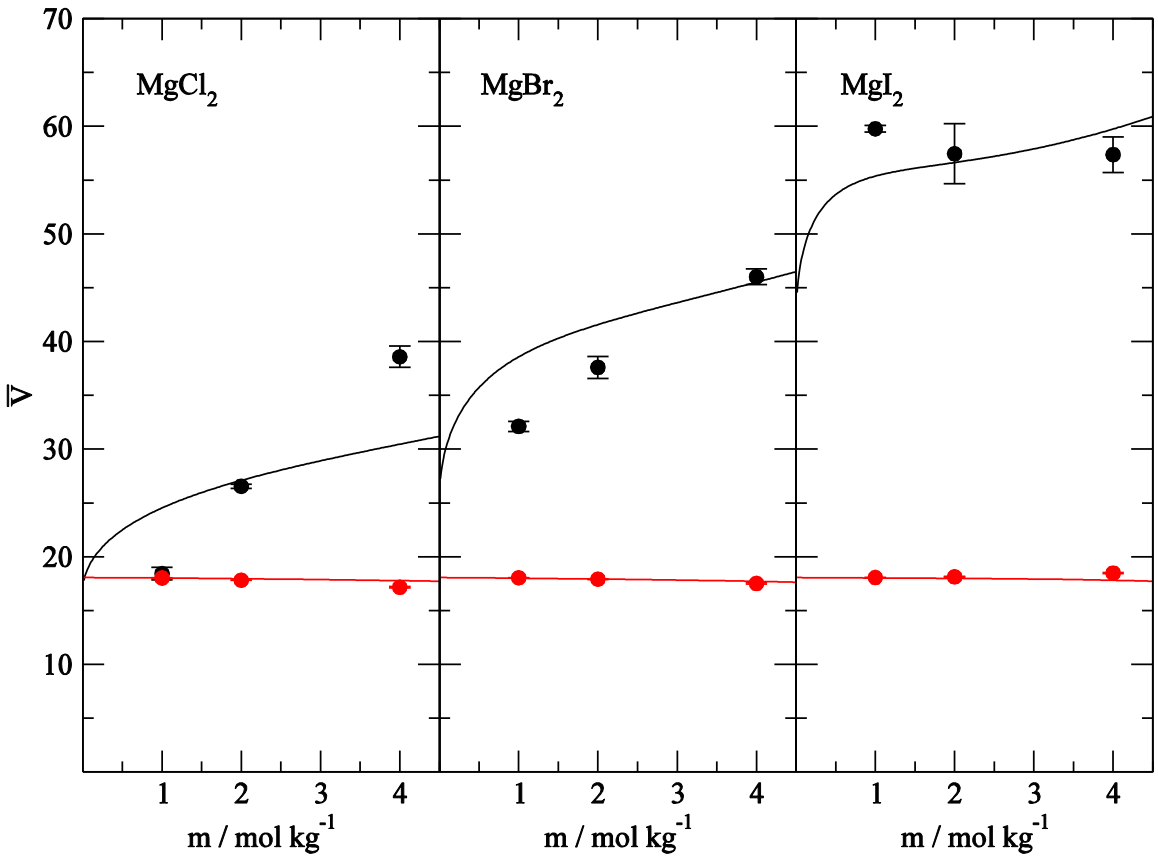

Figure 2.10. Partial molar volumes (cm^3 mol^{-1}) of the salt ions for the magnesium salts as a function of the salt molality. Partial molar volumes of ions displayed in black and partial molar volumes of the water displayed in red color. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

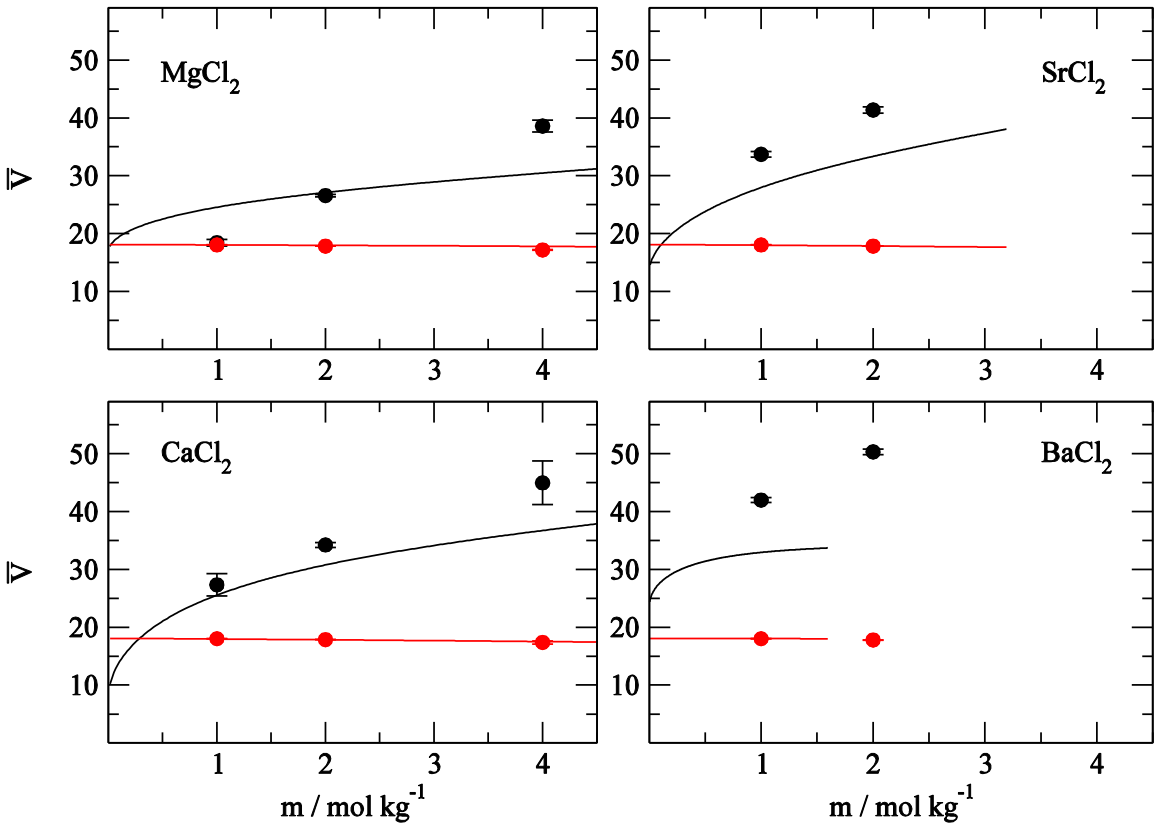

Figure 2.11. Partial molar volumes ($cm^3\ mol^{-1}$) of the salt ions for the chloride salts as a function of the salt molality. Partial molar volumes of ions displayed in black and partial molar volumes of the water displayed in red color. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

Figure 2.10 and Figure 2.11 display the simulated and the experimental partial molar volumes of both the ions and water as a function of concentration for the magnesium salts and for the chloride salts. The simulated and the experimental partial molar volumes for calcium halides, strontium halides and barium halides are displayed in the supporting materials, see Figure S 2.7, Figure S 2.8 and Figure S 2.9, respectively. The simulated values follow the experimental trends. An increase in salt concentration results in an increase of the partial molar volume of the ions, with a slight decrease in the partial molar volume of the water. However, the simulated partial molar volumes display some deviations at high concentrations.

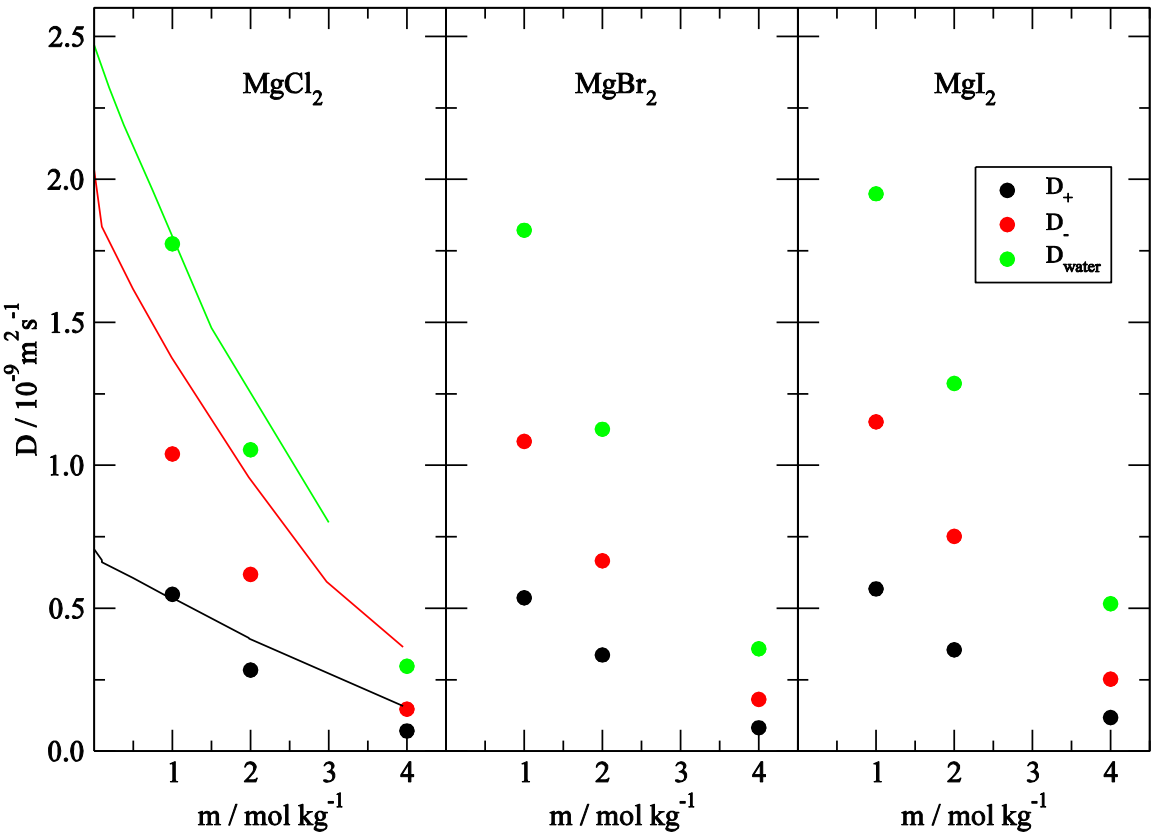

Figure 2.12. Diffusion constants for magnesium salts as a function of the salt molality. The lines represent the experimental diffusion constant data[72], the symbols correspond to the MD simulations.

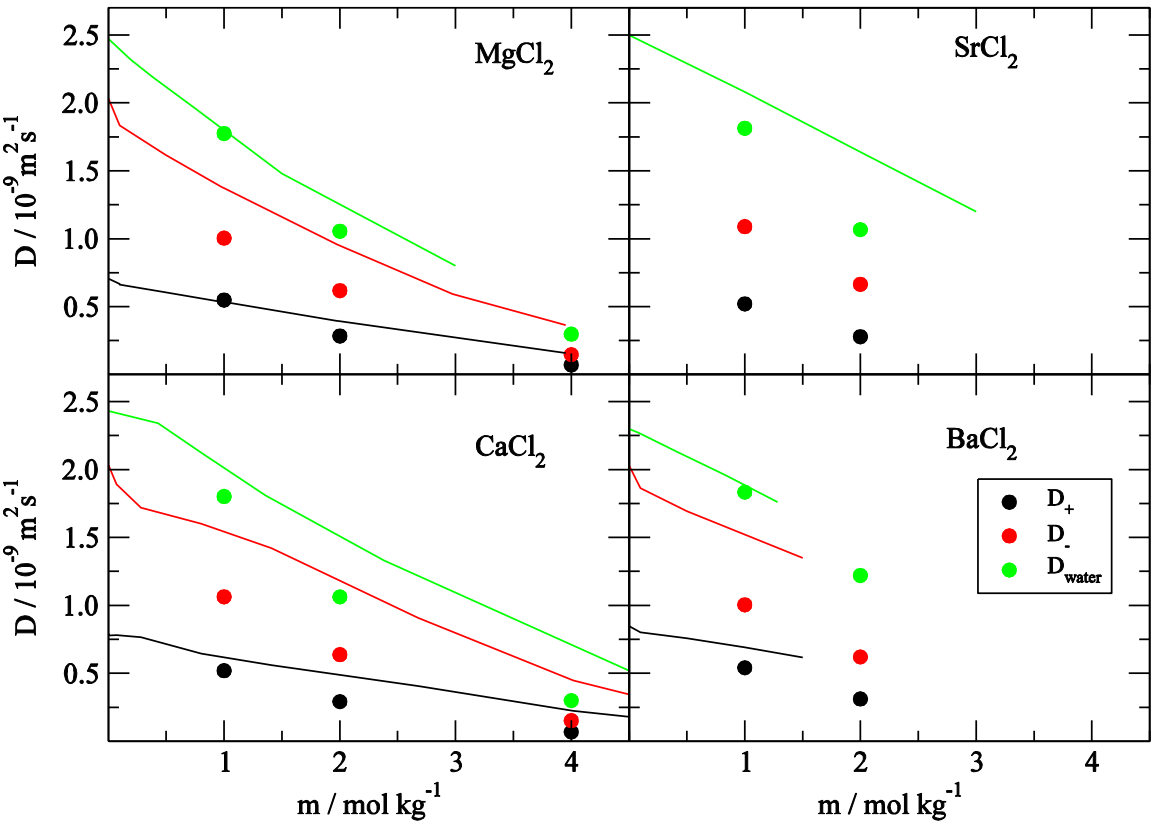

Figure 2.13. Diffusion constants for chloride salts as a function of the salt molality. The lines represent the experimental diffusion constant data[72], the symbols correspond to the MD simulations.

The activity derivatives were used in the parameterization process of the KBFF to reproduce experimental KB integrals. Hence, some properties that are not used in the parametrization process such as the self-diffusion coefficient and dielectric constant have also been calculated in this study. The self-diffusion coefficients are displayed in Figure 2.12 and Figure 2.13 for the magnesium salts and for the chloride salts, and were calculated by using the mean square fluctuation approach as a function of salt concentration. The simulated and the experimental self-diffusion coefficients[72] for calcium halides, strontium halides and barium halides are displayed in the supporting materials, see Figure S 2.10, Figure S 2.11 and Figure S 2.12, respectively. An increase in the salt concentration lowers the self-diffusion coefficient of the

cation, anion and water molecules. As reported in the experimental studies, the self-diffusion coefficients of alkaline earth cations are independent of the molar mass.[73] Furthermore, the self-diffusion coefficients of the halide ions also do not display any correlation with the size of the ion. These trends appear to be correctly reproduced in the MD simulation. In the present study, an overestimation of the diffusion constant for the halide ions were observed in the simulations. The parameters used in this study for halide ions are taken from a previous work for alkaline halides which also displayed the same overestimation trend.[10] Therefore, the parameterization of halide ions to reproduce reasonable diffusion coefficients for both alkaline halides and alkaline earth halides appears to be a challenging task. The highest diffusion is displayed by the water molecules in all salt solutions, which is consistent with the experimental data.

A MD study of ions with a polarizable force field reported the self-diffusion coefficient of the Mg^{2+} and Ca^{2+} ions as 0.3 x 10^{-5} cm^2s^{-1} and 0.8 x 10^{-5} cm^2s^{-1} at infinite dilution, respectively.[62] These values are close to the results of this simulation study. Furthermore, other non-polarizable force field MD studies have reported lower self-diffusion coefficients at infinite dilutions for Mg^{2+} and Ca^{2+} ions.[74] However, experimental values of Mg^{2+} and Ca^{2+} ions are 0.71 x 10^{-5} cm^2s^{-1} and 0.79 x 10^{-5} cm^2s^{-1} , respectively.[72]

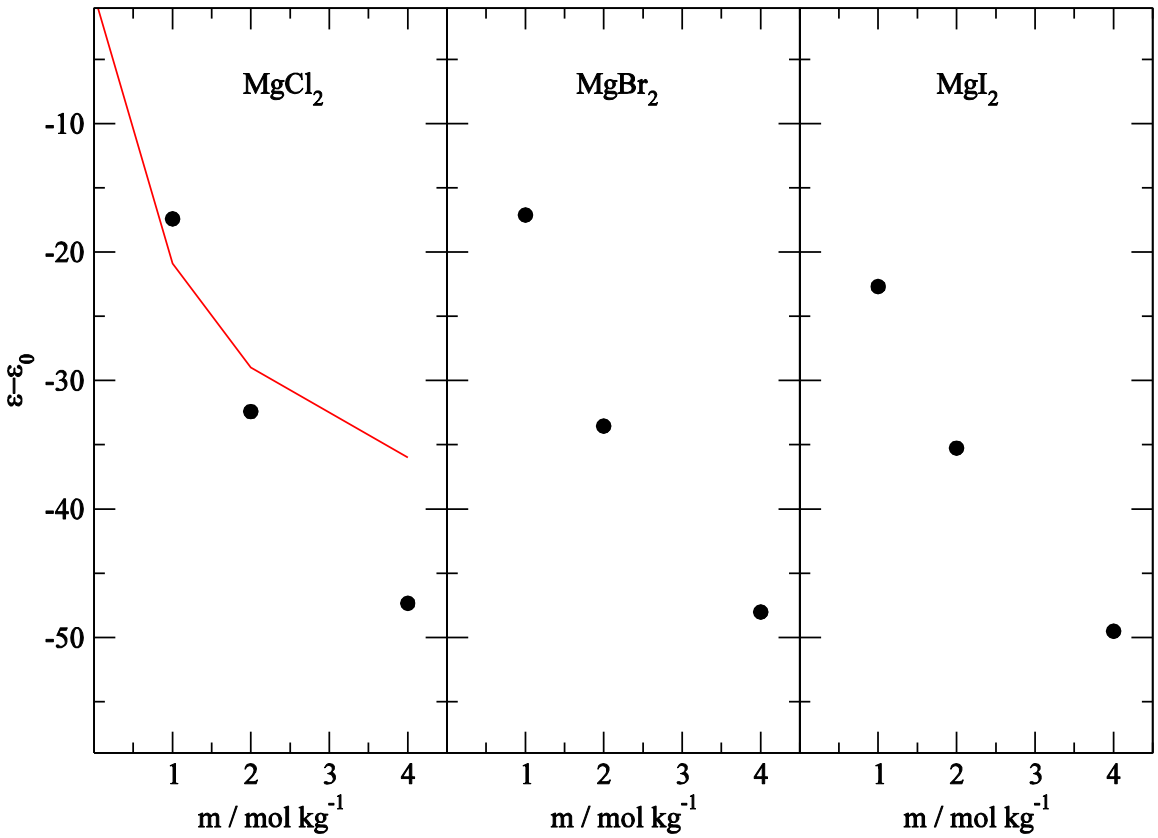

Figure 2.14. Dielectric decrements ($\varepsilon - \varepsilon_0$) for the magnesium salts as a function of the salt molality (mol kg^{-1}). Lines are for the experimental data[75] and symbols are for the MD simulations.

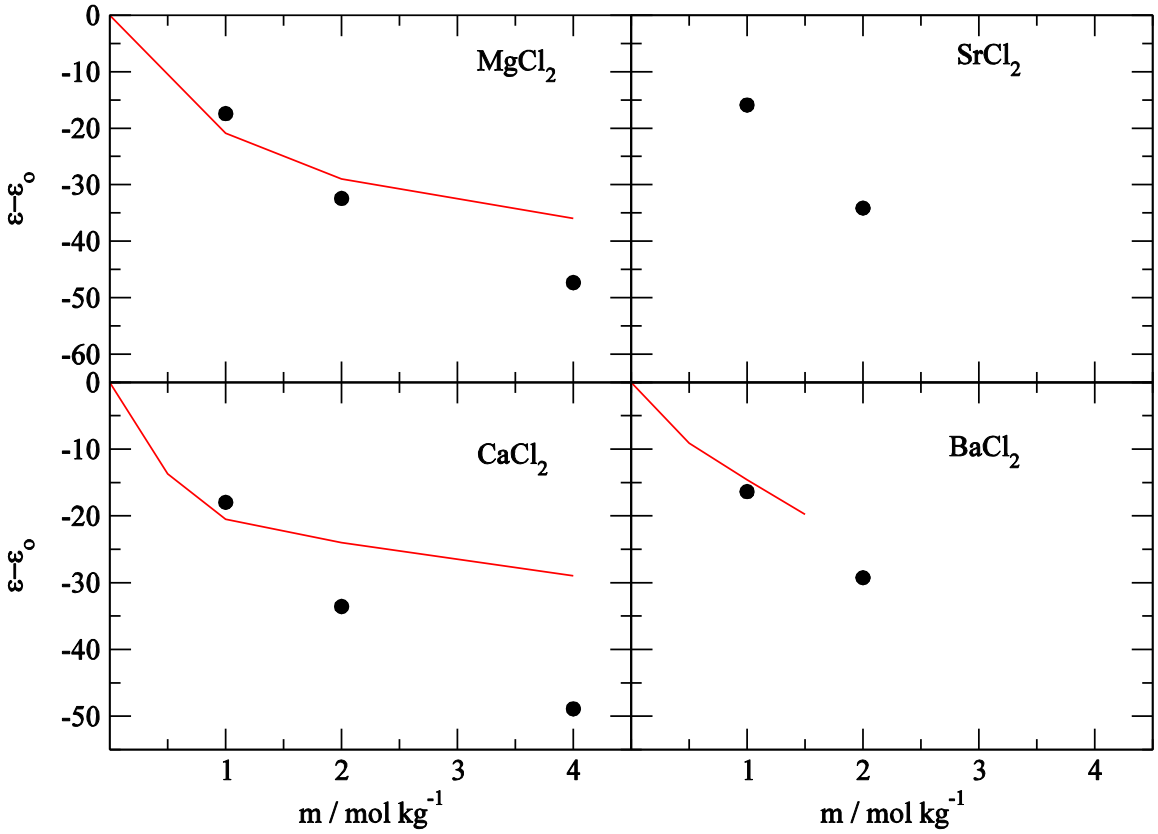

Figure 2.15. Dielectric decrements ($\varepsilon - \varepsilon_0$) for the chloride salts as a function of the salt molality. Lines are for the experimental data[75] and symbols are for the MD simulations.

Figure 2.14 and Figure 2.15 display the dielectric decrements for the magnesium salts and for the chloride salts with respect to an increase in salt concentration. The simulated and the experimental dielectric decrement[75] for calcium halides, strontium halides and barium halides are displayed in the supporting materials, see Figure S 2.13, Figure S 2.14 and Figure S 2.15, respectively. The dielectric constants were calculated using the total dipole moment fluctuations. The SPC/E water model dielectric constant is 63, and it is significantly lower than the experimental dielectric constant of pure water which is 78.[76] Therefore, it is more convenient to use the dielectric decrement instead of the dielectric constant to compare the experimental and simulated results.

Increasing the solution molality results in a decrease in the dielectric decrement because the more ions there are in the medium the larger the effects on the water dipole moment. However, at higher concentrations the simulated values appear to overestimate the experimental dielectric decrement values.

The excess enthalpies of mixing were also calculated in this study. These are sensitive measures of the mixing thermodynamics. However, we could not find the corresponding experimental excess enthalpies of mixing for alkali earth metal halides at finite concentrations. Therefore, the MD simulation results for excess enthalpies were reported in the supporting information (Figure S 2.16, Figure S 2.17, Figure S 2.18, Figure S 2.19 and Figure S 2.20 for chloride salts, magnesium salts, calcium salts, strontium salts and barium salts, respectively.). The alkaline earth halide salts display favorable excess enthalpy of mixing in the MD simulations. Finally, we have also simulated the salt crystal structures. Orthogonal crystal structures were used for $MgCl_2$, $MgBr_2$, MgI_2 and CaI_2 despite the prominent non orthogonal (90º, 90º, 120º) crystal forms being observed in the nature. Unfortunately, the non orthogonal forms were unstable during our simulations. The crystal simulation potential energies per molecule, densities, and unit cell dimensions are compared to the experimental values in the supporting materials, see Table S 2.3 and Table S 2.4. A stable crystal in MD simulations is crucial to perfume salt crystal solubility studies. It appears that the unstable salt crystal forms with KBFF parameters would not be a good choice for crystal or solubility studies.

2.5. Conclusions

In this study we have developed a series of models for aqueous alkali earth halide salt solutions by attempting to reproduce the experimentally derived Kirkwood-Buff integrals using molecular dynamics simulations. It has been clearly demonstrated that the parameters developed for chloride salts are transferable to bromides and iodides. The models perform very well for low salt concentrations. However, there are some deviations observed for high salt concentrations of some salt solutions. Furthermore, some of the salts parameters are clearly not suitable for salt crystal simulations, as unstable crystal structures are obtained.

Considering the importance of the aqueous alkaline earth metal halides solutions, it is surprising that there is a scarcity of experimental data in the literature - such as enthalpy of mixing, self-diffusion constants, and dielectric constant, etc, especially at finite concentrations. Also, it is noticeable that most of the experimental studies do not report statistical errors, which would facilitate a comparison with the simulation results.

This KBFF models should be viewed as providing a reasonable balance between solute-solute, solute-solvent, and solvent-solvent interactions, as inferred by their resulting distributions and KBIs. The models provide the correct salt activities, and also reasonably reproduce experimental densities, diffusion constants and dielectric decrements. Therefore, the KBFF models are suitable for studies of solute activities and co-solvent interactions with biomolecules. Breaking the standard combination rules when determining the cation-water interactions appears to be a reasonable way to include implicit polarization effects in non-polarizable force fields for cations with high charge densities.

2.6. Supporting Information

Table S 2.1. First shell coordination numbers (n_1), second shell coordination numbers (n_2) as a function of the magnesium salt molality. The $R_{1,max}$ and $R_{1,min}$ are the first shell maximum and minimum distances and the $R_{2,max}$ and $R_{2,min}$ are the second shell maximum and minimum distances from the corresponding radial distribution functions.

		m	+/-	+/O	-/O	O/O
$MgCl_2$	$R_{1,max}$			0.185	0.315	0.275
	$R_{1,min}$			0.261	0.385	0.325
	n_1	1		5.94	7.42	4.37
		2		5.93	7.46	4.37
		4		5.9	7.45	4.31
		6		5.78	7.31	4.2
	$R_{2,max}$		0.451	0.406	0.494	0.445
	$R_{2,min}$		0.55	0.484	0.615	0.564
	n_2	1	1.14	18.07	32.76	24.48
		2	2.18	16.93	32.72	23.78
		4	4.52	14.25	31.8	21.85
		6	6.84	11.65	30.29	19.78
$MgBr_2$	$R_{1,max}$			0.185	0.325	0.275
	$R_{1,min}$			0.28	0.395	0.325
	n_1	1		5.95	7.54	4.34
		2		5.94	7.59	4.32
		4		5.9	7.56	4.23
	$R_{2,max}$		0.466	0.41	0.51	0.446
	$R_{2,min}$		0.575	0.48	0.625	0.565
	n_2	1	1.11	18.01	33.63	24.2
		2	2.22	16.84	33.34	23.25
		4	4.71	13.88	32.02	20.94
MgI_2	$R_{1,max}$			0.185	0.345	0.275
	$R_{1,min}$			0.254	0.406	0.325
	n_1	1		5.93	7.46	4..29
		2		5.92	7.47	4.24
		4		5.84	7.38	4.21

	$R_{2,max}$		0.485	0.406	0.515	0.455
	$R_{2,min}$		0.545	0.485	0.645	0.566
	n_2	1	0.67	18.14	35.48	23.67
		2	1.44	16.77	34.69	22.3
		4	3.37	13.42	32.51	19.34

Table S 2.2. First shell coordination numbers (n_1), second shell coordination numbers (n_2) as a function of the chloride salt molality. The $R_{1,max}$ and $R_{1,min}$ are the first shell maximum and minimum distances and the $R_{2,max}$ and $R_{2,min}$ are the second shell maximum and minimum distances from the corresponding radial distribution functions.

		m	+/-	+/O	-/O	O/O
MgCl2	$R_{1,max}$			0.185	0.315	0.275
	$R_{1,min}$			0.261	0.385	0.325
	n_1	1		5.94	7.42	4.37
		2		5.93	7.46	4.37
		4		5.9	7.45	4.31
		6		5.78	7.31	4.2
	$R_{2,max}$		0.451	0.406	0.494	0.445
	$R_{2,min}$		0.55	0.484	0.615	0.564
	n_2	1	1.14	18.07	32.76	24.48
		2	2.18	16.93	32.72	23.78
		4	4.52	14.25	31.8	21.85
		6	6.84	11.65	30.29	19.78
$CaCl_2$	$R_{1,max}$		0.284	0.245	0.315	0.275
	$R_{1,min}$		0.325	0.306	0.385	0.344
	n_1	1	0	7.78	7.48	5.18
		2	0	7.76	7.58	5.18
		4	0.05	7.59	7.68	4.97
	$R_{2,max}$		0.505	0.466	0.495	0.455
	$R_{2,min}$		0.615	0.545	0.625	0.544
	n_2	1	1.58	24.33	33.71	21.79
		2	3.04	22.82	33.33	20.89
		4	6.1	19.52	31.77	18.86

$SrCl_2$	$R_{1, max}$		0.295	0.256	0.315	0.275
	$R_{1, min}$		0.375	0.329	0.385	0.355
	n_1	1	0.04	8.02	7.49	5.61
		2	0.09	7.93	7.62	5.58
	$R_{2, max}$		0.506	0.475	0.495	0.484
	$R_{2, min}$		0.615	0.565	0.616	0.545
	n_2	1	1.59	24.71	31.89	21.66
		2	3.04	24.85	31.23	20.68
$BaCl_2$	$R_{1, max}$		0.315	0.285	0.315	0.275
	$R_{1, min}$		0.385	0.366	0.385	0.365
	n_1	1	0.22	8.48	7.48	6.07
		2	0.46	7.99	7.53	5.99
	$R_{2, max}$		0.523	0.496	0.495	0.445
	$R_{2, min}$		0.635	0.595	0.61	0.565
	n_2	1	1.8	29.74	31.45	23.86
		2	3.38	28.05	29.73	22.71

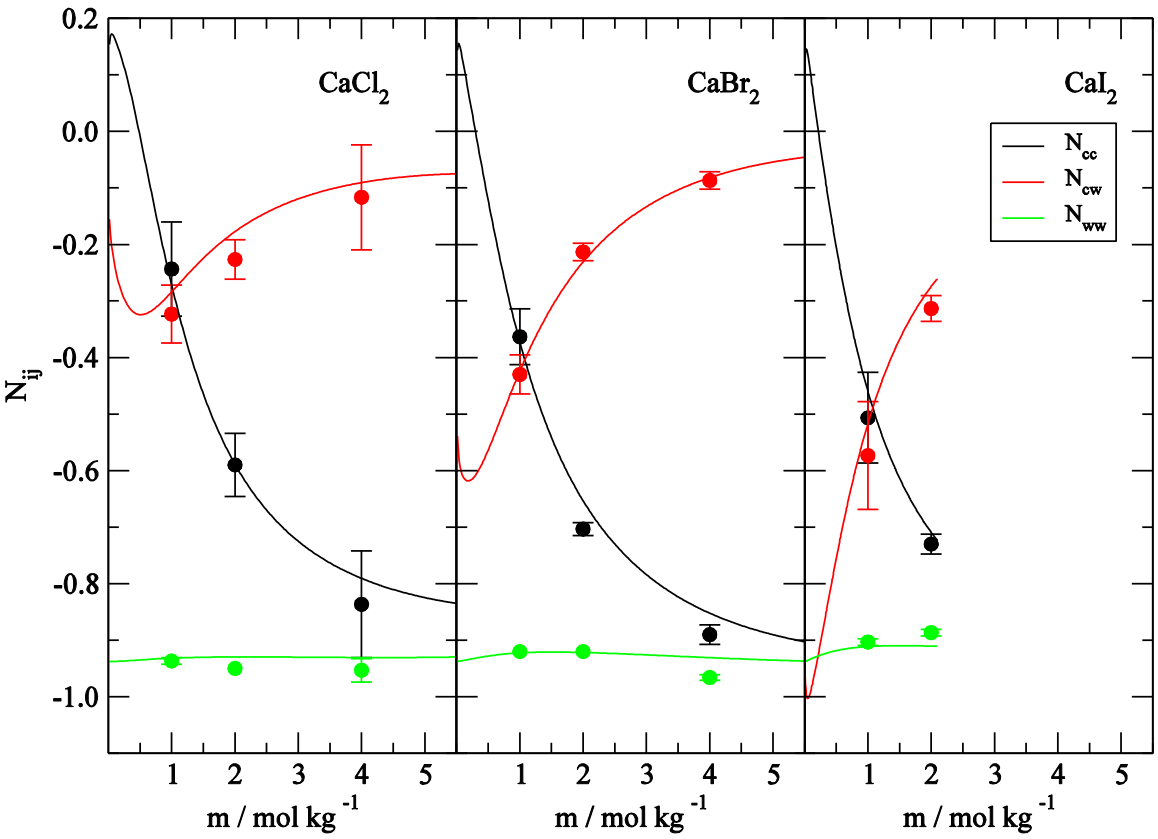

Figure S 2.1. Excess coordination number as a function of calcium salt concentraion. The lines are obtained from the experimental KBI analysis. The symobols are from the simulations performed by using the KBFF models.

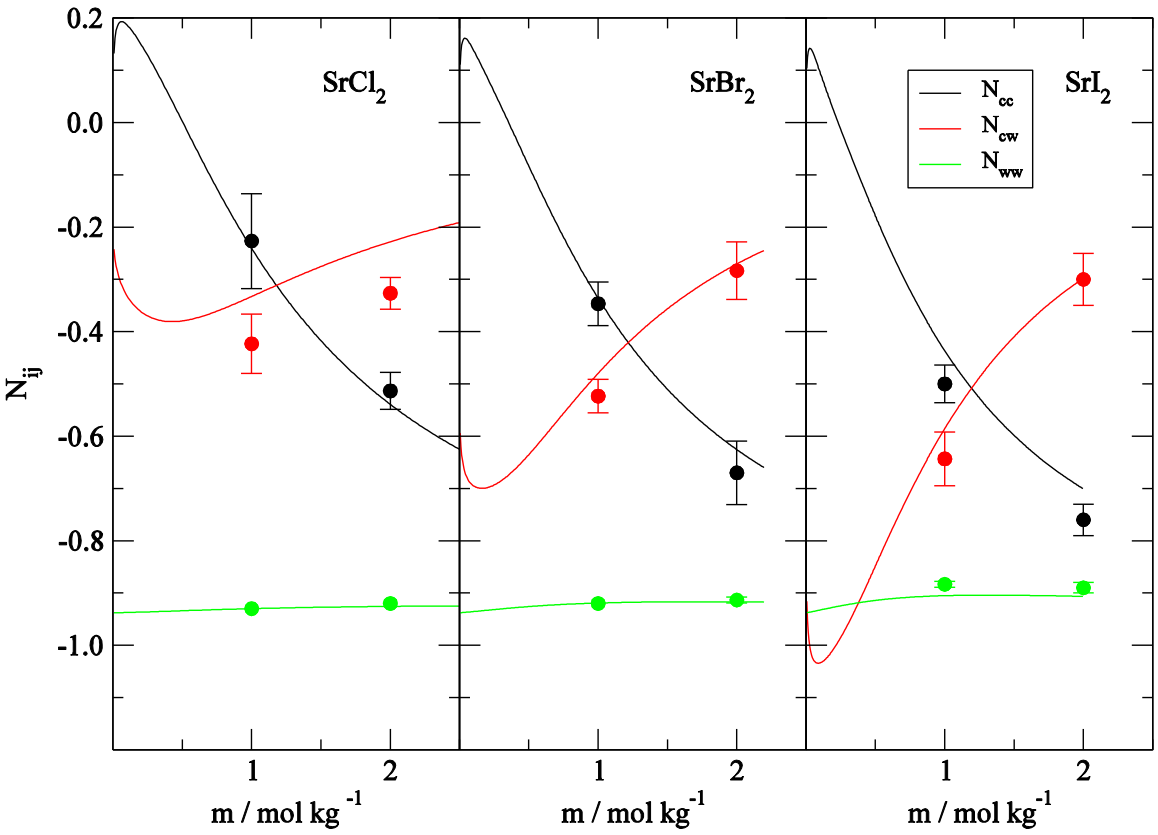

Figure S 2.2. Excess coordination number as a function of strontium salt concentraion. The lines are obtained from the experimental KBI analysis. The symobols are from the simulations performed by using the KBFF models.

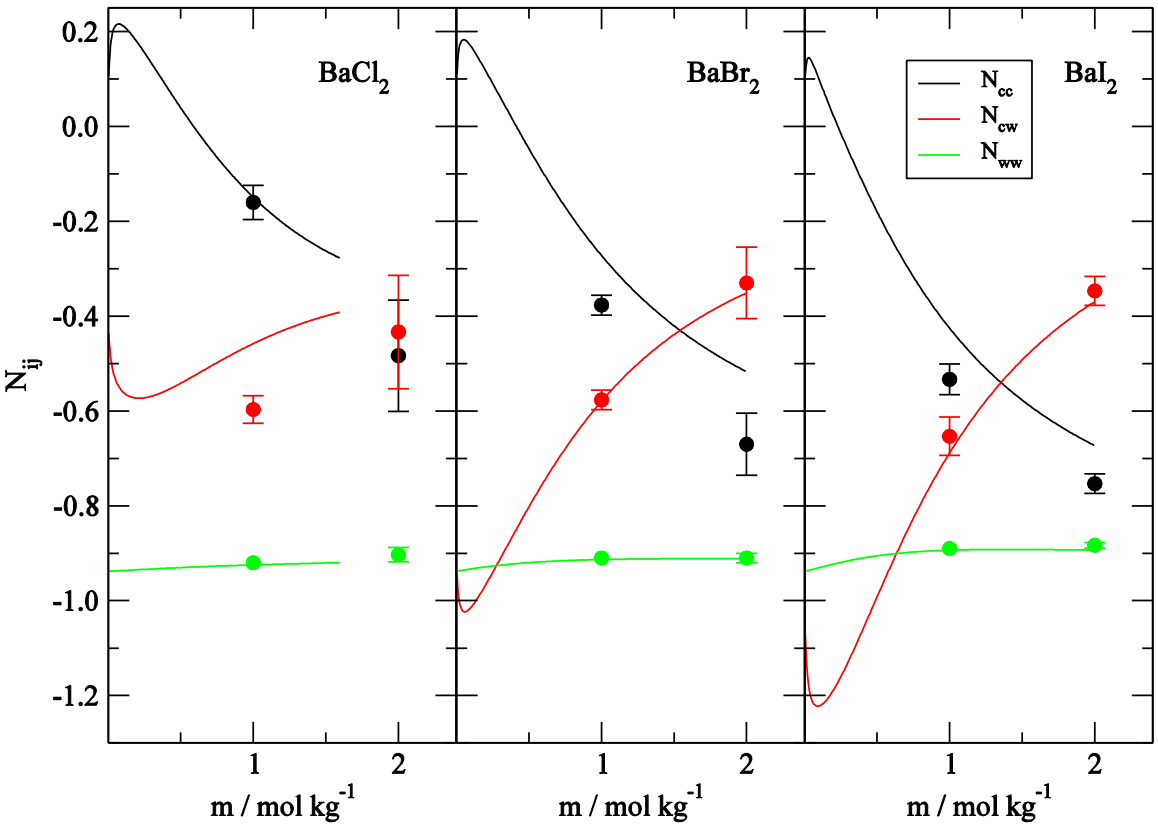

Figure S 2.3. Excess coordination number as a function of barium salt concentraion. The lines are obtained from the experimental KBI analysis. The symobols are from the simulations performed by using the KBFF models.

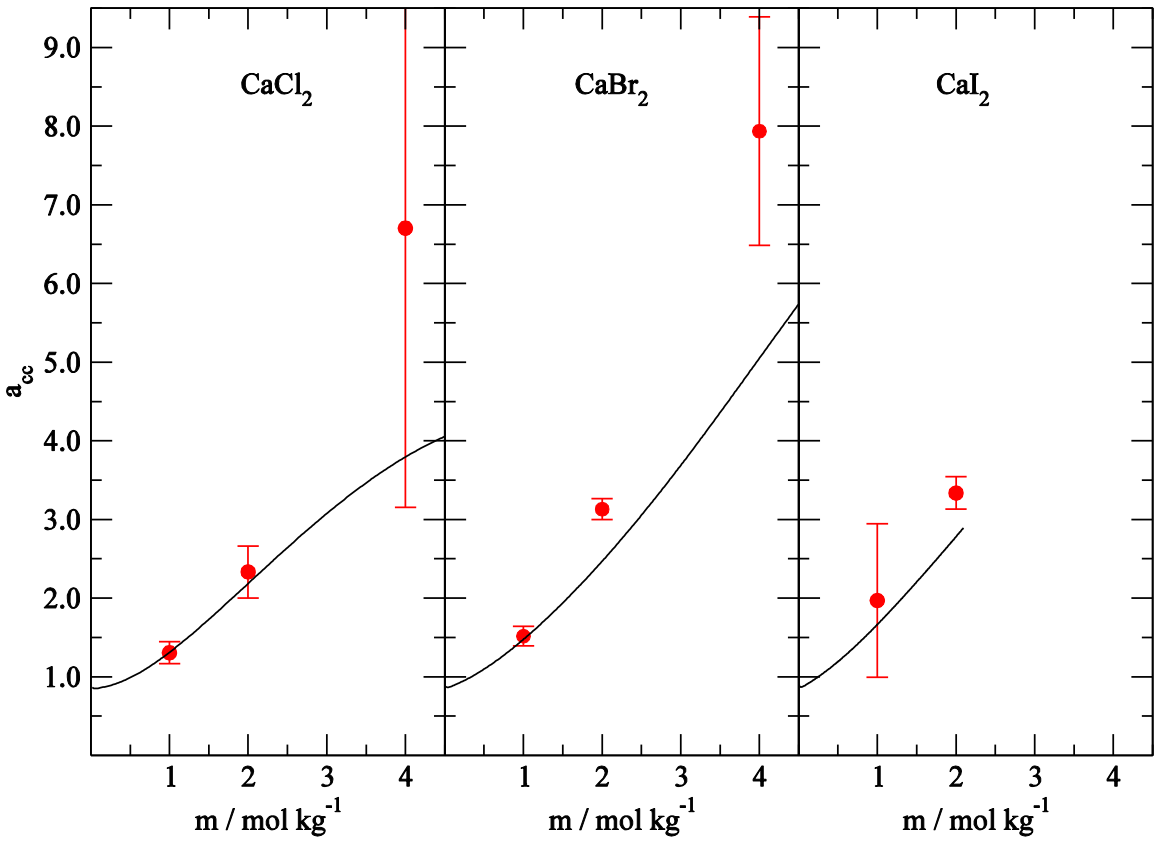

Figure S 2.4. Activity derivatives for calcium saits as a function of the salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

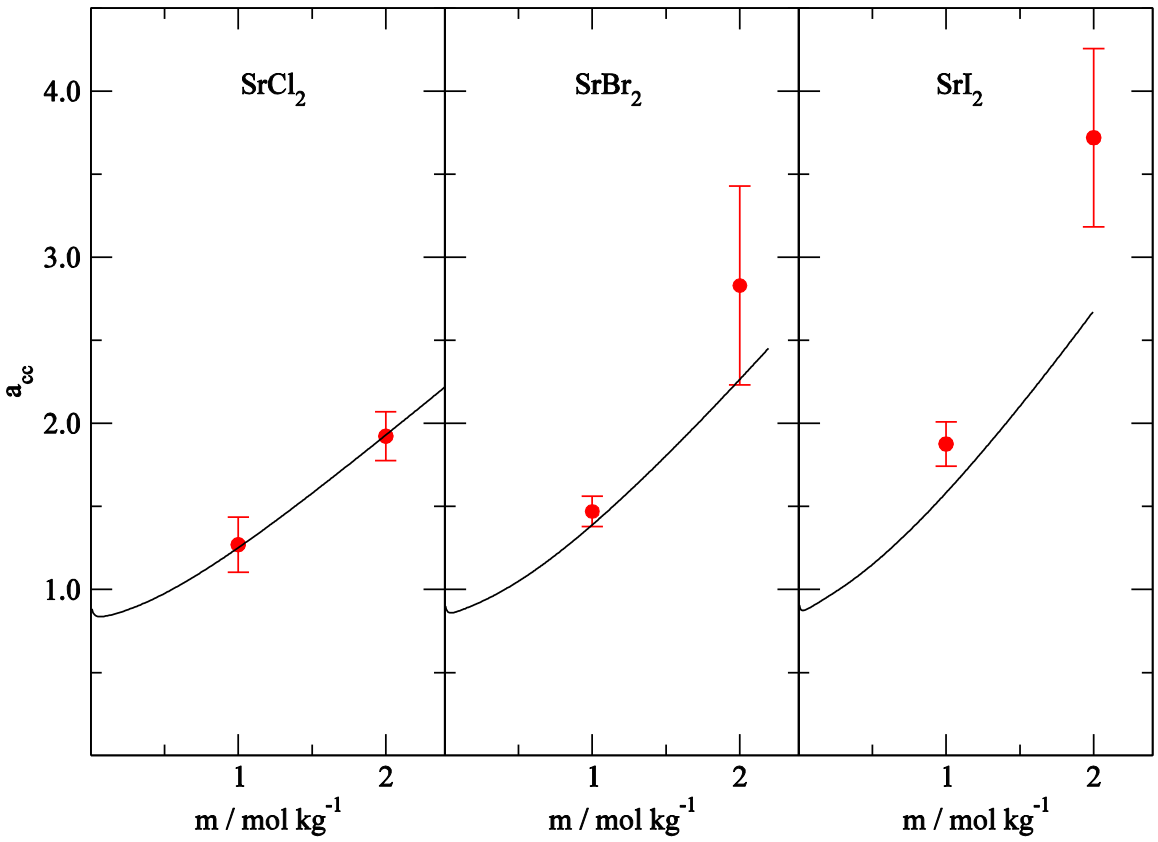

Figure S 2.5. Activity derivatives for strontium salts as a function of the salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

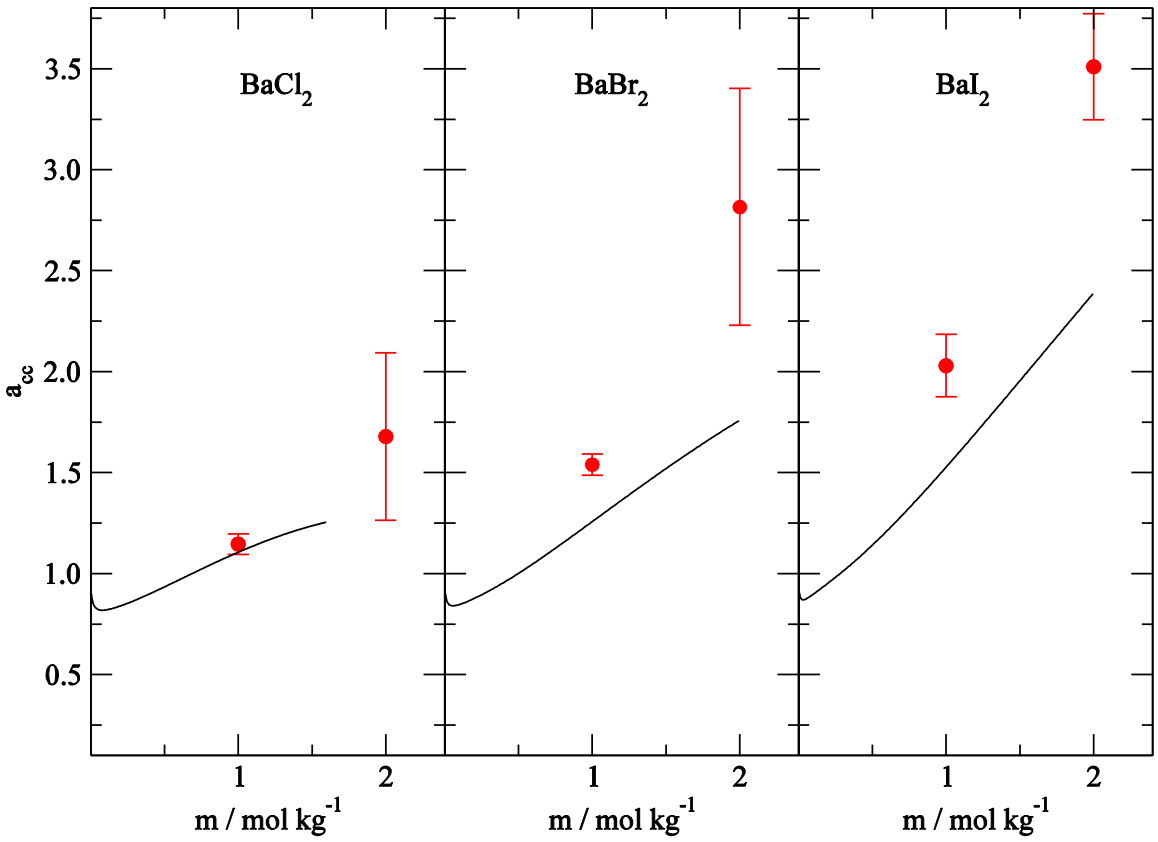

Figure S 2.6. Activity derivatives for barium salts as a function of the salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

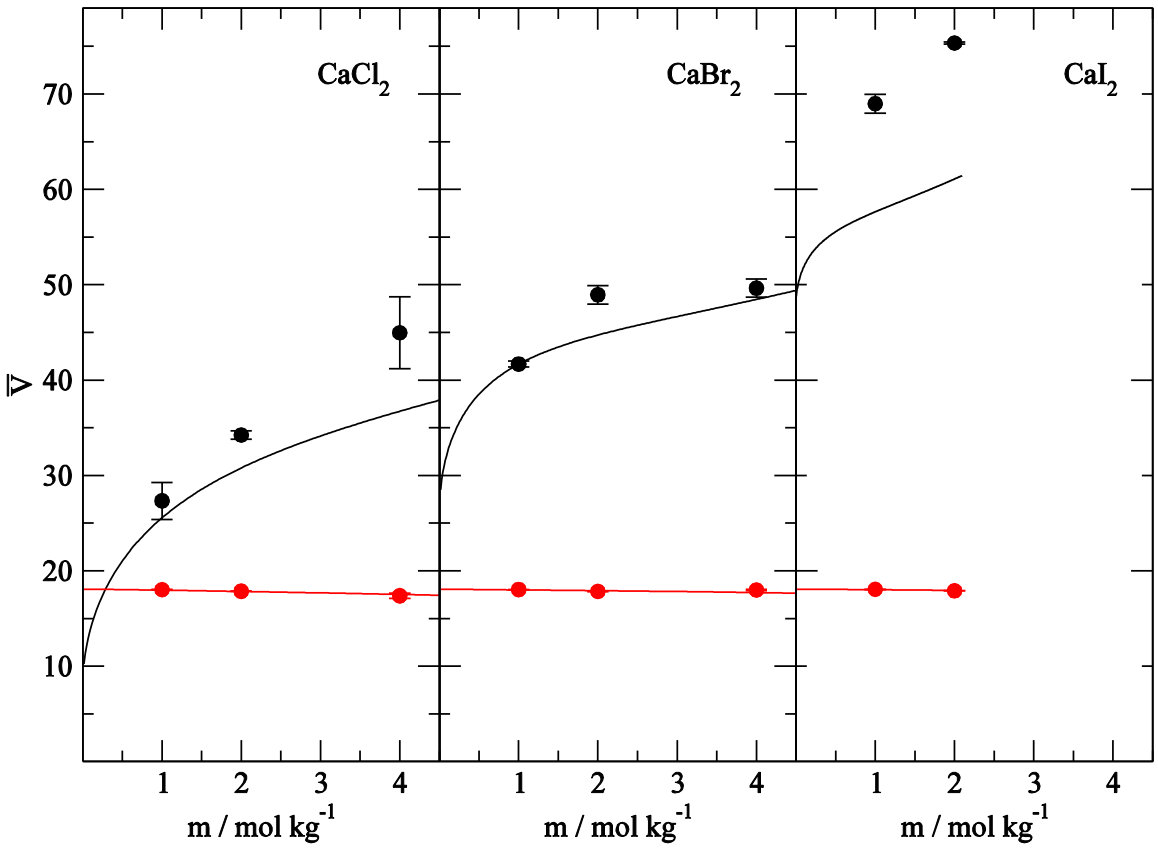

Figure S 2.7. Partial molar volumes ($cm^3\ mol^{-1}$) of the salt ions for the calcium salts as a function of the salt molality. Partial molar volumes of ions displayed in black and partial molar volumes of the water displayed in red color. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

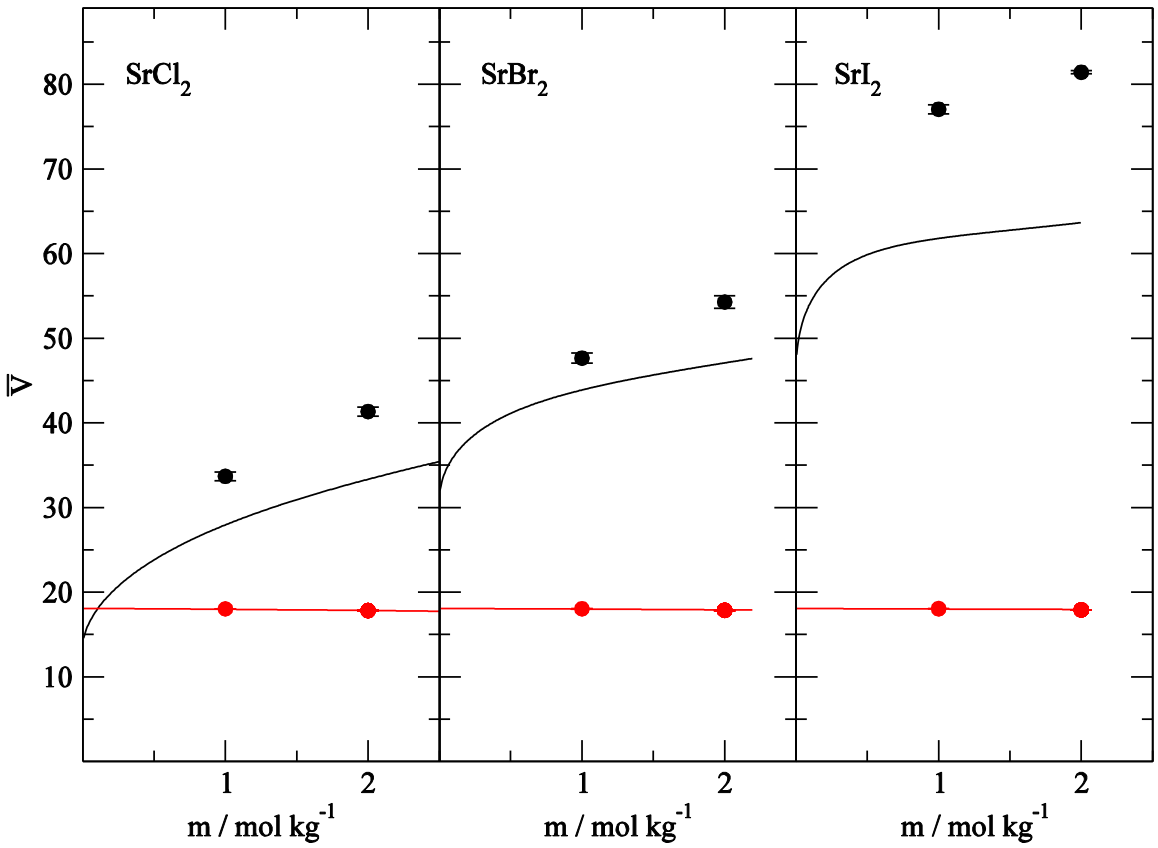

Figure S 2.8. Partial molar volumes ($cm^3\ mol^{-1}$) of the salt ions for the strontium salts as a function of the salt molality. Partial molar volumes of ions displayed in black and partial molar volumes of the water displayed in red color. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

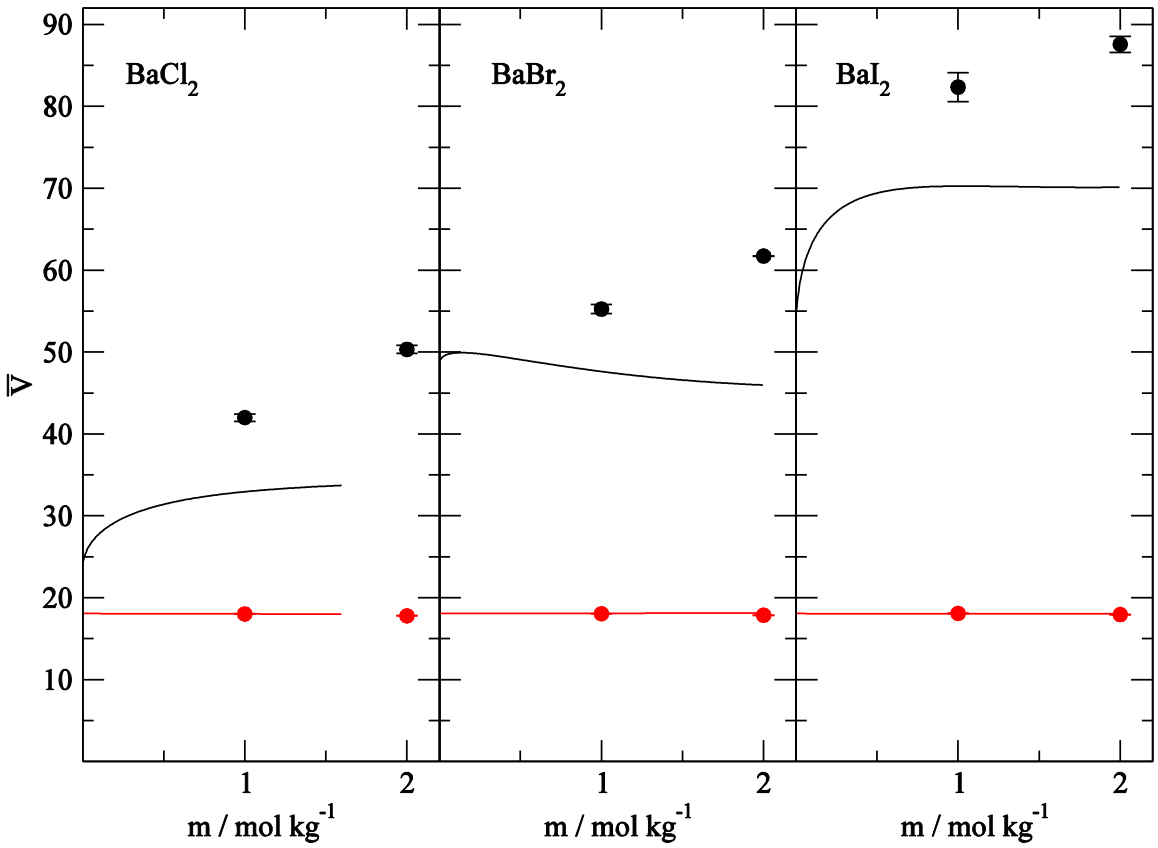

Figure S 2.9. Partial molar volumes ($cm^3\ mol^{-1}$) of the salt ions for the barium salts as a function of the salt molality. Partial molar volumes of ions displayed in black and partial molar volumes of the water displayed in red color. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

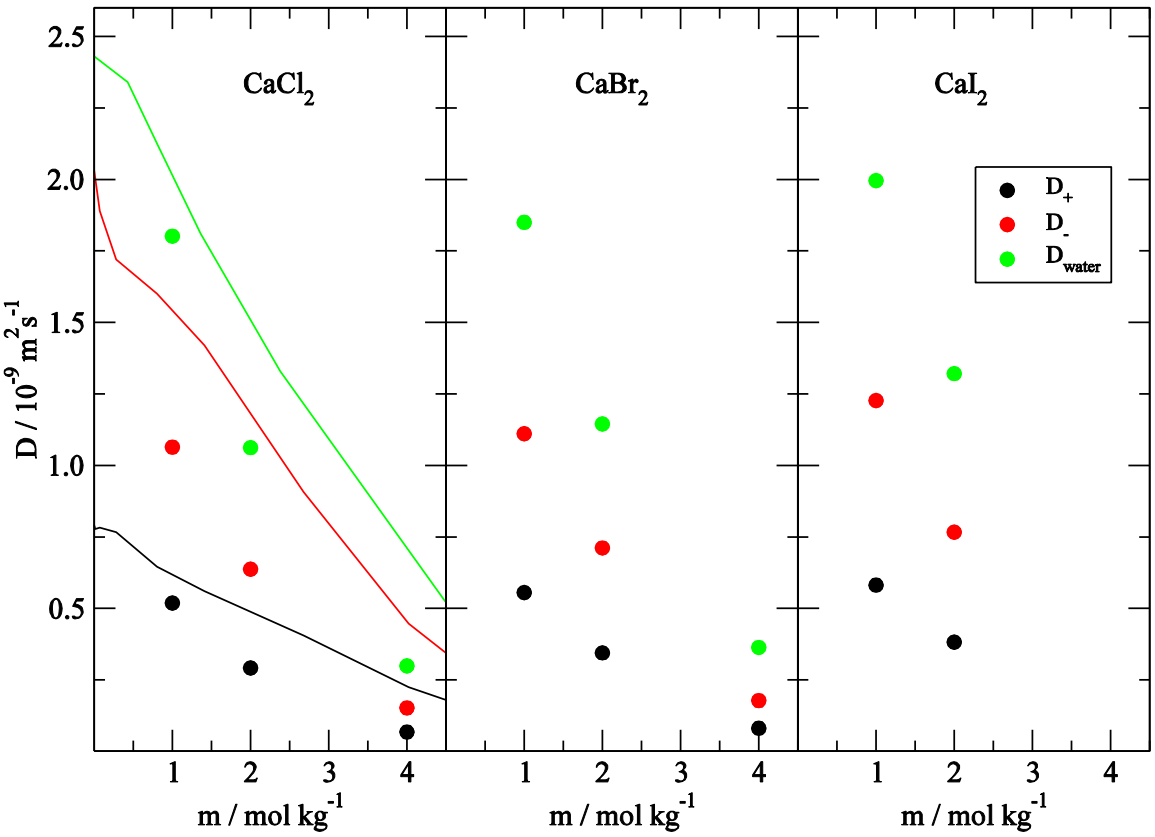

Figure S 2.10. Diffusion constants for calcium salts as a function of the salt molality. The lines represent the experimental diffusion constant data, the symbols correspond to the MD simulations.

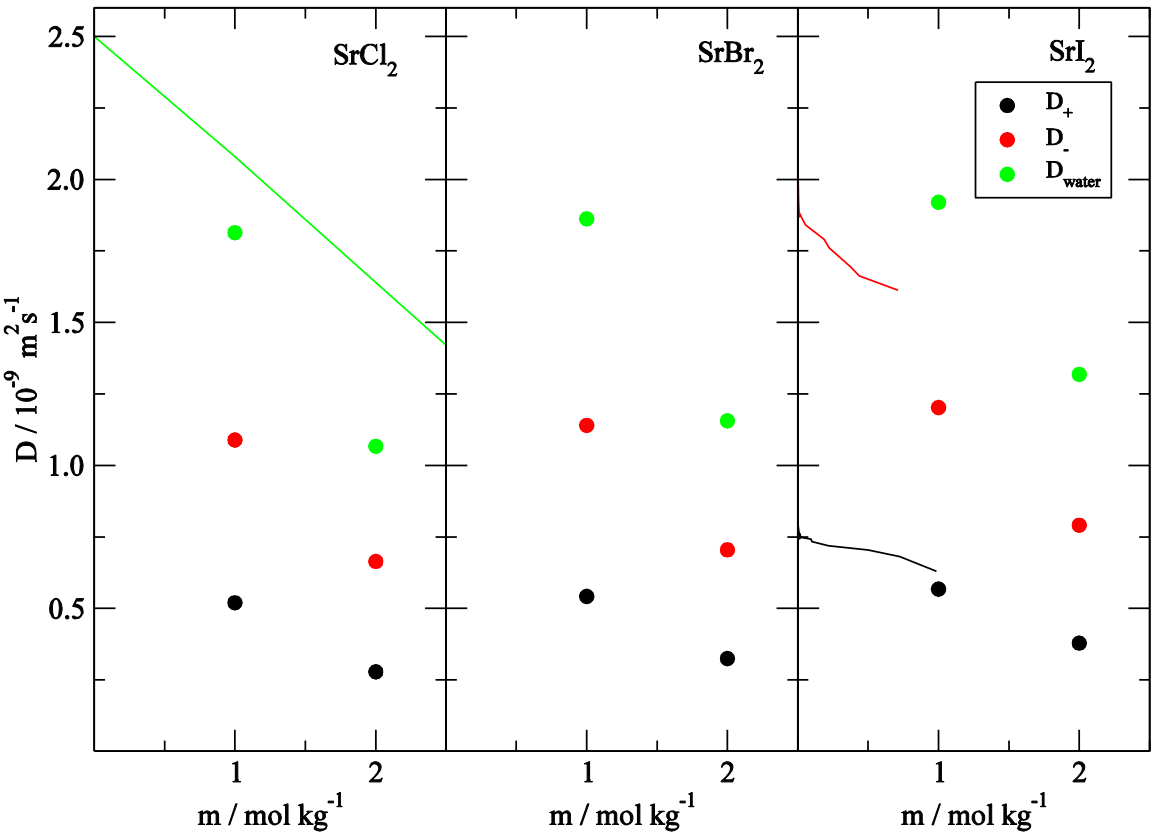

Figure S 2.11. Diffusion constants for strontium salts as a function of the salt molality. The lines represent the experimental diffusion constant data, the symbols correspond to the MD simulations.

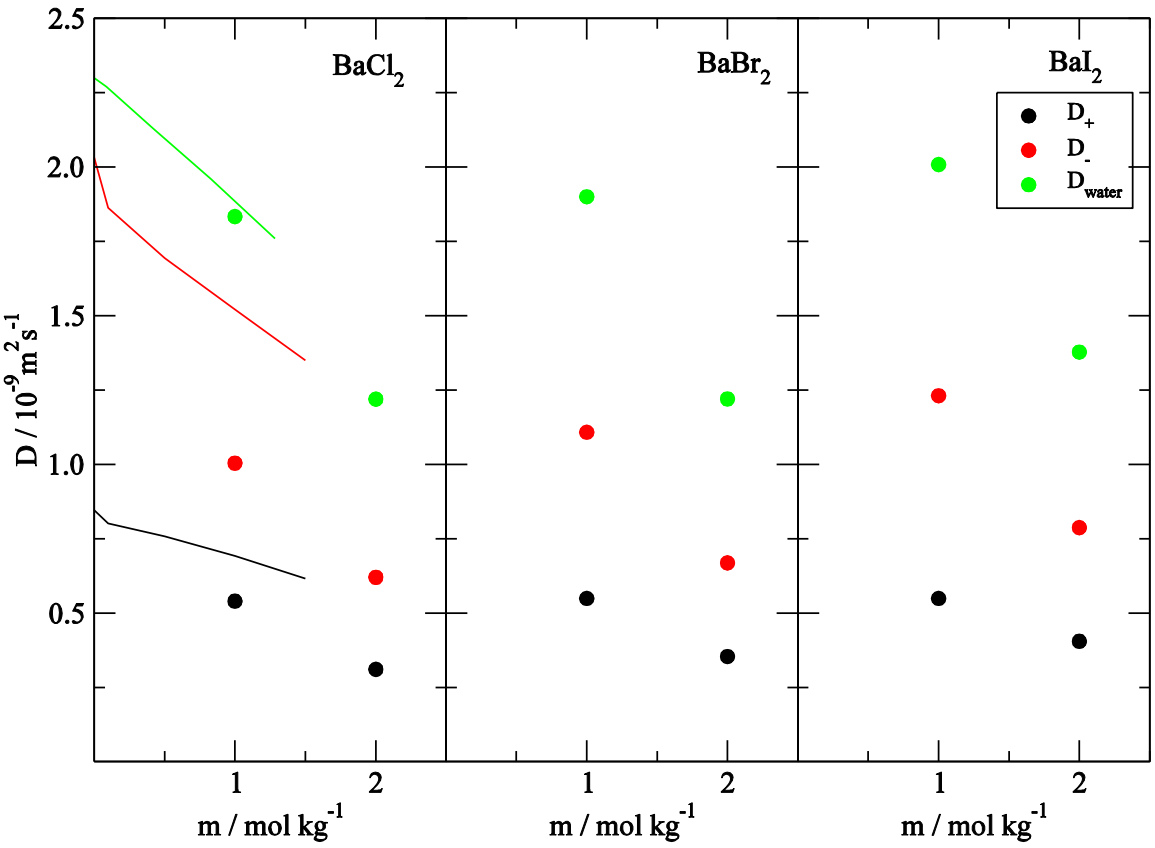

Figure S 2.12. Diffusion constants for barium salts as a function of the salt molality. The lines represent the experimental diffusion constant data, the symbols correspond to the MD simulations.

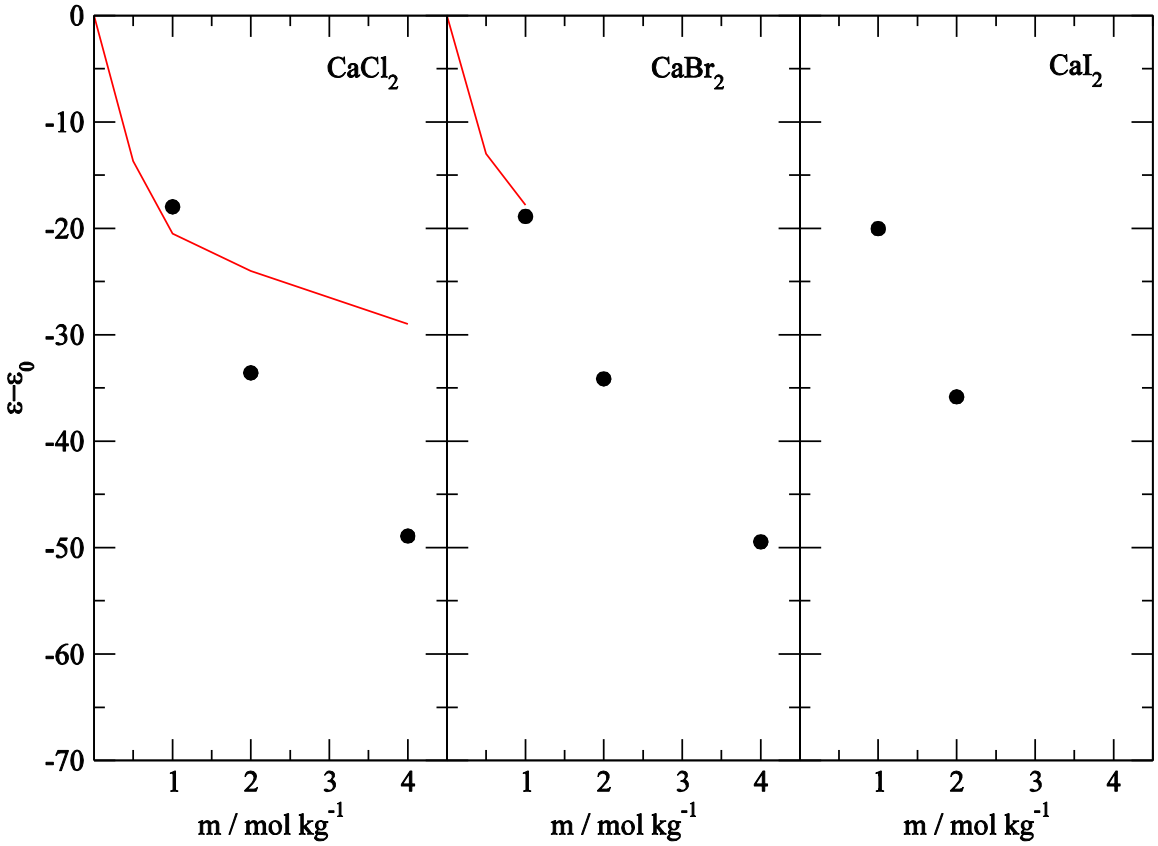

Figure S 2.13. Dielectric decrements ($\varepsilon - \varepsilon_0$)for the calcium salts as a function of the salt molality (mol kg^{-1}). Lines are for the experimental data and symbols are for the MD simulations.

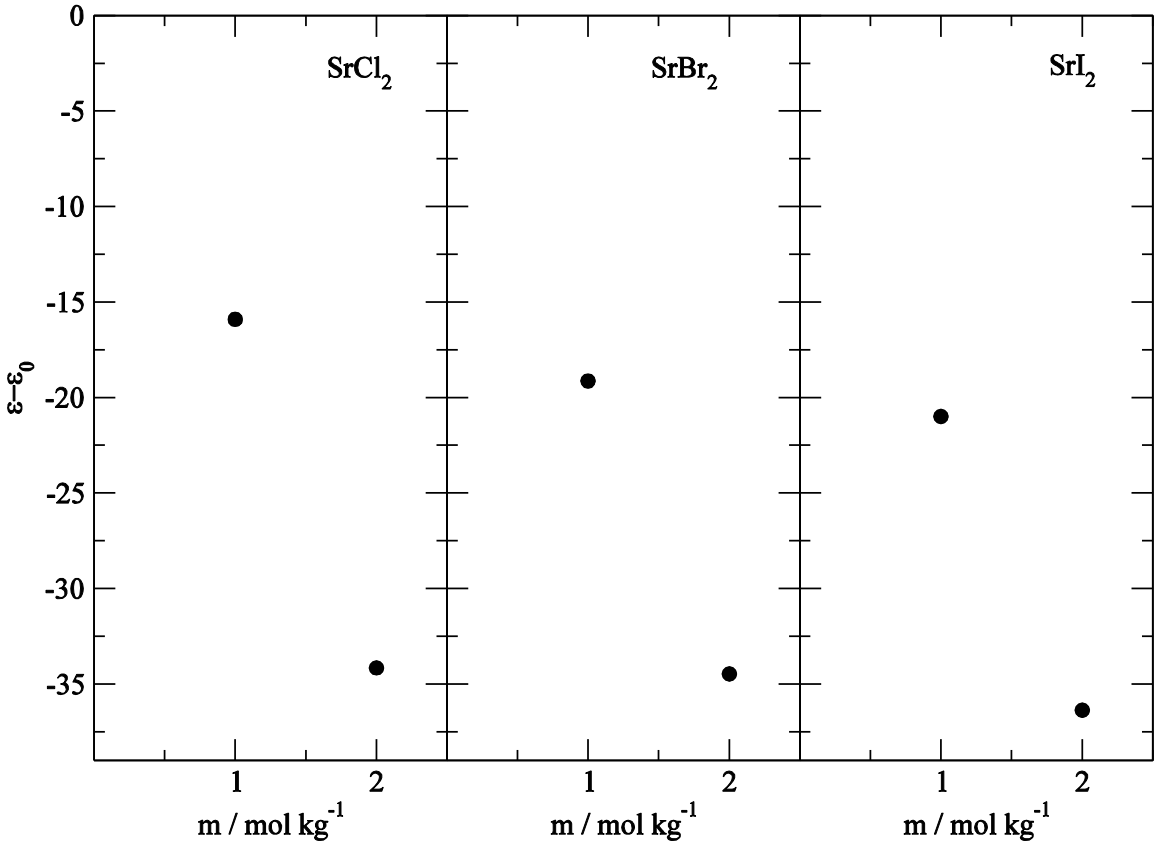

Figure S 2.14. Dielectric decrements (ε - ε_0) for the strontium salts as a function of the salt molality (mol kg^{-1}). Lines are for the experimental data and symbols are for the MD simulations.

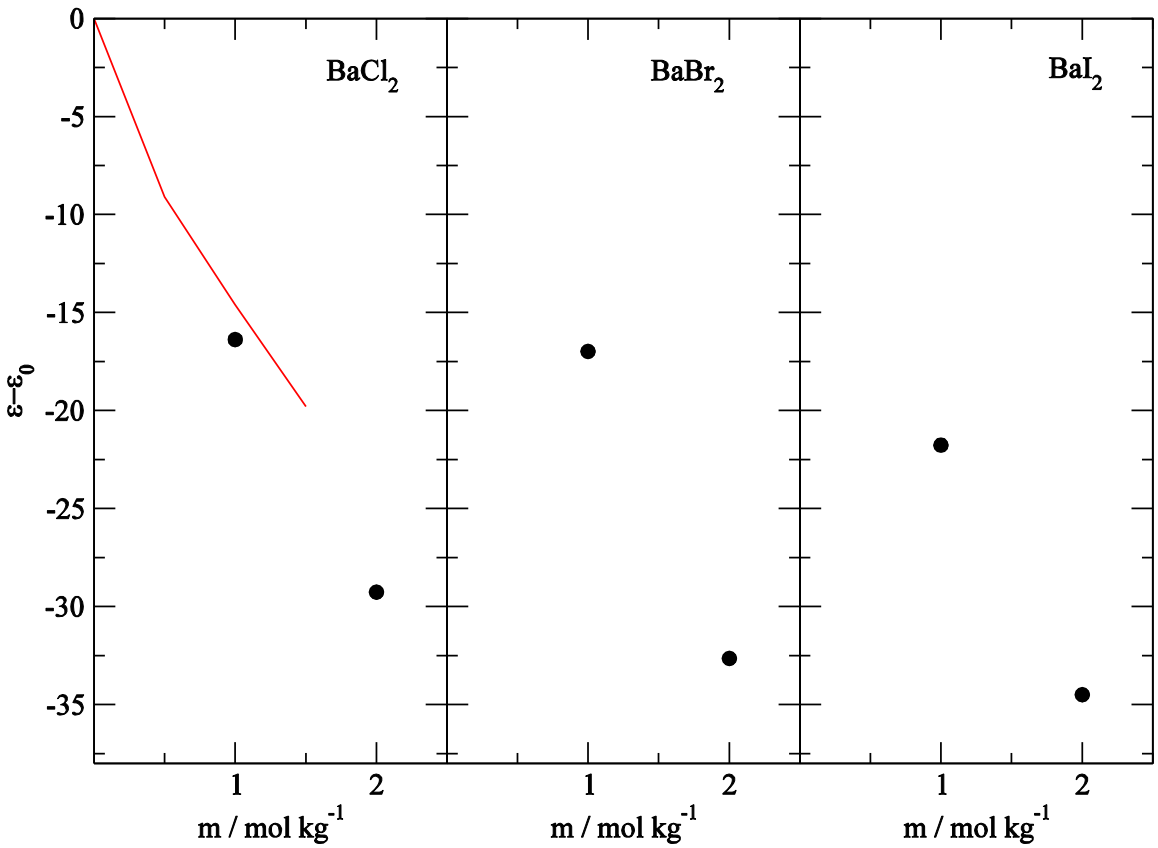

Figure S 2.15. Dielectric decrements ($\varepsilon - \varepsilon_0$) for the barium salts as a function of the salt molality (mol kg^{-1}). Lines are for the experimental data and symbols are for the MD simulations.

Table S 2.3. The comparison of the salt crystals potential energies per molecule and densities of simulation with the experimenla values. The non orthognal crystal (90°, 90°, 120°) forms are denoted by astrix sign (*).

	E_{pot} (kj/mol)	Non Orthogonal E_{pot} (kj/mol)	ρ_{sim} (g/cm^3)	Non Orthogonal ρ_{sim} (g/cm^3)	ρ_{exp} (g/cm^3)
*$MgCl_2$	-2387.54	-2399.38	1.876	2.085	2.408
*$MgBr_2$	-2321.9	-2321.21	3.126	3.364	3.885
*MgI_2	-2162.02	-2179.26	3.738	3.901	4.504
$CaCl_2$	-2187.97		1.918		2.187
$CaBr_2$	-2113.65		3.138		3.38
*CaI_2	-1970.27	-1937.12	3.95	3.494	4.034
$SrCl_2$	-2111.26		2.519		3.101
$SrBr_2$	-2031.05		3.638		3.638
SrI_2	-1928.2		4.124		4.589
$BaCl_2$	-1975.45		3.387		3.498
$BaBr_2$	-1904.6		4.343		4.78
BaI_2	-1790.23		5.135		5.15

Table S 2.4. The comparison of the simulated salt alkaline earth hallide crystals unit cell diamensions with the experimental value. The non orthognal crystal (90°, 90°, 120°) forms are denoted by astrix sign (*).

	a_{sim} (nm)	a_{exp} (nm)	b_{sim} (nm)	b_{exp} (nm)	c_{sim} (nm)	c_{exp} (nm)
*$MgCl_2$		0.3596		0.3596		1.7589
*$MgBr_2$		0.381		0.381		0.626
*MgI_2		0.4157		0.4157		0.6862
$CaCl_2$	0.663	0.624	0.663	0.643	0.4373	0.42
$CaBr_2$	0.686	0.655	0.686	0.688	0.449	0.434
*CaI_2		0.448		0.448		0.696
$SrCl_2$	0.7015	0.69767	0.7427	0.69767	0.8023	0.69767
$SrBr_2$	0.8299	0.92	1.062	1.142	0.529	0.43
SrI_2	1.64	1.522	0.746	0.822	0.899	0.79
$BaCl_2$	0.8313	0.7865	0.504	0.4731	0.9745	0.9421
$BaBr_2$	0.861	0.8275	0.524	0.4956	1.006	0.9919
BaI_2	0.943	0.892	0.6444	0.5304	1.1097	1.0695

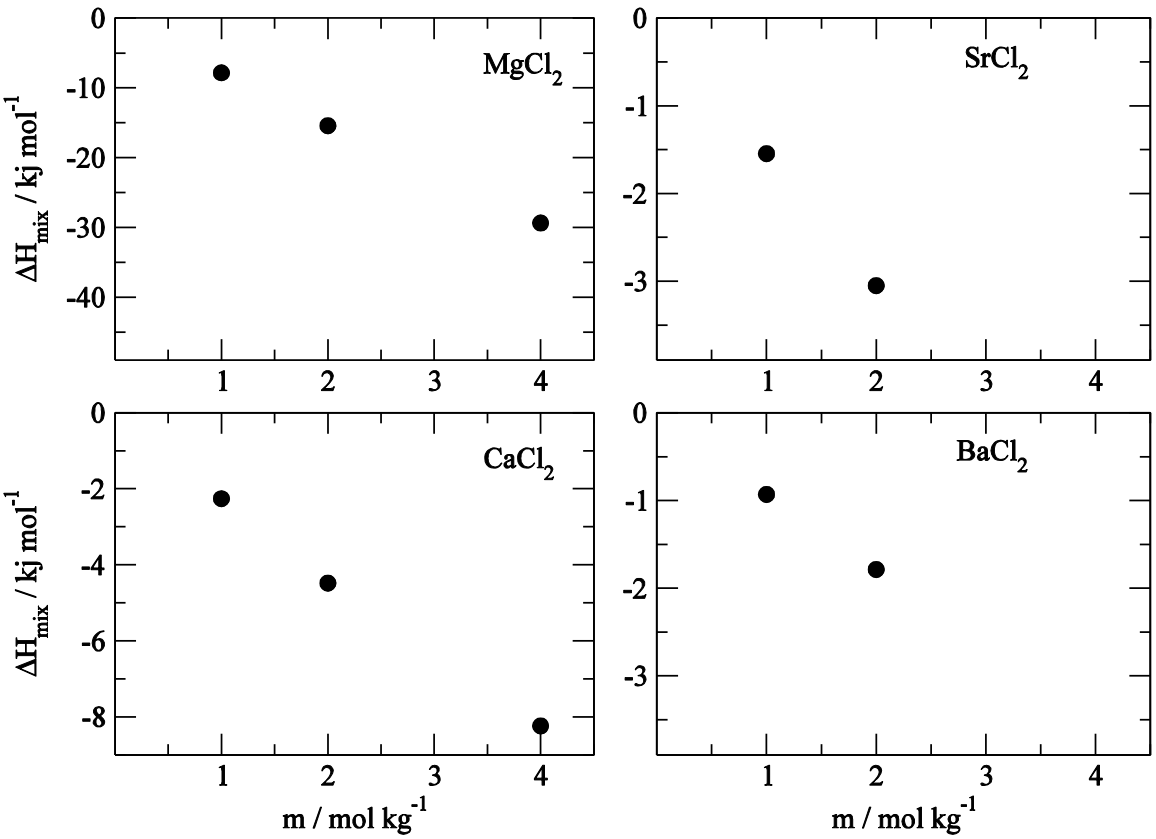

Figure S 2.16. Excess enthalpy of mixing for chloride salts as a function of salt molality. The experimental data not available to compare with the symbols from the MD simulations.

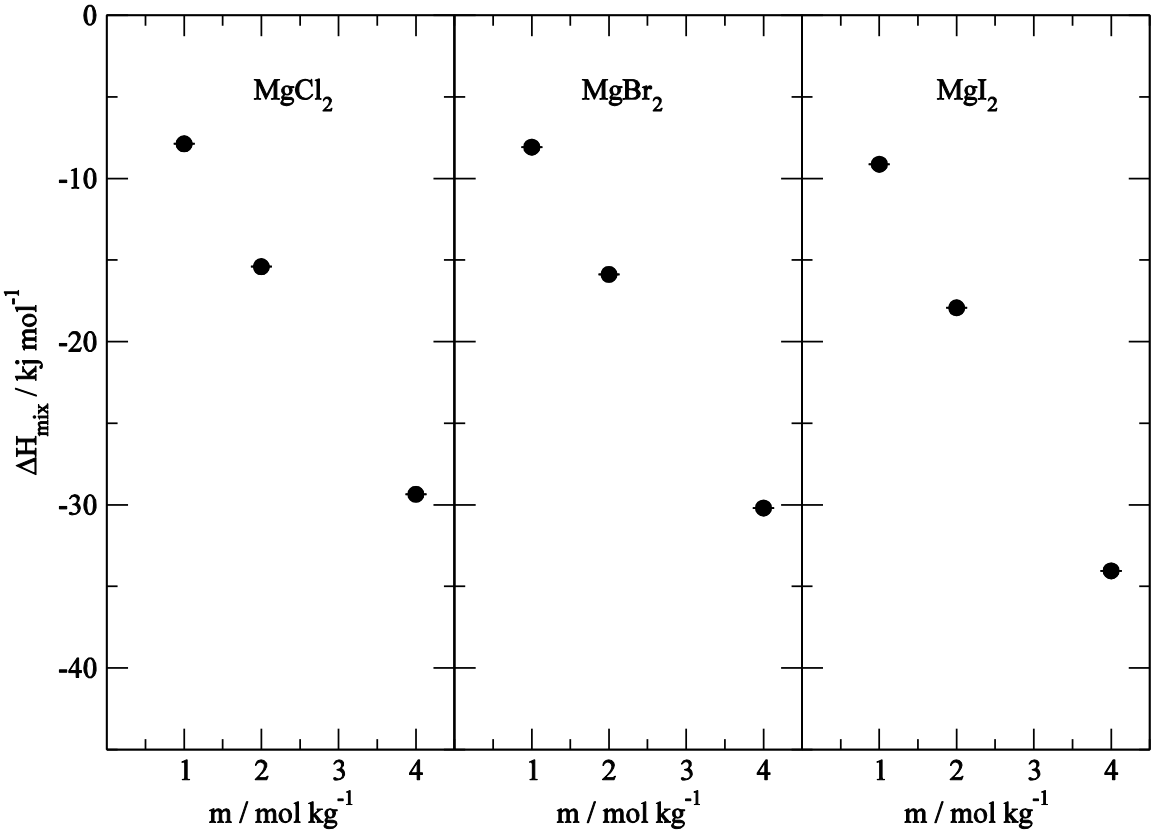

Figure S 2.17. Excess enthalpy of mixing for magnesium salts as a function of salt molality. The experimental data not available to compare with the symbols from the MD simulations.

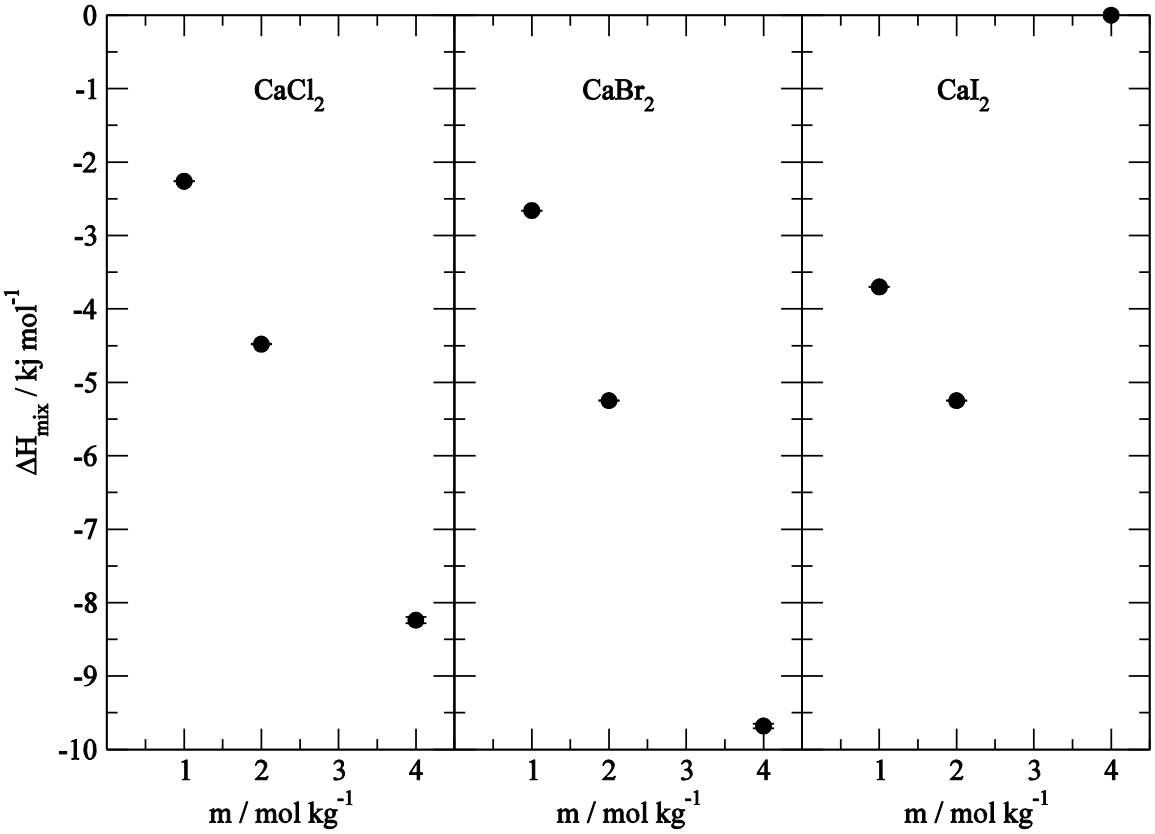

Figure S 2.18. Excess enthalpy of mixing for calcium salts as a function of salt molality. The experimental data not available to compare with the symbols from the MD simulations.

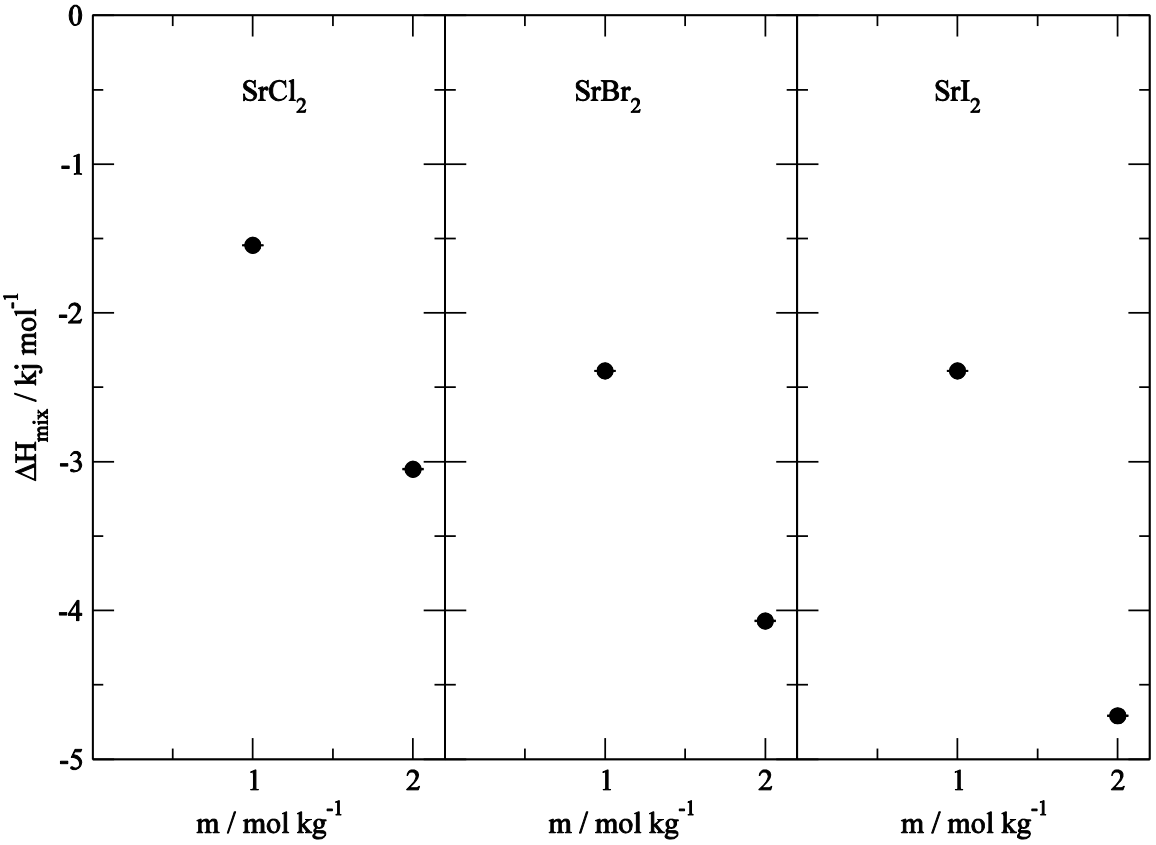

Figure S 2.19. Excess enthalpy of mixing for strontium salts as a function of salt molality. The experimental data not available to compare with the symbols from the MD simulations.

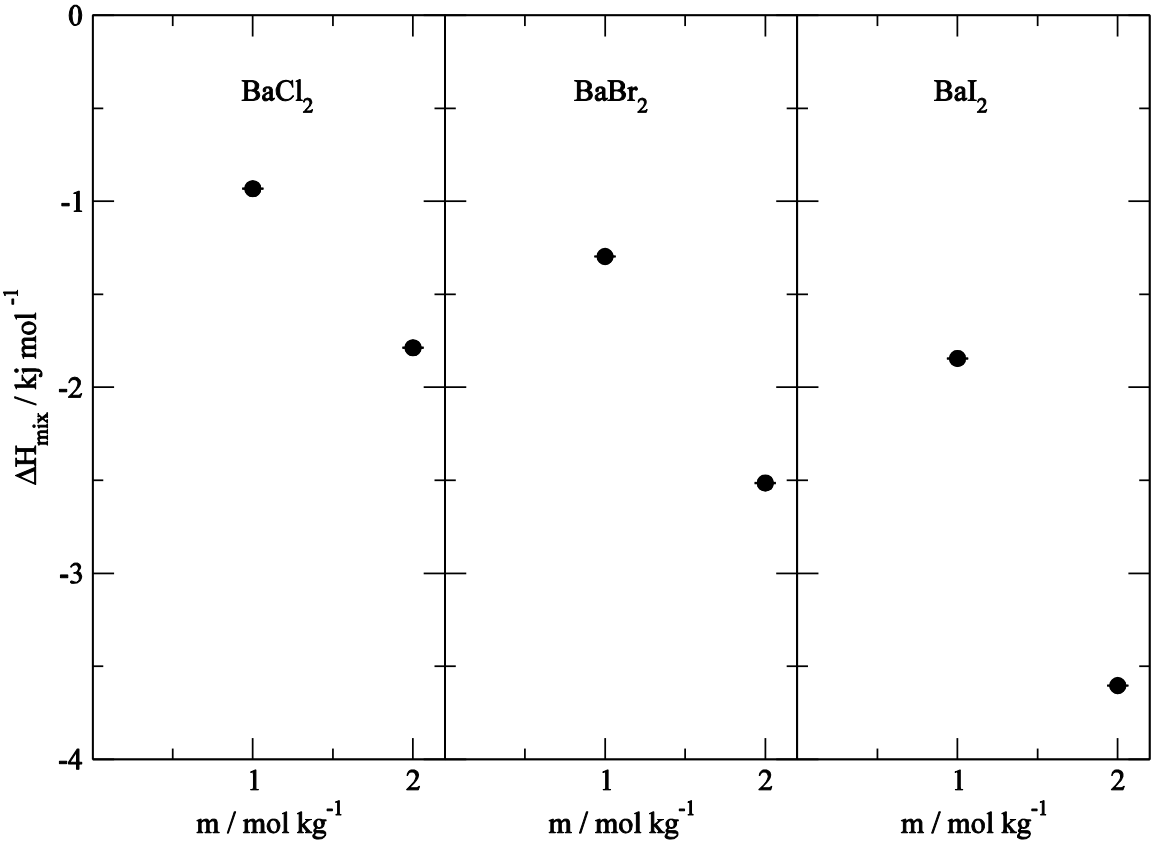

Figure S 2.20. Excess enthalpy of mixing for barium salts as a function of salt molality. The experimental data not available to compare with the symbols from the MD simulations.

References

1. Frausto da Silva, J. J. R.; William, R. J. P., *The Biological Chemistry of the Elements: The Inorganic Chemistry of Life*. 2nd ed.; Oxford University Press: 1991.

2. Jiang, Y.; Lee, A.; Chen, J.; Ruta, V.; Cadene, M.; Chait, B. T.; MacKinnon, R., *Nature* **2003,** *423*, 33-41.

3. Mordasini, T.; Curioni, A.; Andreoni, W., *Journal of Biological Chemistry* **2003,** *278*, 4381-4384.

4. Stephanie X. Wang; Eugene Hur; Carolyn A. Sousa; Linda Brinen; Eric J. Slivka; Fletterick, R. J., *Biochemistry* **2003,** *42* (26), 7959–7966.

5. Cowan, J. A., *Chemical Reviews* **1998,** *98*, 1067-1088.

6. Dudev, T.; Lim, C., *Journal of the American Chemical Society* **2013,** *135*, 17200-17208.

7. Leontyev, I.; Stuchebrukhov, A., *Physical Chemistry Chemical Physics* **2011,** *13* (7), 2613.

8. Ichiye, T., Water in the Liquid State: A Computational Viewpoint. In *Advances in Chemical Physics*, John Wiley & Sons, Inc.: 2014; pp 161-200.

9. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2003,** *119*, 11342.

10. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E., *Journal of Chemical Theory and Computation* **2011,** *7* (5), 1369-1380.

11. J. Šponer; M. Otyepka; P. Banáš; Réblová, K.; Walter, N. G., Molecular Dynamics Simulations of RNA Molecules. In *Innovations in biomolecular modeling and simulations*, Royal Society of Chemistry: 2012; Vol. 2, pp 129-155.

12. Israelachvili, J.; Wennerström, H., *Nature* **1996,** *379*, 219-225.

13. Schneck, E.; Sedlmeier, F.; Netz, R. R., *Proceedings of the National Academy of Sciences of the United States of America* **2012,** *109* (36), 14405-14409.

14. Chen, H.; Cox, J. R.; Ow, H.; Shi, R.; Panagiotopoulos, A. Z., *Nature Publishing Group* **2016,** *6*, 1-10.

15. Dietz, W.; Riede, W. O.; Heinzinger, K., *Zeitschrift für Naturforschung A* **1982,** *37*, 1038-1048.

16. Probst, M. M.; Radnai, T.; Heinzinger, K.; Bopp, P.; Rode, B. M., *The Journal of Physical Chemistry* **1985,** *89*, 753-759.

17. Àqvist, J., *The Journal of Physical Chemistry* **1990,** *94*, 8021-8024.

18. Guàrdia, E.; Robinson, A.; Padró, J. A., *The Journal of Chemical Physics* **1993,** *99*, 4229.

19. Gavryushov, S.; Linse, P., *The Journal of Physical Chemistry B* **2006,** *110*, 10878-10887.

20. Hashem, Y.; Westhof, E.; Auffinger, P., Milestones in Molecular Dynamics Simulations of RNA Systems. In *Computational Structural Biology*, WORLD SCIENTIFIC: 2008; pp 363-399.

21. Ohtaki, H., *Monatshefte fuer Chemie* **2002,** *132* (11), 1237-1268.

22. Kirmizialtin, S.; Elber, R., *The Journal of Physical Chemistry B* **2010,** *114*, 8207-8220.

23. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. SOC* **1995,** *117*, 5179-5197.

24. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M., *Journal of Computational Chemistry* **1983,** *4*, 187-217.

25. Yoo, J.; Aksimentiev, A., *The Journal of Physical Chemistry Letters* **2012,** *3*, 45-50.

26. Auffinger, P., Ions in Molecular Dynamics Simulations of RNA Systems. In *RNA 3D Structure Analysis and Prediction*, Springer Berlin Heidelberg: 2012; Vol. 27, pp 299-318.

27. Auffinger, P.; Grover, N.; Westhof, E., 1 Metal Ion Binding to RNA. In *Structural and Catalytic Roles of Metal Ions in RNA*, The Royal Society of Chemistry: 2011; Vol. 9, pp 1-36.

28. Yoo, J.; Wilson, J.; Aksimentiev, A., *Biopolymers* **2016,** *105*, 752-763.

29. Jiang, Y.; Zhang, H.; Feng, W.; Tan, T., *J. Chem. Inf. Model* **2015,** *55* (12), 2575-2586.

30. Saxena, A.; Sept, D., *Journal of Chemical Theory and Computation* **2013,** *9*, 3538-3542.

31. Saxena, A.; García, A. E., *The Journal of Physical Chemistry B* **2015,** *119*, 219-227.

32. Panteva, M. T.; Giambaşu, G. M.; York, D. M., *Journal of computational chemistry* **2015,** *36*, 970-82.

33. Marchand, S.; Roux, B., *Proteins: Structure, Function, and Genetics* **1998,** *33*, 265-284.

34. Babu, C. S.; Lim, C., *The Journal of Physical Chemistry A* **2006,** *110*, 691-699.

35. Mamatkulov, S.; Fyta, M.; Netz, R. R., *The Journal of Chemical Physics* **2013,** *138*, 024505.

36. Van Gunsteren, W. F.; Berendsen, H. J. C., *Library manual (Nijenborgh, Groningen)* **1987**, 1-221.

37. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19*, 774.

38. Kusalik, P. G.; Patey, G. N., *The Journal of Chemical Physics* **1987,** *86* (9), 5110-5116.

39. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67*, 4884.

40. Smith, P. E., *The Journal of chemical physics* **2008,** *129*, 124509.

41. Söhnel, O.; Novotný, P., *Densities of aqueous solutions of inorganic substances*. Elsevier: Amsterdam ;, 1985; p 335.

42. Lobo, V. M. M.; Quaresma, J. L., *Handbook of electrolyte solutions*. Elsevier: 1989; p 2354.

43. Pitzer, K. S., *The Journal of Physical Chemistry* **1973,** *77*, 268-277.

44. Pitzer, K. S.; Wang, P.; Rard, J. A.; Clegg, S. L., *Journal of Solution Chemistry* **1999,** *28*, 265-282.

45. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2003,** *107*, 3891-3898.

46. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

47. Marcus, Y., *Chemical Reviews* **1988,** *88*, 1475-1498.

48. Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A., *Journal of applied crystallography* **2009,** *42*, 726-729.

49. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Computer Physics Communications* **1995,** *91*, 43-56.

50. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81*, 3684.

51. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

52. Darden, T.; York, D.; Pedersen, L., *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

53. Smith, P. E.; Matteoli, E.; O' Connell, J. P., *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*. CRC Press: Boca Raton, 2013.

54. Chandrasekhar, S., *Reviews of Modern Physics* **1943,** *15* (1), 1-89.

55. Chitra, R.; Smith, P. E., *The Journal of Physical Chemistry B* **2000,** *104* (24), 5854-5864.

56. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: USA, 1987.

57. Smith, P. E.; Gunsteren, W. F. v., *The Journal of Chemical Physics* **1994,** *100* (1), 577-585.

58. Caminiti, R.; Licheri, G.; Piccaluga, G.; Pinna, G., *Chemical Physics Letters* **1977,** *47*, 275-278.

59. Ohtaki, H.; Radnai, T., *Chemical Reviews* **1993,** *93*, 1157-1204.

60. Larentzos, J. P.; Criscenti, L. J., *The Journal of Physical Chemistry B* **2008,** *112*, 14243-14250.

61. Piquemal, J. P.; Perera, L.; Cisneros, G. A.; Ren, P.; Pedersen, L. G.; Darden, T. A., *Journal of Chemical Physics* **2006,** *125*, 054511.

62. Jiao, D.; King, C.; Grossfield, A.; Darden, T. A.; Ren, P., *The Journal of Physical Chemistry B* **2006,** *110*, 18553-18559.

63. Jalilehvand, F.; Spångberg, D.; Lindqvist-Reis, P.; Hermansson, K.; Persson, I.; Sandström, M., *Journal of the American Chemical Society* **2001,** *123*, 431-441.

64. Badyal, Y. S.; Barnes, A. C.; Cuello, G. J.; Simonson, J. M., *Journal of Physical Chemistry A* **2004,** *108*, 11819-11827.

65. Fulton, J. L.; Heald, S. M.; Badyal, Y. S.; Simonson, J. M., *Journal of Physical Chemistry A* **2003,** *107*, 4688-4696.

66. Pham, V.-T.; Fulton, J. L., *The Journal of Chemical Physics* **2013,** *138*, 044201.

67. Smirnov, P. R.; Trostin, V. N., *ISSN Russian Journal of General Chemistry Original Russian Text © Zhurnal Obshchei Khimii* **2009,** *79*, 1070-3632.

68. Neilson, G. W.; Broadbent, R. D., *Chemical Physics Letters* **1990,** *167*, 429-431.

69. Dang, L. X.; Schenter, G. K.; Fulton, J. L., *Journal of Physical Chemistry B* **2003,** *107*, 14119-14123.

70. Ramos, S.; Neilson, G. W.; Barnes, A. C.; Capitán, M. J., *The Journal of Chemical Physics* **2003,** *118*, 5542.

71. D'Angelo, P.; Migliorati, V.; Sessa, F.; Mancini, G.; Persson, I., *The Journal of Physical Chemistry B* **2016,** *120*, 4114-4124.

72. Mills, R. R.; Lobo, V. M. M., *Self-diffusion in electrolyte solutions : a critical examination of data compiled from the literature*. Elsevier: 1989; p 346.

73. Obst, S.; Bradaczek, H., *The Journal of Physical Chemistry* **1996,** *100*, 15677-15687.

74. Koneshan, S.; Rasaiah, J. C.; Lynden-Bell, R. M.; Lee, S. H., *The Journal of Physical Chemistry B* **1998,** *102*, 4193-4204.

75. Barthel J.; Buchner R.; Münsterer M., *Electrolyte Data Collection, Part 2: Dielectric Properties of Water and Aqueous Electrolyte Solutions,* . Wiley-VCH Verlag GmbH & Co. KGaA: Frankfurt, 1996; Vol. 100, p 1922-1922.

76. Smith, P. E.; van Gunsteren, W. F., *The Journal of Chemical Physics* **1994,** *100*, 3169.

Chapter 3 - Molecular Dynamics Simulations of Oxo Anion Salt Solutions

3.1. Abstract

The oxo anions, such as NO_3^-, ClO_4^-, SO_4^{2-} and CO_3^{2-}, play a vital role in physiological systems by regulating biological processes. These ions form part of the Hofmeister series, which is used to explain biological processes such as protein crystallization, nucleic acid and protein stability, enzyme activity, and many others. However, at a molecular level the mechanism of these Hofmeister ions is still not fully understood. Therefore, computer simulations are often used to assist experimentalists to understand the dynamics of these hydrated anion and their interactions with biomolecules. Here, we develop classical non-polarizable force fields for the NO_3^-, ClO_4^-, and SO_4^{2-} ions which are compatible with the SPC/E water model. These new force field parameters are specifically developed to reproduce the experimental Kirkwood-Buff integrals. Use of these new models in MD simulations also leads to reasonable diffusion constants and dielectric decrements. Unfortunately, attempts to develop force fields for CO_3^{2-} ions were unsuccessful due to an excessive aggregation behavior in the simulations.

3.2. Introduction

Some oxo anions play a vital role in physiological systems by helping to regulate biological processes. In the present work anions, the activity of anions, such as NO_3^-, ClO_4^-, SO_4^{2-}, and CO_3^{2-}, in aqueous solutions is investigated. These ions form part of the Hofmeister series.[1] Hydrogen bonds between the water molecules are the most abundant electrostatic interaction in the nature. In physical chemistry one of the longstanding and fundamental problems is to modeling ion

interactions in water. In particular, capturing ion specificity, which is very common in biological systems, is very difficult to reproduce in computer simulations.

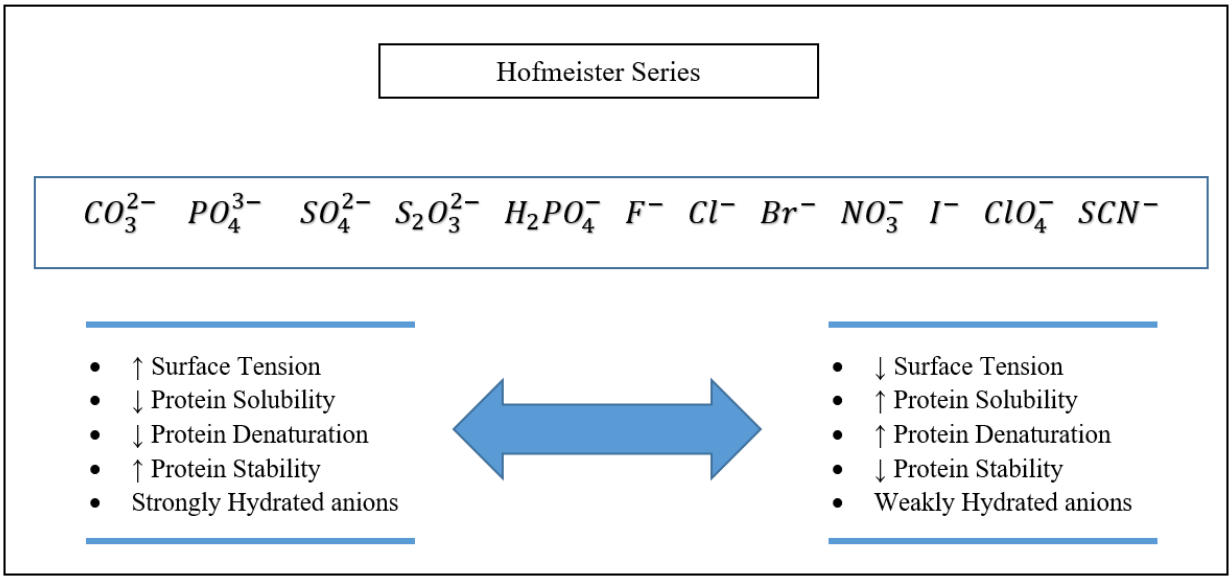

Figure 3.1. The Hofmeister Series for Anions

The Hofmeister series, as displayed in the Figure 3.1, is often used to help explain some important biological processes such as protein crystallization[2], nucleic acid and protein stability[3], enzyme activity,[4] and many more. However, at a molecular level the mechanism of action of these Hofmeister ions is still not fully understood. Therefore, computer simulations are often used to assist experimentalist to understand the dynamics of the hydrated anion and its interactions with biomolecules.

The addition of ions to water disrupts the hydrogen bond network between water molecules since the ion interacts strongly with any surrounding water molecules. The formation of water

spheres around the ion is known as hydration. The hydration of ions is a very important phenomenon in nature which extends to the hydration of minerals in oceans, and to ion binding to biomolecules inside a living organism. Negative ions are claimed to be more stably solvated compared to positive ions of the same size in some experimental studies and in some MD simulation studies.[5-6]

Hydrogen bonds are weaker by several orders of magnitude compare to covalent bonds. At room temperature, the hydrogen bond angle and length can fluctuate on the sub pico second time scale.[7] Experimental studies have shown that water molecule reorientation happens (on average) every 2.5 ps or so.[8] Hence, the hydrogen bond fluctuates several times before reorientation of a water molecule happens. Despite decades of research to understand the hydration process, fundamental questions remain unresolved. For instance, what is the impact of the exchange of water molecules in a solvation shells on the mobility of an ion.

There are significantly more experimental and theoretical studies on metal ions compared to oxo anions in aqueous solutions. In a 2014 publication, it is stated that the oxo anions which have been studied by both experimental and theoretical approaches includes only sulfate, sulfite, thiosulfate, selenate and chromate ions.[9] The relatively weak hydrogen bonds between the water molecules and hydrated anions, and the faster exchange rates, limit the use of most experimental techniques. However, some experimental techniques, such as large and x-ray scattering, have given some insight in to water structure and the water exchange rates of some oxo anions.[10]

The perchlorate salts are widely used in the physiochemical studies since perchlorate ions do not react with other chemical species in aqueous solutions. The low charge density of the mono charge per chlorate anions leads them to behave as structure breakers, with water exchange rates for these ions being on the same order of magnitude or slightly faster than in pure water.[9, 11-13]

The nitrate salts are readily soluble in water which makes it useful in the commercial and laboratory applications. The nitrate ions are also used in a research context because they plays an important role in the nitrogen cycle. Therefore, the nitrate ion has been extensively studied by experimental spectroscopic studies.[14-15] However, there is a growing interest in the scientific community to use both QM and MD computational simulations to help understand the hydration dynamics of aqueous nitrate ions.[16-19] There are also studies to help understand the correlation between the adsorption propensity of a material with the bulk hydration properties of the anions. As an example, one experimental study in 2012 reported that ions in aqueous sodium nitrate solutions display a special behavior.[20] That is, the surface adsorption of the ions of sodium nitrate solution happens as an ion pair. Therefore, there are efforts to determine the possibility of the adsorption of nitrate ions to the air-water interfaces as a contact ion pair with the sodium counter ion.

The sulfate ion is located near the salting out end of the Hofmeister series. Sulfate ions have high solvation free energy which implies a strong hydration of the sulfate ions in the aqueous solutions.[21] However, modeling divalent oxo anion like sulfate in non-polarizable MD simulations is a challenging task. Indeed, most of the MD simulations of sulfate ions with different force fields have reported an unrealistic aggregation behavior.[22-23] However, with polarizable force fields this aggregating behavior is not observed in the simulations.[24] Therefore, missing polarization effects may be a reason for the aggregation behavior in non-polarizable force fields, yielding excessive ion pairing and clustering. Dielectric relaxation spectra indicate that solvent separated and double solvent separated ion pairs are dominant over the contact ion pairs in sodium sulfate solutions.[25] Furthermore, Raman spectroscopy data does not indicate stronger contact ion pairs in the sodium sulfate solutions.[26]

Carbonate ions are also a very important type of ion in nature. Carbonate ions play an important role in biomineralization processes in nature that have intrigue scientist for decades.[27] Computer simulations can be used to compliment experiment to help understand the crystallization of carbonates. Also in the area of the geosequestration, dissolved carbon dioxide is captured as metal carbonates.[28] The oceans act as a sink to absorb about a quarter of the carbon dioxide gas released by industrialization.[29] This process generates 95% of monovalent hydrogen carbonate and 5% of divalent carbonate ions. However, this process lowers the pH level of sea water, which causes a lowering of carbonate ions available for sea creatures. There are efforts to use computer simulations to model this phenomenon and help forecast the impacts of this process.

However, to model carbonate ions, which have a high charge density and strong polarization effects on surrounding water molecules, is a challenging task. Irrespective of the ion or water model used, these carbonate ions tend to form unrealistic aggregation in the MD simulations.[30] The missing polarization effects in the non-polarizable force fields are argued to be the reason for this unrealistic behavior. One of the earliest attempt to model calcium carbonate was reported in the 1978 by Yuen et al.[31] There are studies that have tried to use different interatomic potentials combined with commonly used Morse potential to model the thermal expansion of calcite.[32] This study used a four state valance bond approach to capture distortion and polarization effects of the carbonate ions.[33] Therefore, there is a necessity to develop a force field for oxo anions which can reproduce several experimental chemical properties in MD simulations of these ions at finite concentrations.

3.3. Experimental Analysis and MD simulations

The same strategy as explained in the previous chapter, i.e. the Kirkwood Buff inversion procedure, was used to extract the experimental Kirkwood Buff integral (KBI).[34-35] Hence, to obtain experimental KBI values we used the experimental composition dependent activity and density data, after fitting to the Pitzer equation and a polynomial equation, respectively.[36] The experimental partial molar volumes were obtained from the experimental densities using previously established standard approaches.[37-38] Although sodium salts were used to parameterize the anions, whenever the experimental data were available the transferability of the same parameters for modeling the corresponding potassium salts also investigated. The optimization of the charge distribution for the anion to reproduce the experimental KB integrals for the respective salt is the main focus in this project. Therefore, a set of MD simulations were performed with different charge distributions for the anion in aqueous sodium salt solutions. Once the correct charge distribution for the sodium salt is found, the same charge destitution is attempted for the aqueous potassium salt solution.

All the molecular dynamics simulations were performed with Gromacs[39] 4.6 software. The SPC/E[40] water model used to model water in the simulations. All the ions were modeled using the KBFF force field for $NaNO_3$, KNO_3, $NaClO_4$, Na_2SO_4 and Na_2CO_3 systems. All simulations in this study maintained an isothermal isobaric ensemble at 300 K temperature and 1 atm pressure. The temperature and pressure were weakly coupled to a bath with relaxation times of 0.1 ps and 0.5 ps, respectively.[41] A 2 fs was used as the time step to integrate the equations of motion. All the bonds in the simulations were constrained with LINCS algorithm.[42] To calculate electrostatic interactions, the particle mesh Ewald technique[43] was used with a 1 nm cut off distance for real space calculations, while a 1.5 nm cutoff was used for the van der Waals interactions. A random

initial configuration of ions and waters in a cubic box with length of 10 nm was generated to provide a known concentration of aqueous salt solution using a custom written Fortran code. An energy minimization using the steepest descend method followed by 2 ns of equilibration was performed before the production run. The production run of 15 ns was used to calculate ensemble averages. In the production run the configurations were saved for analysis every 0.1 ps.

A custom written Fortran program was used to calculate the center of mass radial distribution function and to calculate the simulated KB integrals. The simulated KBI integrals were then used to calculate the activity derivatives, excess coordination numbers and partial molar volumes. Finally, translational self-diffusion coefficients were calculated using the mean square fluctuation approach,[44-45] and dielectric coefficients were calculated by analysis of the mean dipole moment fluctuations.[46-47]

3.4. Results and Discussion

The NO_3^- ion has a trigonal planer shape with negative one charge. The simulated and experimental excess coordination numbers $N_{ij} = \rho_i G_{ij}$, where the density is ρ_i and the KB integral is G_{ij}, of $NaNO_3$ solutions are shown in Figure 3.2. The best charge distribution for the NO_3^- ion, which reproduced the experimental KBIs in the simulations, is displayed in the Table 3.1. The MD simulations of aqueous $NaNO_3$ salts can reproduce experimental excess coordination numbers reasonably.

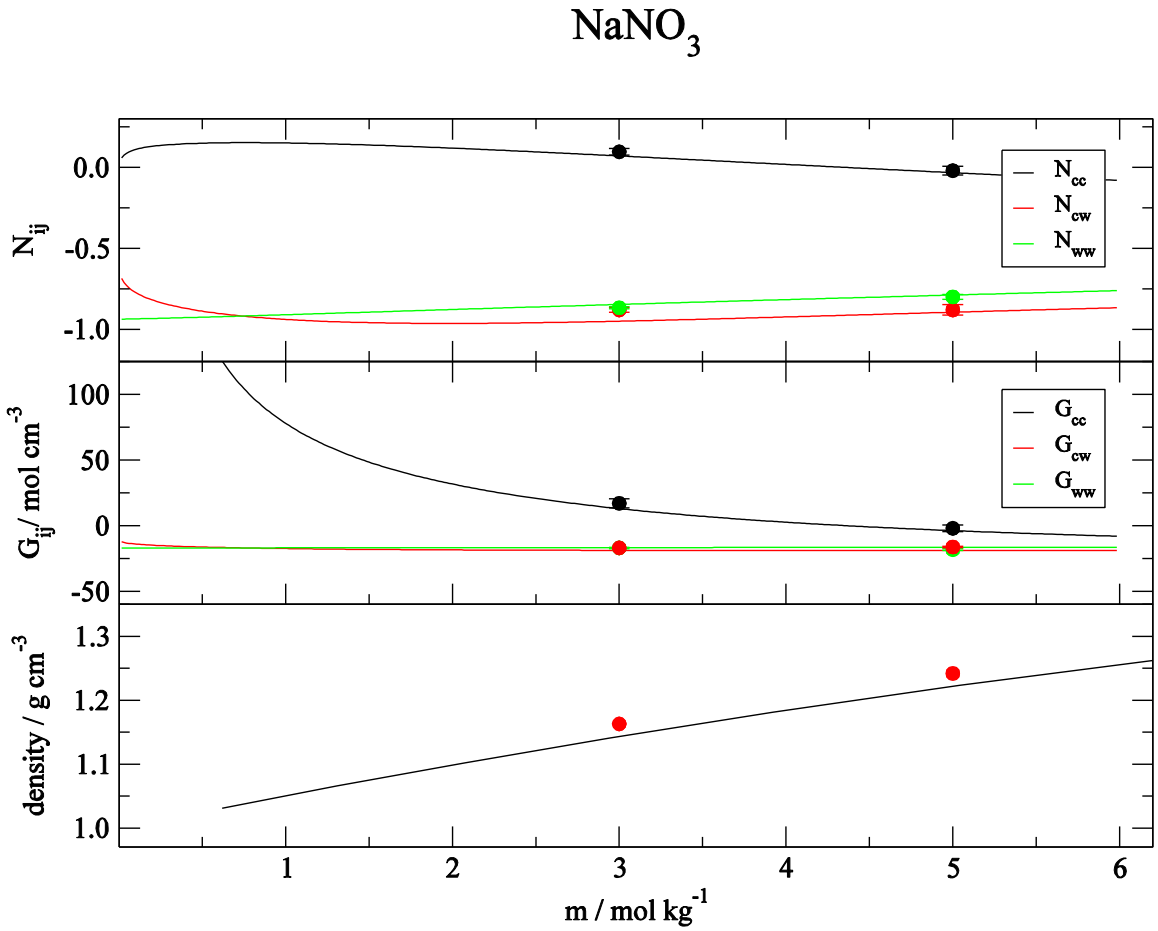

Figure 3.2. The density, KB integrals and the excess coordination number of the aqueous $NaNO_3$ salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

Table 3.1. The optimal charge distribution for the NO_3^- ion for reproducing the experimental KB integrals in the simulations

N	O	O	O
-0.4	-0.2	-0.2	-0.2

The transferability of the NO_3^- ion charge distribution obtained for the sodium salt to the corresponding potassium salts was investigated and the results are displayed in Figure 3.3. Here, the KB values are slightly underestimated for the potassium salts, resulting in a deviation in the excess coordination numbers too. A single charge distribution for the NO_3^- ion which can reproduce the experimental KBI for both sodium and potassium could not be achieved in this study.

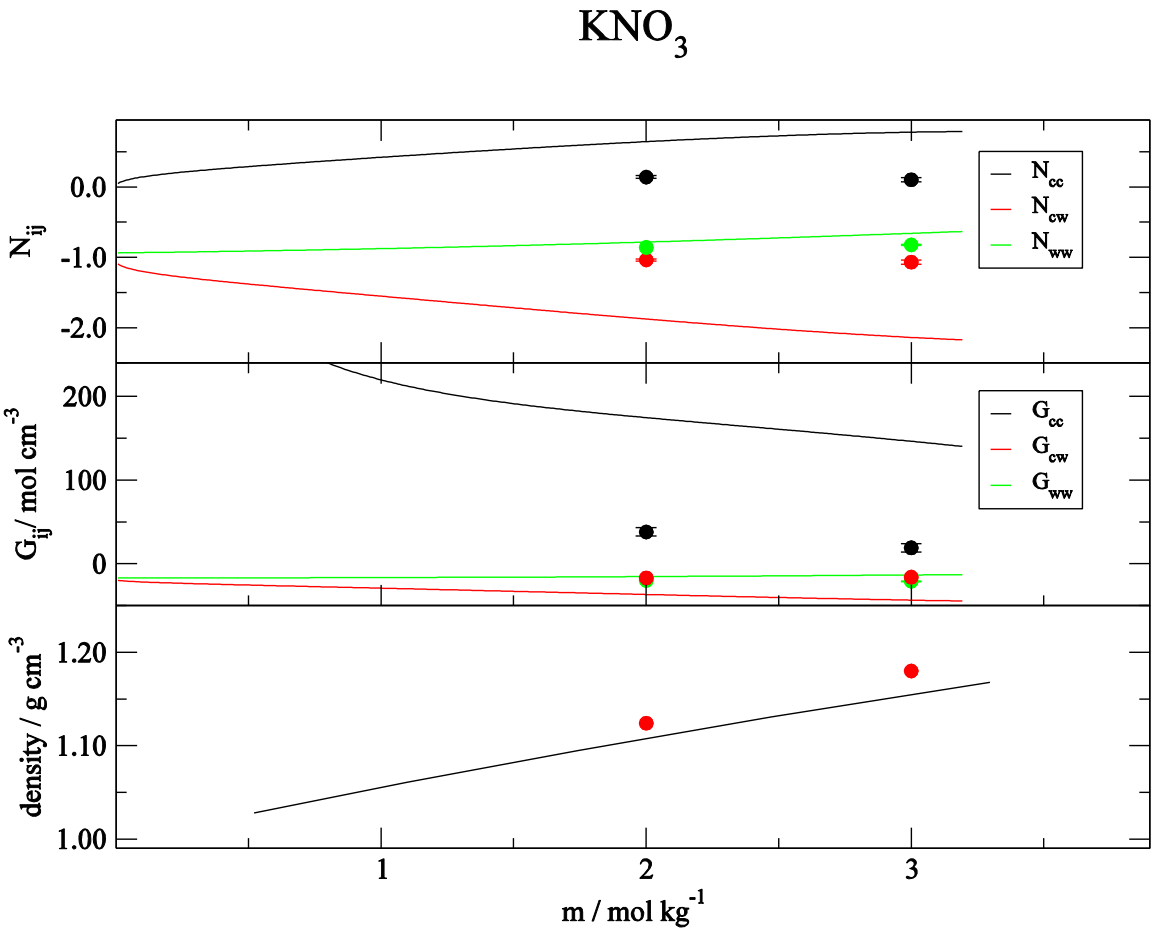

Figure 3.3. The density, KB integrals and the excess coordination number of aqueous KNO_3 salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

The ClO_4^- ion is tetrahedral in shape and has a negative one charge. Table 3.2 displays the best charge distribution found for the ClO_4^- ion. A comparison of MD simulation results and experimental data for the density, KB integrals and the excess coordination numbers for $NaClO_4$ salt solutions are displayed in Figure 3.4. The KB integrals are closer to the experimental values at lower concentrations. However, the KB integrals start to deviate at higher concentrations. The same trend is displayed by the simulated densities too. However, we found that adjustments to the charge distribution alone do not help to bring down the simulated densities to the experimental values.

Table 3.2. The optimal charge distribution for the ClO_4^- ion in reproducing the experimental KB integrals in the simulations

Cl	O	O	O	O
1.4	-0.6	-0.6	-0.6	-0.6

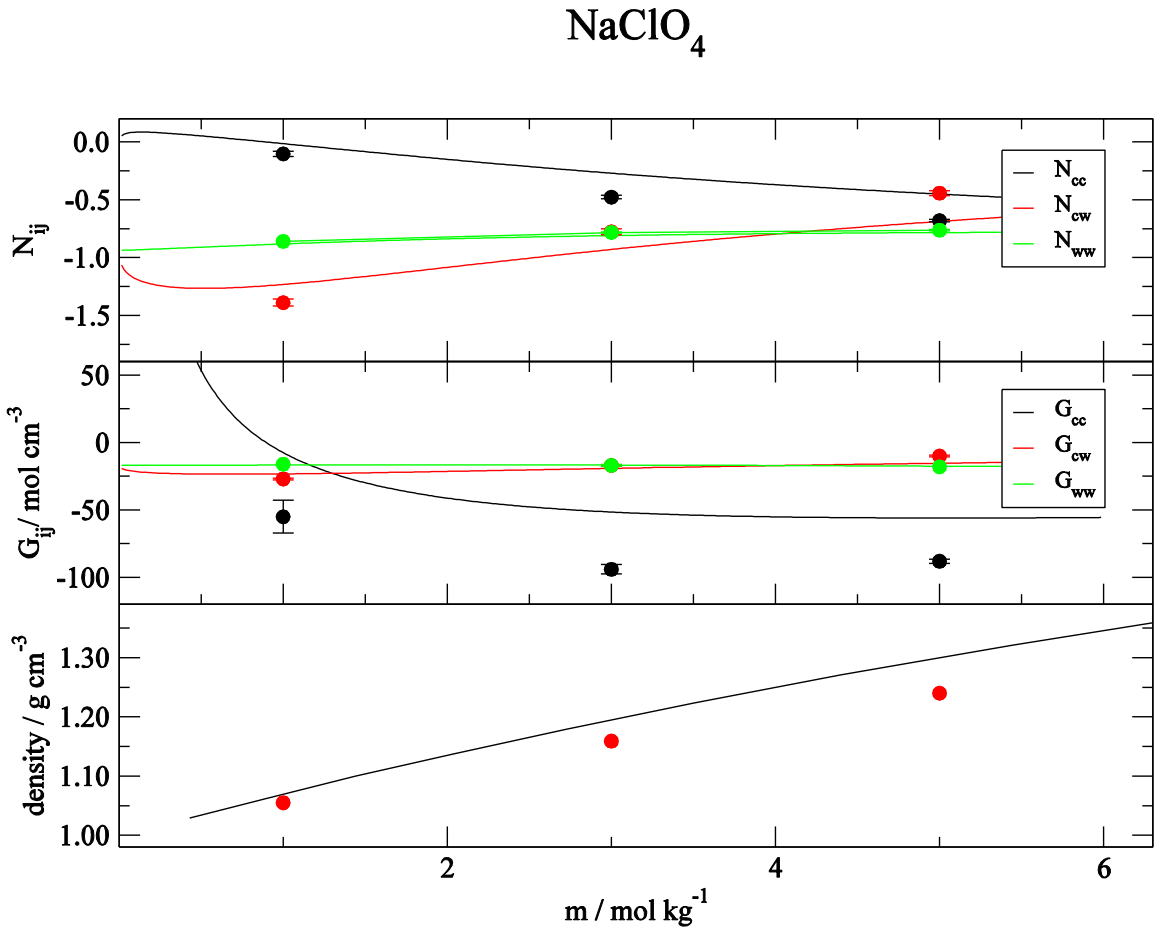

Figure 3.4. The density, KB integrals and the excess coordination number of aqueous $NaClO_4$ salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

The SO_4^{2-} ion is tetrahedral in shape with a negative two charge resulting in a high charge density. The best charge distribution for SO_4^{2-} ion is displayed in Table 3.3. A comparison of the MD simulations and the experimental data for the density, KB integral and the excess coordination number for the aqueous Na_2SO_4 salt solution is displayed in Figure 3.5. However, the KB integral are overestimated in the MD simulations compared to the experimental numbers. This leads to a

deviation in the excess coordination number in the MD simulations. These deviations are more prominent at the lower concentrations.

Table 3.3. The optimal charge distribution for the SO_4^{2-} ion to reproduce the experimental KB integrals in the simulations

S	O	O	O	O
2.4	-1.1	-1.1	-1.1	-1.1

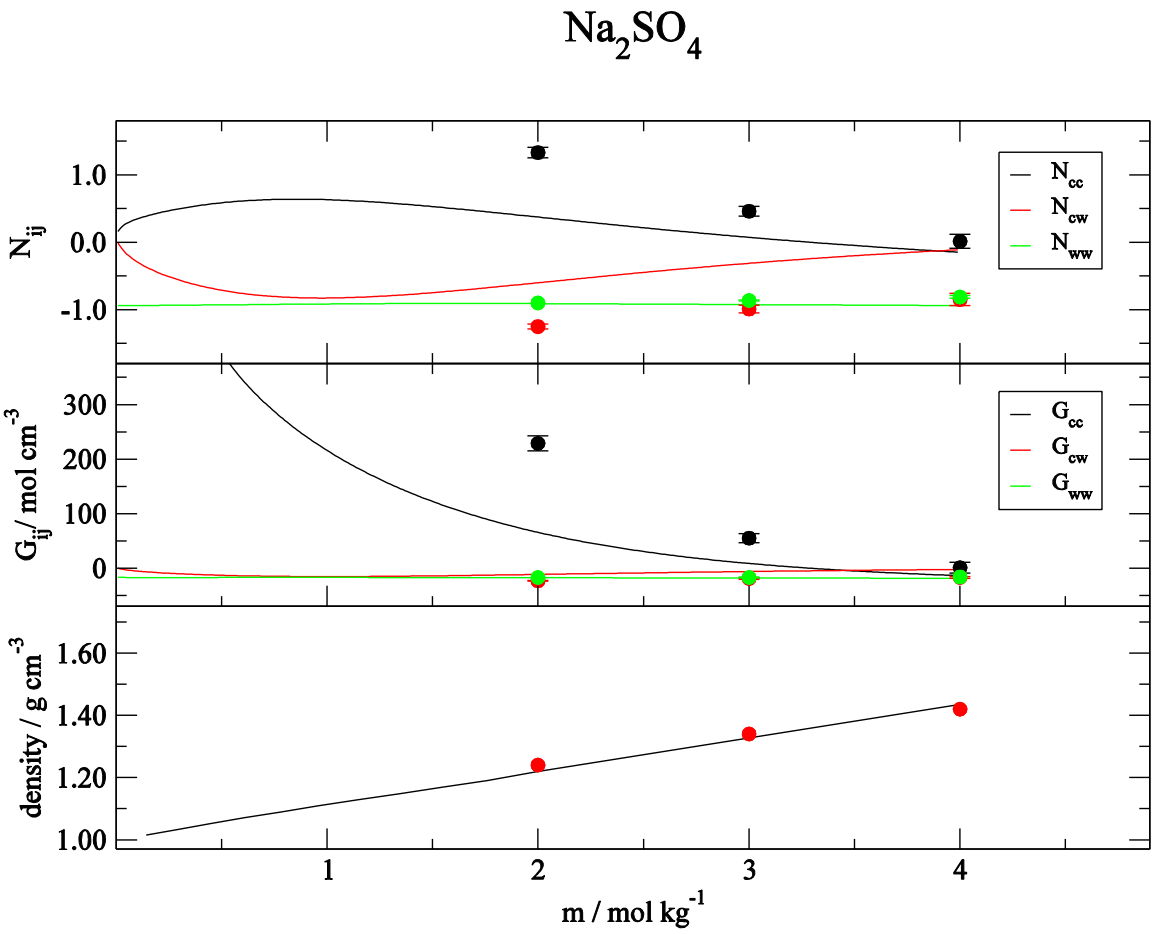

Figure 3.5. The density, KB integrals and the excess coordination number of the aqueous Na_2SO_4 salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

Table 3.4. Two charge distributions for the CO_3^{2-} ion that display unrealistic aggregation behavior in finite soluble concentrations. High values for ion-ion KB integral reflect the aggregation. The units for the KBI are cm^3mol^{-1}.

C	O		SPC/E		
			G_{cc}	G_{cw}	G_{ww}
		Exp	114	-4	-17
1	-1	MD	34654	-973	12
0.4	-0.8	MD	31113	-1035	17

The CO_3^{2-} ion is a trigonal planer shaped molecule with a negative two charge. This makes the CO_3^{2-} ion a highly polarized molecule with a high charge density. The difficulty of modeling ions with a high charge density is very prominent in the MD simulations of aqueous Na_2CO_3 salt solutions. All the charge distributions tried in this study displayed aggregation behavior at the finite concentrations that are lower than the experimental solubility limit. Table 3.4 is displays two of the charge distributions, and resulting KB integrals, investigated in the MD simulations and their comparison with the experimental values. The larger positive G_{cc} values reflect the excessive ion clustering in the MD simulations. Figure 3.6 displays the experimental density, KB integrals, and excess coordination numbers of aqueous Na_2CO_3 salt solutions as a function of salt molality.

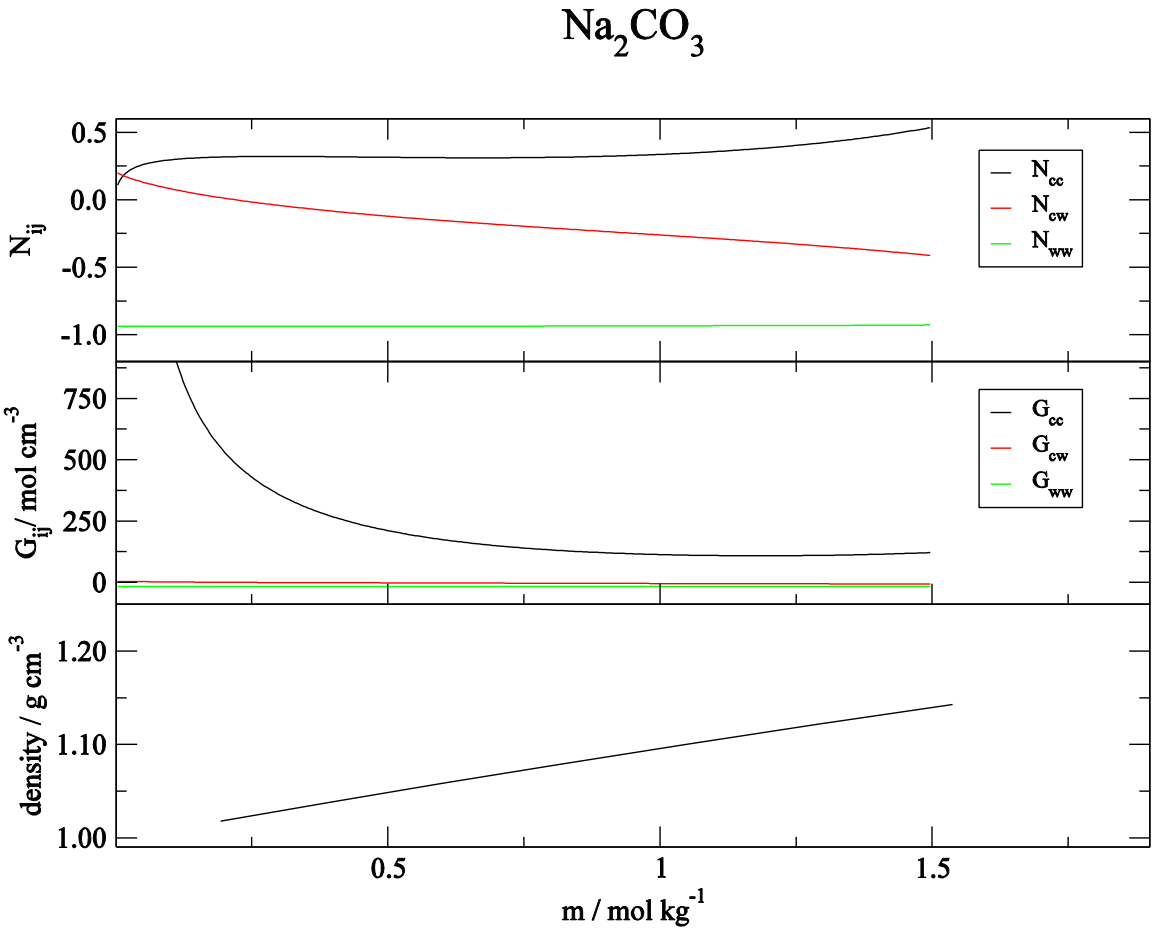

Figure 3.6. The density, KB integrals and the excess coordination number of aqueous Na_2CO_3 salt solutions as a function of salt molality. Lines are from a KB analysis of the experimental data. Simulation results are not displayed as they result in an unrealistic aggregation behavior even in the soluble concentrations.

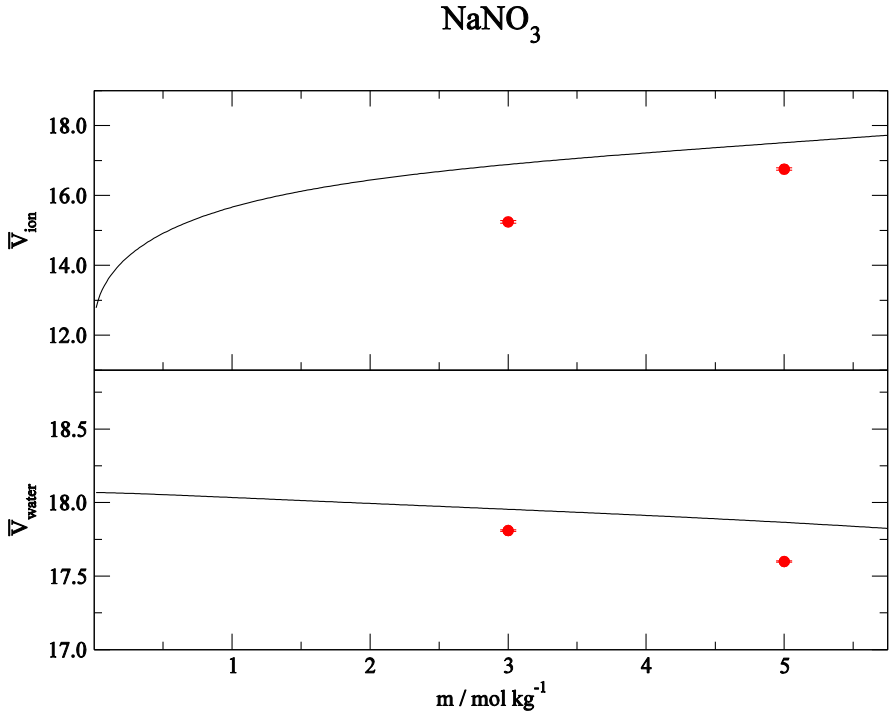

Figure 3.7. Partial molar volumes ($cm^3\ mol^{-1}$) of the ions and water for aqueous $NaNO_3$ salt solutions as a function of the molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

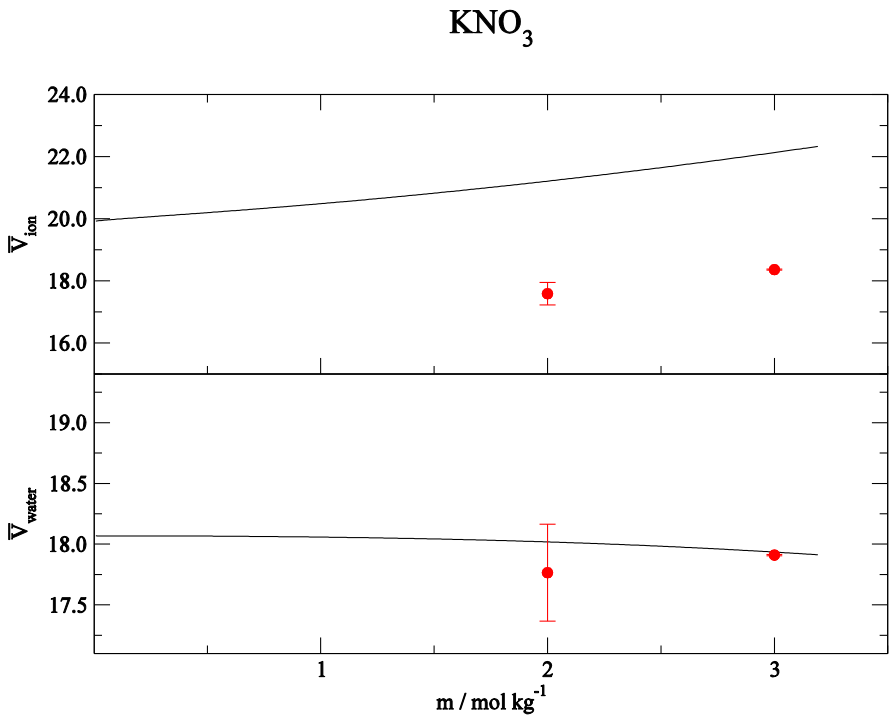

Figure 3.8. . Partial molar volumes ($cm^3\ mol^{-1}$) of the ions and water for aqueous KNO_3 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

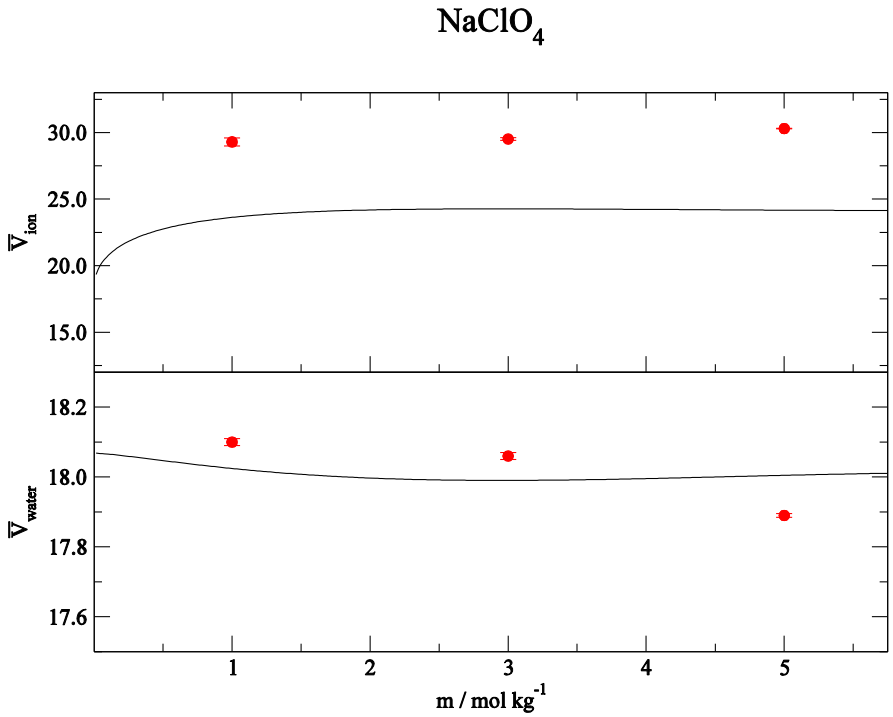

Figure 3.9. Partial molar volumes (cm^3 mol^{-1}) of the ions and water for aqueous $NaClO_4$ salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

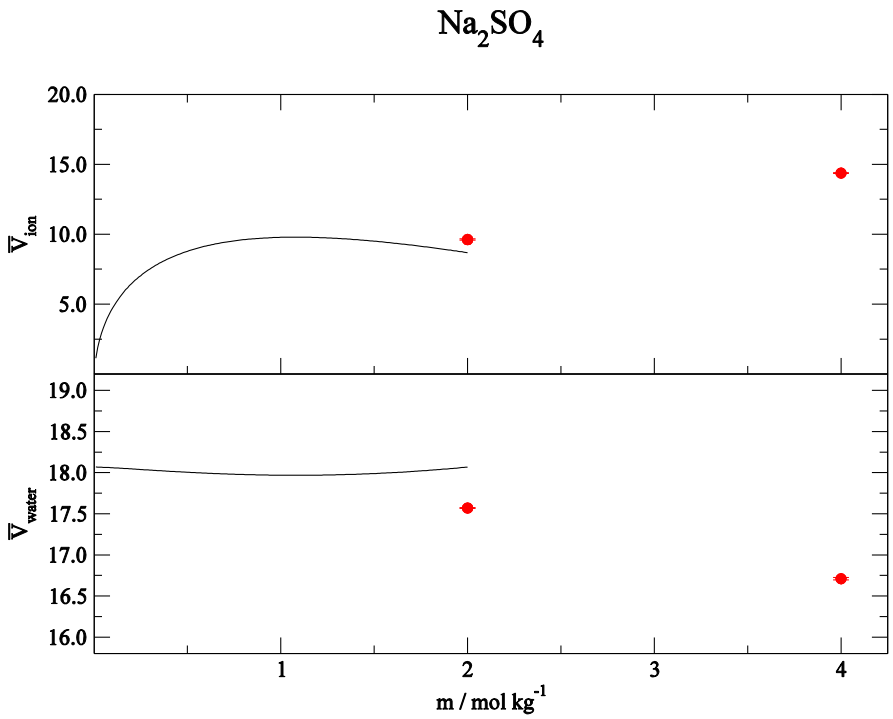

Figure 3.10. Partial molar volumes (cm^3 mol^{-1}) of the ions and water for aqueous Na_2SO_4 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

Displayed in Figure 3.7, Figure 3.8, Figure 3.9 and Figure 3.10 are a comparison of the simulated partial molar volumes of the ions and water molecules for the $NaNO_3$, KNO_3, $NaClO_4$ and Na_2SO_4 salt solutions as a function of the salt molality with the experimental values. Although the $NaNO_3$ salt solutions give partial molar volumes of the ions that are less deviated from the experimental values, the corresponding values for the KNO_3 salt solutions are considerably under estimated. Since the KB integrals are slightly deviated from the experimental values for the $NaClO_4$ salt solutions, the partial molar volumes also display the same trend. However, despite the higher charge density of the ions in the Na_2SO_4 salt solutions, the experimental partial molar volumes are well reproduced at lower concentrations. However, at higher concentrations of the Na_2SO_4 salt solutions the partial molar volumes of the ions are overestimated.

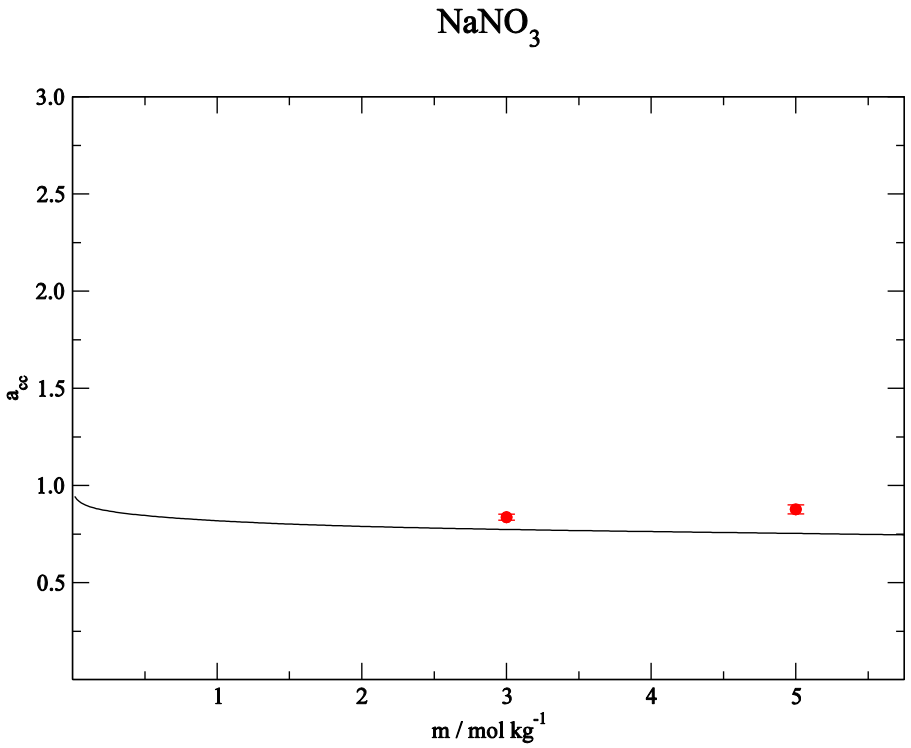

Figure 3.11. Activity derivatives of the aqueous $NaNO_3$ salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

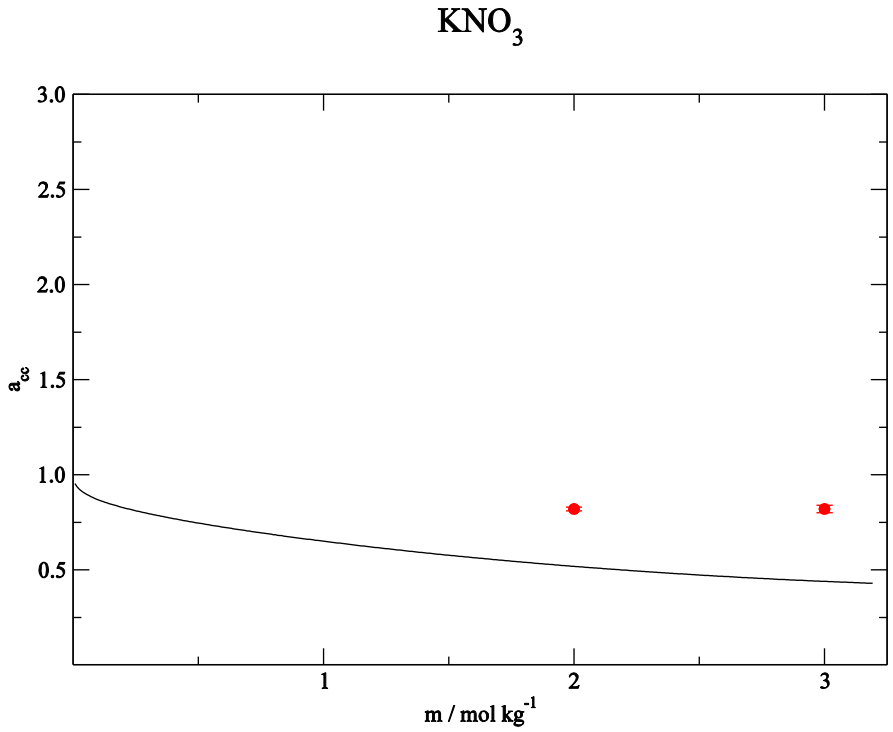

Figure 3.12. Activity derivatives of the aqueous KNO_3 salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

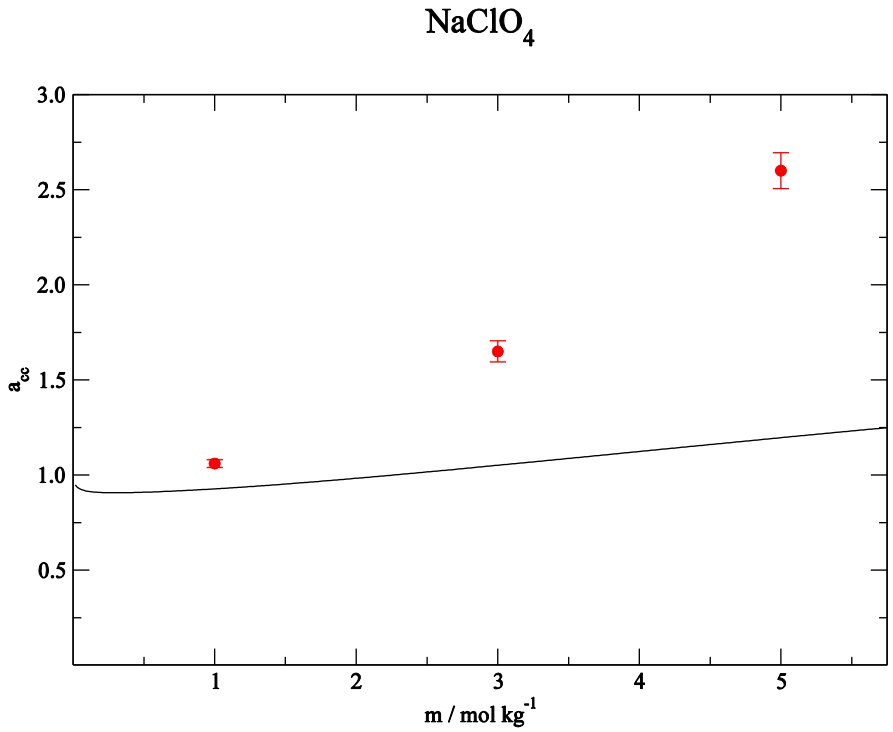

Figure 3.13. Activity derivatives of the aqueous $NaClO_4$ salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

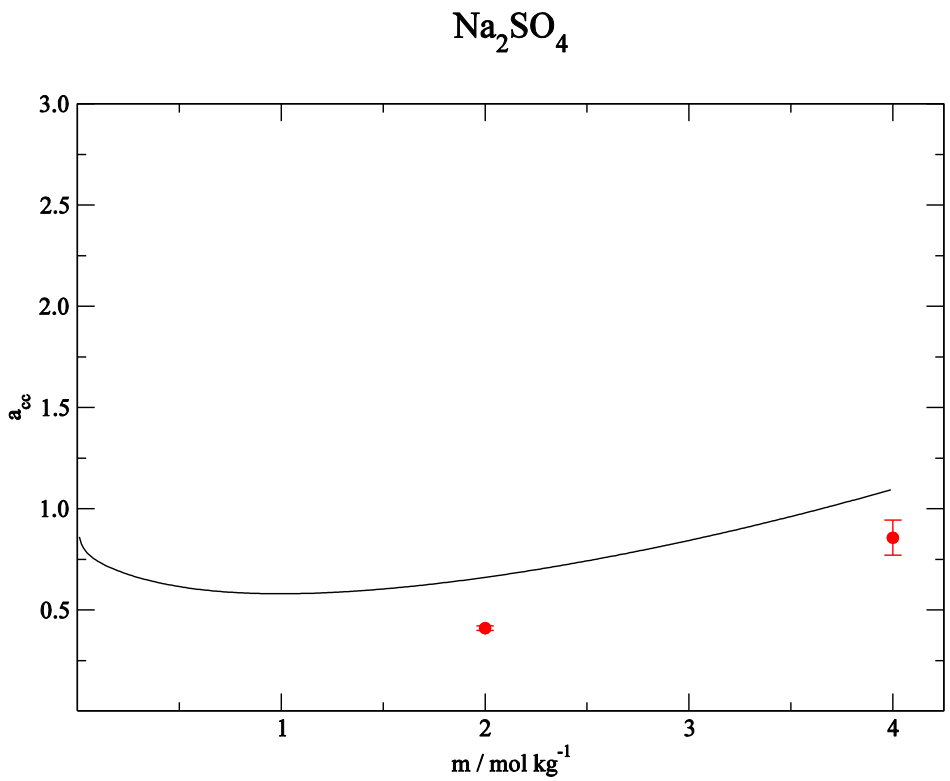

Figure 3.14. Activity derivatives of the aqueous Na_2SO_4 salt solutions as a function of salt molality. Lines are from the KB analysis of the experimental data and symbols are for the MD simulations.

In Figure 3.11, Figure 3.12, Figure 3.13 and Figure 3.14 we display the activity derivatives for aqueous $NaNO_3$, KNO_3, $NaClO_4$ and Na_2SO_4 salt solutions as a function of salt molality. Since the KNO_3 model could not reproduce the experimental KB integrals in the MD simulations, it cannot not reproduce the experimental activity derivatives too. However, the $NaNO_3$ salt solutions did reproduce the experimental activity derivatives successfully. For the $NaClO_4$ salt solutions the activity derivatives are closer to the experimental values at lower concentrations. However, increasing salt concentration leads to significant deviations in the activity derivatives from the experimental values. This could be the result from higher salt concentrations leading to excessive

ion interactions in the simulation box. The Na_2SO_4 activity derivatives are under estimated in the MD simulations, but at high concentrations the values are closer to the experimental values.

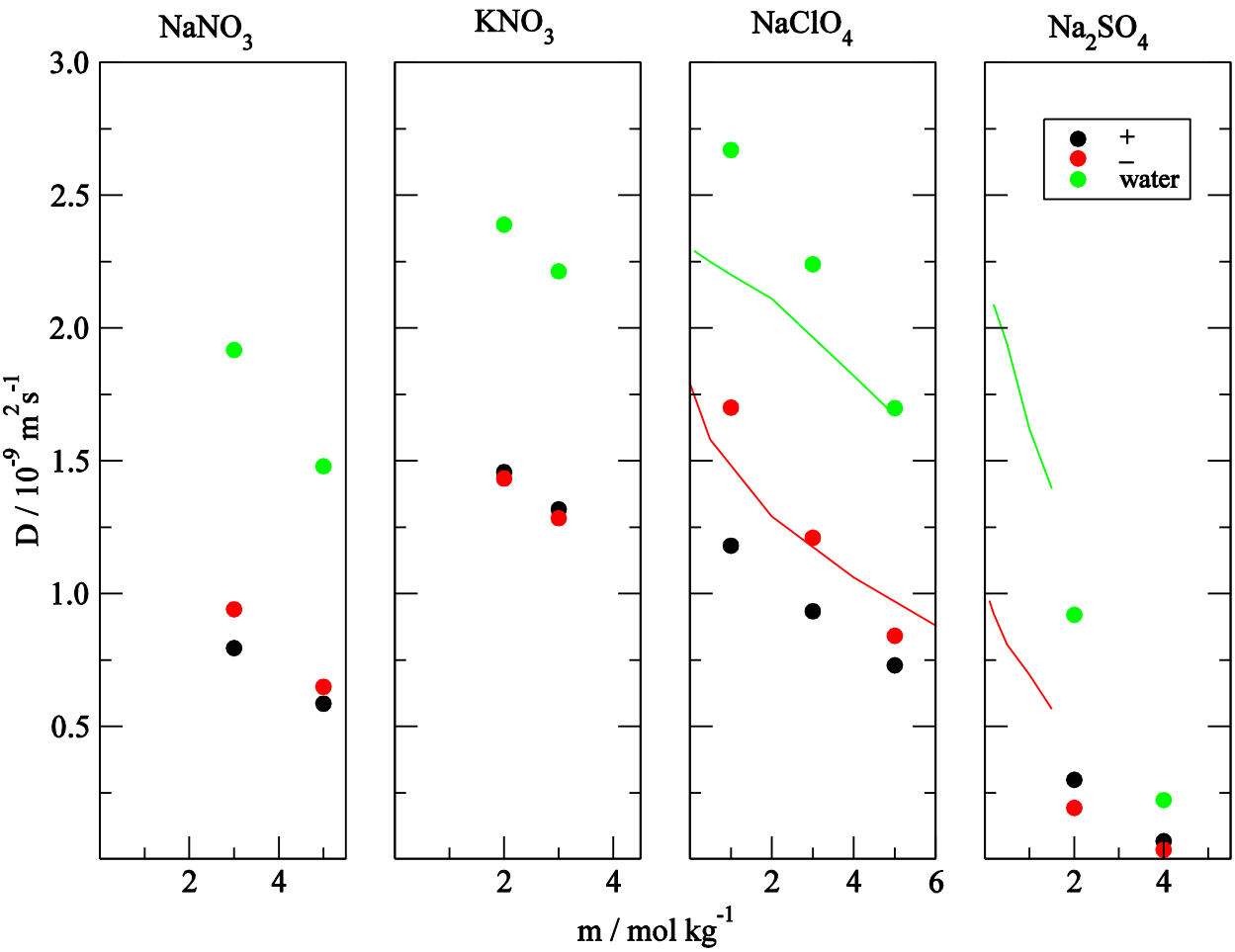

Figure 3.15. The diffusion coefficients of the NaNO3, KNO3, NaClO4 and Na2SO4 salt solution as a function of salt molality. Lines are for the experimental data[48-49] and symbols are for the MD simulations.

All the properties displayed in the previous figures are related to KB integrals. Hence, always in the KBFF development we have tried to investigate the reproducibility of some other properties which are not related to KB integrals, such as the diffusion coefficients and dielectric

constants. In Figure 3.15 we displays a comparison of the experimental and simulated diffusion coefficients for the $NaNO_3$, KNO_3, $NaClO_4$ and Na_2SO_4 salt solutions. The water displayed the highest diffusion coefficient. Increasing the salt concentration results in a decrease in the water and ion diffusion coefficients. Furthermore, with increasing salt concentrations the anion and cation diffusion coefficients become closer in value. For the NO_3^- ion salt systems the experimental diffusion coefficient information not available. In the ClO_4^- ion systems the diffusion of water and anion is close to the experimental values. Surprisingly, despite the high charge density, the diffusion coefficients in the SO_4^{2-} systems appear to follow the experimental trends.

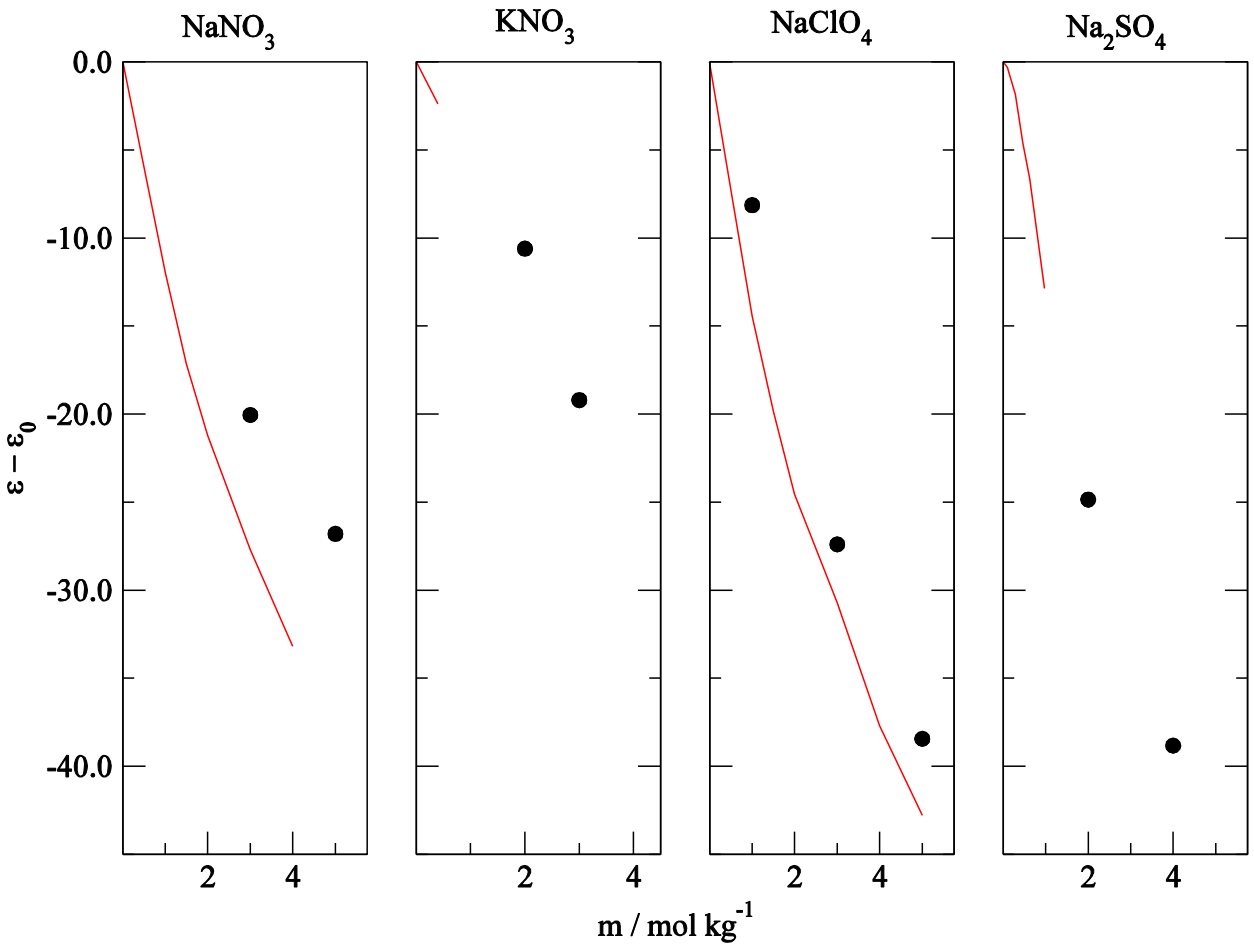

Figure 3.16. The dielectric decrements of the $NaNO_3$, KNO_3, $NaClO_4$ and Na_2SO_4 salt solutions as a function of salt molality. Lines are for the experimental data[50-51] and symbols are for the MD simulations.

Figure 3.16 displays the experimental and simulated dielectric decrements for the $NaNO_3$, KNO_3, $NaClO_4$ and Na_2SO_4 salt solutions. Despite the correct reproducibility of the KB integrals in the $NaNO_3$ salt solutions, the dielectric decrement is underestimated in the MD simulations. For the $NaClO_4$ salt solutions, the experimental and MD simulations dielectric decrement values are very similar. The Na_2SO_4 salt solution dielectric decrement follows the experimental trends in the simulations.

3.5. Conclusions

In this study the parameterization and validation of KBFF models for aqueous $NaNO_3$, KNO_3, $NaClO_4$, Na_2SO_4 and Na_2CO_3 salt solutions was attempted. However, aqueous Na_2CO_3 salt solutions displayed unrealistic aggregation behavior at all concentrations. All other salt solutions were parameterized to reproduce the experimental KB integrals. The charge distribution for the NO_3^- ions, which reproduce the experimental KBIs for aqueous $NaNO_3$, showed reasonable but not perfect transferability to aqueous KNO_3 salt solutions. Aqueous $NaClO_4$ salt solution results were close to the experimental values at lower concentrations, but started to deviated from the experimental values at higher concentrations. The aqueous Na_2SO_4 models provided reasonable, although slightly overestimated, KBI values without showing any aggregation in the MD simulations. However, aqueous Na_2CO_3 models were not successful because of the unrealistic aggregation behavior in the MD simulations. This is probably a result of the higher charge density of the small trigonal planner CO_3^{2-} ion, compared to the larger tetrahedral SO_4^{2-} ion.

The KBFF models should be viewed as providing a reasonable balance between solute-solute, solute-solvent, and solvent-solvent interactions, as inferred by their resulting distributions. The compatibility of parameters for different types of molecules in a given force field is achieved by employing the same approach in the parametrization process. Hence, the extending compatible parameters for oxo anions in the KBFF enables to study solute activities and co-solvent interactions with biomolecules in MD simulations.

References

1. Hofmeister, F., *Archiv für Experimentelle Pathologie und Pharmakologie* **1888,** *24*, 247-260.

2. Collins, K. D., *Methods* **2004,** *34* (3), 300-311.

3. Perez-Jimenez, R.; Godoy-Ruiz, R.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M., *Biophysical Journal* **2004,** *86*, 2414-2429.

4. Vrbka, L.; Jungwirth, P.; Bauduin, P.; Touraud, D.; Kunz, W., *J. Phys. Chem. B* **2006,** *110*, 7036-7043.

5. Latimer, W. M.; Pitzer, K. S.; Slansky, C. M., *The Journal of Chemical Physics* **1939,** *7*, 108.

6. Hummer, G.; Pratt, L. R.; García, A. E., *The Journal of Physical Chemistry* **1996,** *100*, 1206-1215.

7. Eaves, J. D.; Loparo, J. J.; Fecko, C. J.; Roberts, S. T.; Tokmakoff, A.; Geissler, P. L., *Proceedings of the National Academy of Sciences of the United States of America* **2005,** *102*, 13019-22.

8. Rezus, Y. L. A.; Bakker, H. J., *The Journal of Chemical Physics* **2005,** *123*, 114502.

9. Eklund, L.; Hofer, T. S.; Persson, I., *Dalton Trans.* **2015,** *44*, 1816-1828.

10. Eklund, L.; Hofer, T. S.; Persson, I., *Dalton transactions (Cambridge, England : 2003)* **2014,** *44*, 1816-28.

11. Bergstrom, P.-A.; Lindgren, J.; Kristiansson, O., *Journal of physical chemistry 95*, 8575-8580.

12. Walrafen, G. E., *The Journal of Chemical Physics* **1970,** *52* (8), 4176-4198.

13. Pribil, A. B.; Hofer, T. S.; Vchirawongkwin, V.; Randolf, B. R.; Rode, B. M., *Chemical Physics* **2008,** *346*, 182-185.

14. Thøgersen, J.; Réhault, J.; Odelius, M.; Ogden, T.; Jena, N. K.; Jensen, S. J. K.; Keiding, S. R.; Helbing, J., *The Journal of Physical Chemistry B* **2013,** *117*, 3376-3388.

15. Smith, J. W.; Lam, R. K.; Shih, O.; Rizzuto, A. M.; Prendergast, D.; Saykally, R. J., *The Journal of Chemical Physics* **2015,** *143*, 084503.

16. Ebner, C.; Sansone, R.; Hengrasmee, S.; Probst, M., *International Journal of Quantum Chemistry* **1999,** *75*, 805-814.

17. Salvador, P.; Curtis, J. E.; Tobias, D. J.; Jungwirth, P., *Physical Chemistry Chemical Physics* **2003,** *5*, 3752.

18. Pye, C. C.; Rudolph, W. W., *The Journal of Physical Chemistry A* **2003,** *107*, 8746-8755.

19. Banerjee, P.; Subramanian, Y.; Bagchi, B., *ArXiv e-prints* **2016,** *1*.

20. Otten, D. E.; Onorato, R.; Michaels, R.; Goodknight, J.; Saykally, R. J., *Chemical Physics Letters* **2012,** *519*, 45-48.

21. Marcus, Y., *J. Chem. Soc., Faraday Trans* **1991,** *87*, 2995-2999.

22. Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P., *Biochemistry* **2008,** *47*, 12065-12077.

23. Chialvo, A. A.; Gruszkiewicz, M. S.; Cole, D. R., *Journal of Chemical & Engineering Data* **2010,** *55*, 1828-1836.

24. Jungwirth, P.; Curtis, J. E.; Tobias, D. J., Polarizability and aqueous solvation of the sulfate dianion. In *Chemical Physics Letters*, 2003; Vol. 367, pp 704-710.

25. Richard Buchner; Capewell, S. G.; Glenn Hefter; May, P. M., *J. Phys. Chem. B* **1999,** *103* (7), 1185-1192.

26. Daly, F. P.; Brown, C. W.; Kester, D. R., *The Journal of Physical Chemistry* **1972,** *76*, 3664-3668.

27. Lowenstam, H., *Science* **1981,** *211*, 1126-1131.

28. Lackner, K. S.; Wendt, C. H.; Butt, D. P.; Joyce, E. L.; Sharp, D. H., *Energy* **1995,** *20*, 1153-1170.

29. Takahashi, T.; Sutherland, S. C.; Sweeney, C.; Poisson, A.; Metzl, N.; Tilbrook, B.; Bates, N.; Wanninkhof, R.; Feely, R. A.; Sabine, C.; Olafsson, J.; Nojiri, Y., *Deep Sea Research Part II: Topical Studies in Oceanography* **2002,** *49*, 1601-1622.

30. Mason, P. E.; Wernersson, E.; Jungwirth, P., *Journal of Physical Chemistry B* **2012,** *116*, 8145-8153.

31. Yuen, P. S.; Lister, M. W.; Nyburg, S. C., *The Journal of Chemical Physics* **1978,** *68*, 1936-1941.

32. Pavese, A.; Catti, M.; Parker, S. C.; Wall, A., *Physics and Chemistry of Minerals* **1996,** *23*, 89-93.

33. Thackeray, D. J.; Siders, P. D., *Journal of the Chemical Society, Faraday Transactions* **1998,** *94*, 2653-2661.

34. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67*, 4884.

35. Smith, P. E., *The Journal of chemical physics* **2008,** *129*, 124509.

36. Pitzer, K. S., *The Journal of Physical Chemistry* **1973,** *77*, 268-277.

37. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19*, 774.

38. Söhnel, O.; Novotný, P., *Densities of aqueous solutions of inorganic substances*. Elsevier: Amsterdam ;, 1985; p 335.

39. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Computer Physics Communications* **1995,** *91*, 43-56.

40. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

41. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81*, 3684.

42. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

43. Darden, T.; York, D.; Pedersen, L., *The Journal of Chemical Physics* **1993,** *98*, 10089.

44. Chandrasekhar, S., *Reviews of Modern Physics* **1943,** *15* (1), 1-89.

45. Chitra, R.; Smith, P. E., *The Journal of Physical Chemistry B* **2000,** *104* (24), 5854-5864.

46. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: USA, 1987.

47. Smith, P. E.; Gunsteren, W. F. v., *The Journal of Chemical Physics* **1994,** *100* (1), 577-585.

48. Tanaka, K., *Journal of the Chemical Society, Faraday Transactions 1: Physical Chemistry in Condensed Phases* **1988,** *84*, 2895.

49. Heil, S. R.; Holz, M.; Kastner, T. M.; Weingartner, H., *Journal of the Chemical Society, Faraday Transactions* **1995,** *91*, 1877.

50. Barthel J.; Buchner R.; Münsterer M., *Electrolyte Data Collection, Part 2: Dielectric Properties of Water and Aqueous Electrolyte Solutions, .* Wiley-VCH Verlag GmbH & Co. KGaA: Frankfurt, 1996; Vol. 100, p 1922-1922.

51. Feng, L.; Shu, J.; Yang, X.-Q., *Asian Journal of Chemistry* **2013,** *25*, 7653-7654.

Chapter 4 - Molecular Dynamics Simulations of Aqueous Phosphate Solutions

4.1. Abstract

The different types of phosphate ions, such as $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}, and $(CH_3)_2PO_4^-$, are crucial for many important physiological processes. The $H_2PO_4^-$, HPO_4^{2-} and , PO_4^{3-} ions are included in the Hofmeister series which is used to help explain important biological processes. The $(CH_3)_2PO_4^-$ ion is an important structural moiety in the phospholipids and in some post translational modified proteins. Computer simulations of these ions can be used to assist experimentalists to gain insight in to molecular level mechanism of these ions. Here we attempted to develop classical non polarizable force fields for the $H_2PO_4^-$, HPO_4^{2-}, PO_4^{3-}, and $(CH_3)_2PO_4^-$, ions which are compatible with the SPC/E water model. These new force field parameters are specifically developed to reproduce the experimental Kirkwood-Buff integrals. Attempts to develop force fields for PO_4^{3-}, and HPO_4^{2-} ions, which have a relatively high charge density, were unsuccessful due to an excessive aggregation behavior in the simulations.

4.2. Introduction

The oxo anions of phosphorous play a very important structural and functional role in many biological systems.[1] A good example is the ATP molecule which is the fundamental species for energy storage and transfer in a living cell. Phosphate ions are also used as structural components in bone, teeth, DNA and RNA molecules. Furthermore, the phosphate ion moiety acts as one of the most important post translational modifications to protein side chains, and also an integral structural part of lipids. The dihydrogen phosphate (DHP) ion plays an important role as as

intermediate species in the signal transduction process.[2] These hydrogen phosphate charge groups are claimed to act as hydration sites in macromolecules.[3] A monomeric phosphate ion, which can switch between multiple negative charges, possesses unique characteristics that enable phosphorous oxygen ions to play a very important role as an intermediate species in many biological processes.

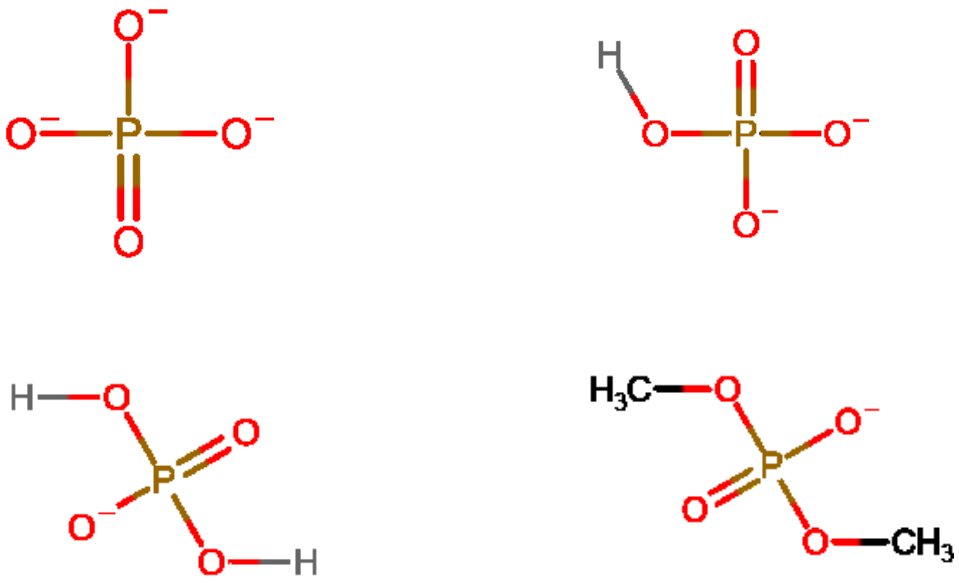

Figure 4.1. The wireframe representation of the phosphate ion (top left), monohydrogenphosphate ion (MHP) (top right), dihydrogenphosphate ion (DHP) (bottom left) and dimethylphosphate ion (DMP) (bottom right)

Despite the importance of phosphate derived ions, it is challenging task to model these types of ions, with high charge densities, in the MD simulations. Unfortunately, standard force

field parameters for these ions are observed to overestimate cation and anion interactions resulting in excessive ion pair formation.[4] This adversely affects the osmotic pressure and other chemical properties of the salt solutions in the MD simulations. Most of the force fields fail to reproduce correct thermodynamic properties in MD simulations and one of the most recent publications appeared in 2017 by Margreitter et al. and reports an updated version of the parameters for the GROMOS force field for the dihydrogenphosphates (DHP), methyl phosphates, dimethylphosphates (DMP) and phenylphosphate ions to reproduce correct experimental solvation free energies.[5]

The possible high negative charge of phosphate ions, and the incompatibility of the standard water models, leads excessive ion clustering in the MD simulations. As an recent example, a MD study in 2016 which investigated the phosphate ion interactions with calcium ions reported that the CHARMM and AMBER force fields form artificial clusters in the MD simulations.[6] Phosphate ions are one of the building blocks of DNA, RNA and phospholipids and many other molecules that play important roles in the biological systems. There are still ongoing studies to understand the nature of the first formed mineral phase of bone and some studies claim it is amorphous calcium phosphate.[7] There are molecular mechanics simulations

There are many experimental spectroscopic studies concerning DHP and MHP ions because of their biological importance.[8-10] Also, there are first principal molecular dynamics simulations of phosphate, MHP and DHP ions to help interpret experimental spectral features in the literature.[11-12] One of the early MD study investigated the solvation of sodium DHP salt solutions as reported in 1982.[13]

The DMP molecule is a model compound for the phosphodiester linkage in phospholipids and in nucleic acids. There are ab initio studies which suggest the geometry and the electronic

structure of the DMP molecule are significantly different in the gas phase and in bulk water.[14-15] Furthermore, computer simulations and experimental studies have shown that different conformers of the DMP molecule are stabilized depending on the solvation effects in the aqueous medium.[16-18] Computer simulations have shown that water molecules prefer anionic oxygen over ester oxygens in the DMP molecule.[16, 19] There is a MD study in 2012 which used the DMP molecule to investigate the interactions of phosphate backbone of DNA molecules with surrounding counter ions in solution.[4] This work claims that the standard CHARMM and AMBER force fields parameters lead to cation to DNA phosphate structural artifacts, and also fail to reproduce experimental osmotic pressure. There are molecular dynamics studies to compare the stability and solvation behavior of the sodium DMP and sodium methylphosphate ions using potential of mean force calculations.[20] Therefore, there is a clear demand to develop a force field for phosphate ions which can reproduce experimental properties in aqueous solutions for future use in biomolecular simulations.

4.3. Experimental Analysis and MD simulations

The same strategy explained in the second chapters, the Kirkwood Buff inversion procedure, was used to extract the experimental Kirkwood Buff integral (KBI).[21-22] Hence, to obtain experimental KBI values the experimental composition dependent activity and the experimental density data were used after fitting to the Pitzer equation and a polynomial equation, respectively,[23] while for the DMP salts the composition dependent osmotic coefficient was used.[24-25] The experimental partial molar volumes are obtained from the experimental densities using previously established standard approaches.[26-27] The optimization of the partial charge distribution

of the anion to reproduce the experimental KB integral of the respective salt is (again) the main strategy used in this project.

All the molecular dynamics simulations were performed with Gromacs[28] 4.6 software. The SPC/E water[29] used to model water in the simulations. All the ions were modeled using KBFF force field for NaDHP, KDHP Na_2MHP, Na_3PO_4, NaDMP, LiDMP and KDMP systems. All simulations in this study were in the isothermal isobaric ensemble of 300 K temperature and 1 atm pressure.[30] Here, temperature and pressure are weakly coupled to a bath with relaxation time of 0.1 ps and 0.5 ps, respectively. A 2 fs timestep was used to integrate the equation of motion. All the bonds in the simulations were constrained with LINCS algorithm.[31] To calculate electrostatic interactions, the particle mesh Ewald technique[32] was used with 1 nm cut off distance for real space calculations, and a 1.5 nm cutoff was used for the Van der Waals interactions. A random initial configuration of ions and water in a cubic box with length of 10 nm was generated to provide a known concentration of aqueous salt solution using a custom written Fortran code. An energy minimization using the steepest descend method followed by 2 ns of equilibration was performed before the production run. The production run of 15 ns was used to calculate ensemble averages. In the production run configurations were saved for analysis every 0.1 ps.

A custom written Fortran program was used to calculate the center of mass radial distribution function and to calculate the simulated KB integrals. The simulated KBI integrals were then use to calculate the activity derivatives, excess coordination numbers and partial molar volumes. Finally, translational self-diffusion coefficients were calculated using the mean square fluctuation approach,[33-34] and dielectric coefficients were calculated by analysis of the mean dipole moment fluctuations.[35-36]

4.4. Results and Discussion

4.4.1. Dihydrogenphophate (DHP) ion

4.4.1.1. NaDHP

The simulations of aqueous NaDHP salts solutions behaved normally and did not display any unrealistic aggregation behavior as indicated in Figure 4.2. The charge density, resulting in a net negative one charge on the DHP ions, did not seem to lead to excessive ion clusters. Furthermore, from all the charge distributions attempted for the DHP ion, the charge distribution displayed in Table 4.1 provided the best experimental KB integrals for the NaDHP simulations.

The density, KB integrals and the excess coordination numbers calculated for the NaDHP salt solution with the charge distribution displayed in the Table 4.1 are compared with the experimental values in the Figure 4.3. This shows the ability of the chosen charge distribution to reproduce the experimental numbers, with reasonable accuracy, in simulations of aqueous NaDHP salt solutions at both lower and higher concentration.

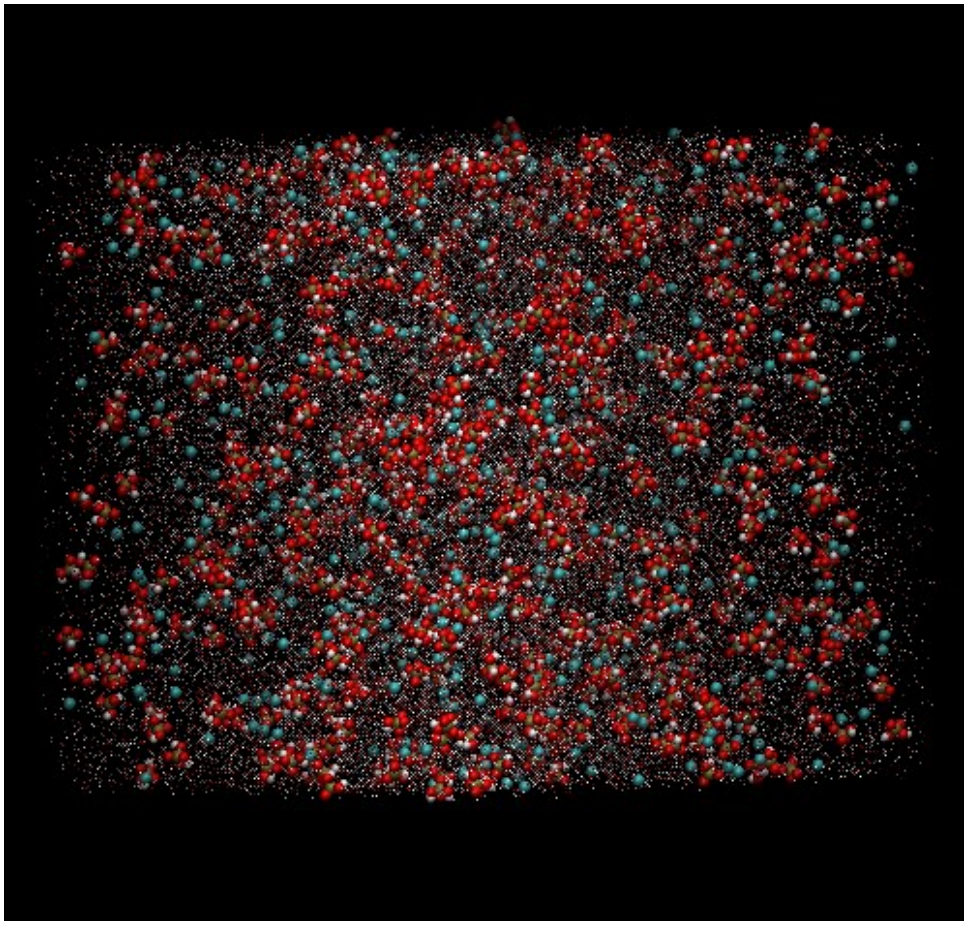

Figure 4.2. The 1 m NaDHP salt solution aggregation with the standard SPC/E water model. Sodium ions are colored in cyan color and oxygen atoms of the DHP are displayed in red color. The charge distribution for the DHP is given in Table 4.1.

Table 4.1. The charge distribution that best reproduced the experimental KBIs for the DHP ion in MD simulations of aqueous NaDHP salt solutions.

H	O	P	O	O	O	H
0.4	-0.63	0.75	-0.645	-0.645	-0.63	0.4

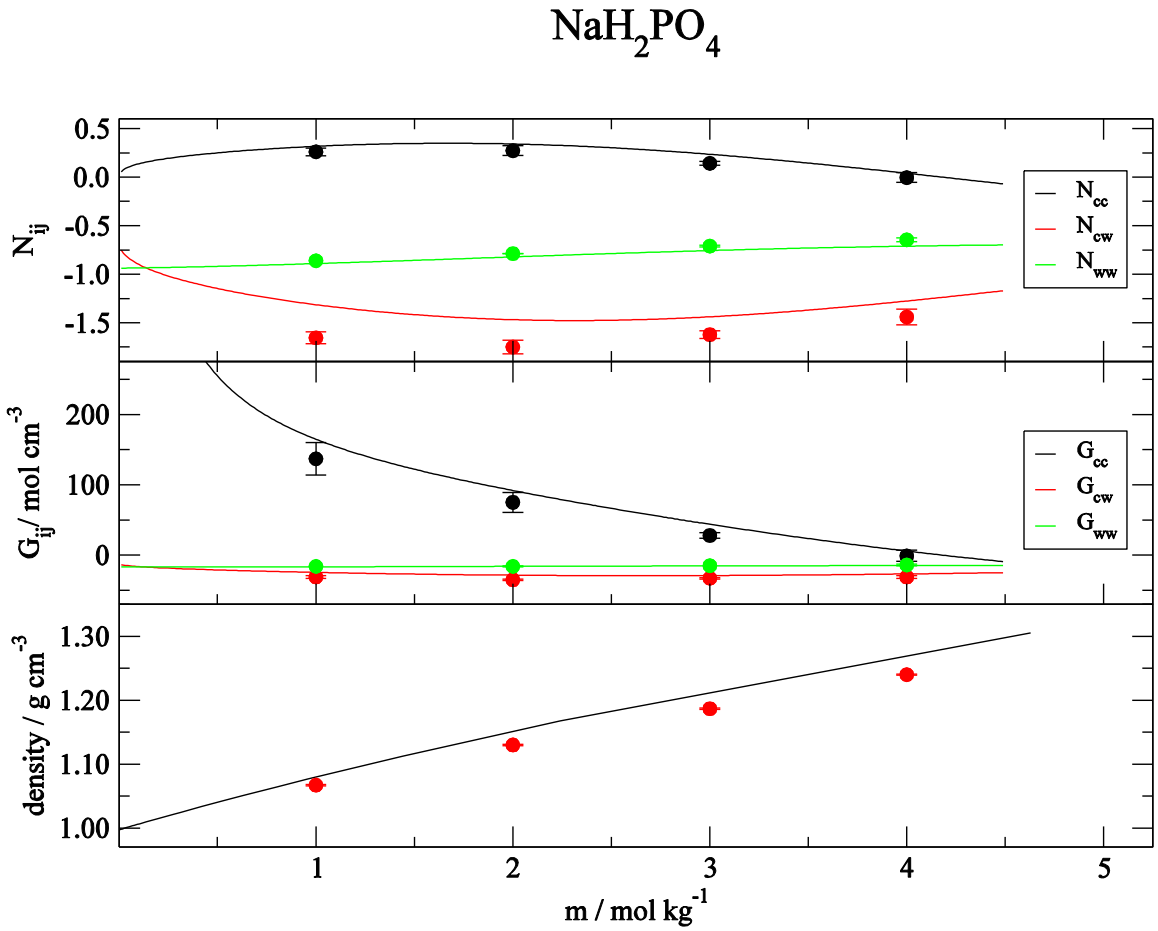

Figure 4.3. The density, KB integrals and the excess coordination number for aqueous NaH_2PO_4 salt solutions as a function of salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

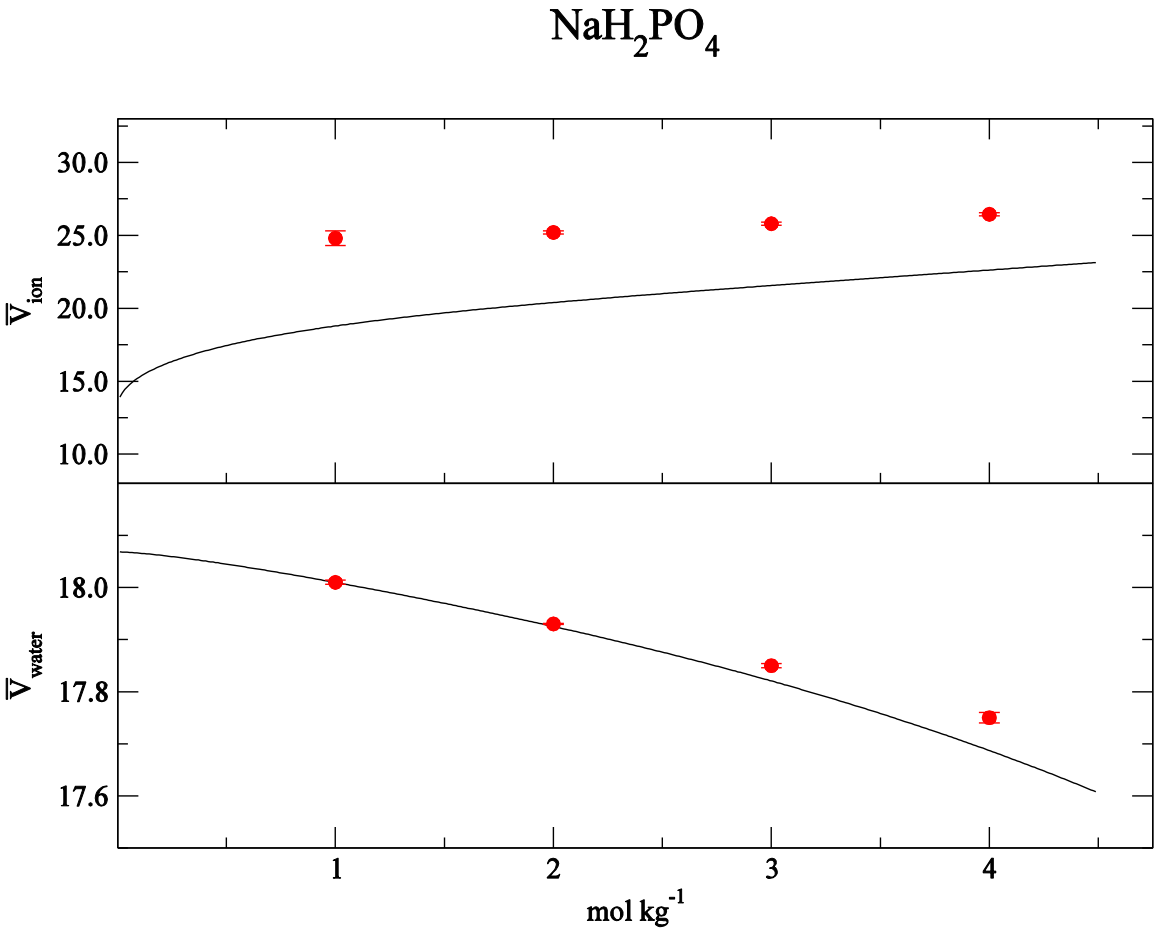

Figure 4.4. Partial molar volumes ($cm^3\ mol^{-1}$) of the ions and water from aqueous NaH_2PO_4 salt solutions as a function of salt molality. Lines are obtained from the KB analysis of the experimental data and symbols are for the MD simulations.

The partial molar volumes of the ions and water obtained from the MD simulations are compared with the experimental values in Figure 4.4. The cation and anion are considered as indistinguishable in the KB analysis,[21-22] and the simulated partial molar volume is calculated as an average value of the individual Na^+ and DHP ions. The deviation of the simulations from the experimental density at higher salt concentrations also leads to the deviation of partial molar values

at higher concentrations. However, we observed that the charge distribution alone cannot be used to adjust the density in the MD simulation to reproduce the experimental density.

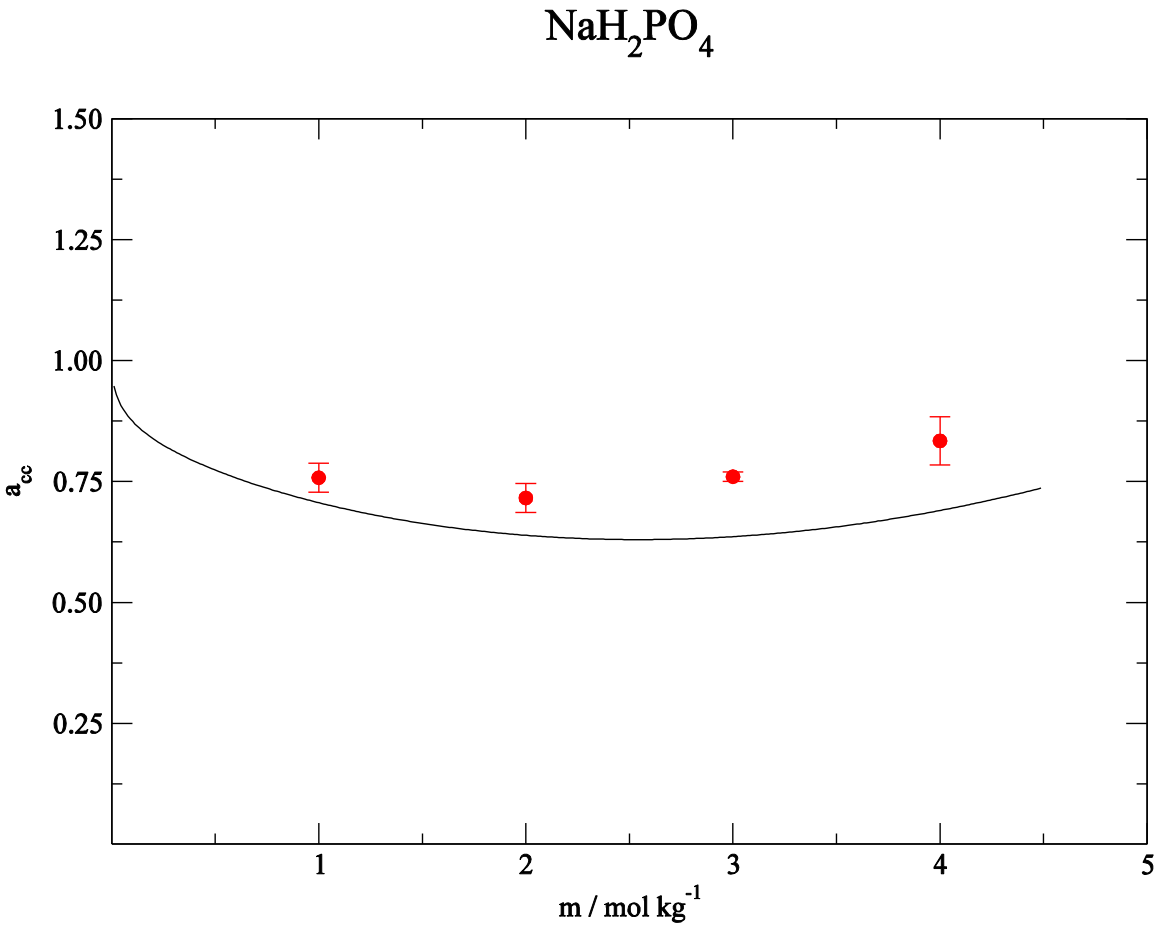

Figure 4.5. Activity derivatives for aqueous NaH_2PO_4 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

The activity derivatives for the simulated aqueous NaDHP salt solutions are compared to the experimental values in Figure 4.5. The experimental activity derivatives trends are reproduced in the MD simulations, with a slight overestimation at higher concentrations.

All the chemical properties discussed so far are directly related to the KB integrals. Therefore, we have examined some other properties, such as diffusion and dielectric constants, which are not related to the KB integrals. These are calculated for the MD simulation and are displayed in Figure 4.6. As displayed in Figure 4.6., the diffusion coefficient of water is higher than for the ions. The Na^+ ion and the DHP anion diffusion is similar for all the salt concentrations. An increase in salt concentration gradually decreases the rate diffusion of ions and water as expected.

Furthermore, the general trend observed for salt solutions is that an increase in salt concentration results in a decrease in the dielectric constant. However, for DHP salts the experimental data indicates an initial increase in the dielectric constant followed by a decrease at higher concentrations. The DHP anion hydration properties are significantly affected by the type of the cation in the salt, as reported in the experimental studies.[37] A kosmotrope ion is usually small with a high charge density and are therefore able to bind with water strongly. On the other hand, a chaotrope ion is a big ion with a low charge density that binds with water relatively weakly. The DHP ion and Na^+ ion are on the borderline of this classification. According to the "matching water affinities" concept, two kosmotrope ions, or two chaotrope ions, in an aqueous solution prefer to dominantly form contact ion pairs.[38-41] However, a combination of a kosmotrope and a chaotrope ion could lead to the preferential formation of solvent separated ion pairs. These solvation properties could impact the experimental dielectric decrement trends observed here. Unfortunately, the current MD simulations follow the general trend of a gradual decrease in the dielectric with salt concentration. Furthermore, it is not clear how to reproduce this type of experimental trend for the dielectric decrement with a classical non-polarizable force field.

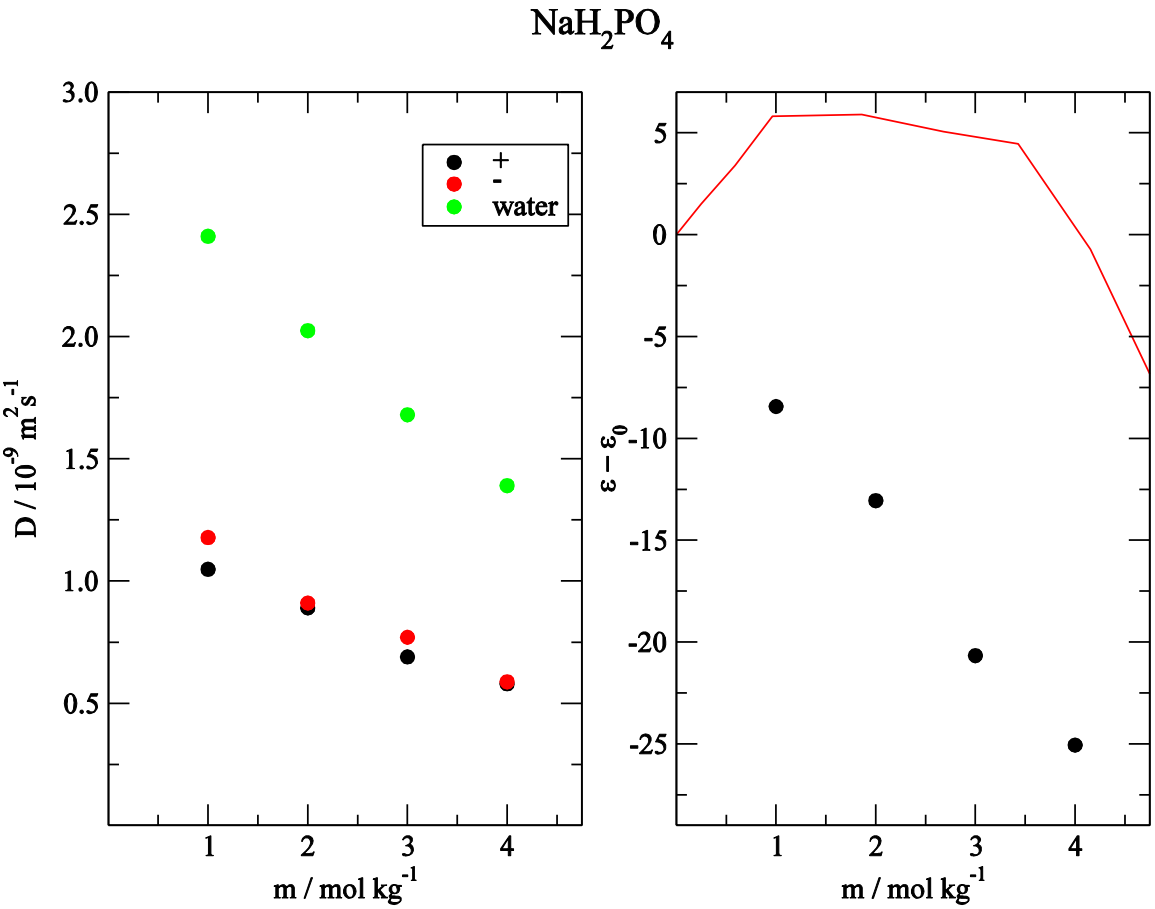

Figure 4.6. The diffusion coefficients and the dielectric decrements for aqueous NaH_2PO_4 salt solutions as a function of salt molality. The lines are displayed for the experimental data[37] and symbols are for the MD simulations.

4.4.1.2. KDHP

The transferablity of the best charge distribution found for NaDHP salt solutions was investigated using the results for KDHP salt solutions. The KDHP salts also did not display any unrealistic aggregation behavior in the present MD simulations. A comparison of the experimental KB integrals for NaDHP and KDHP solutions reflects the changes in the aqueous salt environments. The G_{cc}, G_{cw} and G_{ww} values for a 1m salt solution of NaDHP and KDHP are 177, -24, -17 and 251, -34, -16, respectively. The change of cation from sodium to potassium ion has slightly increased the ion-ion interactions, as implied by an increase in the positive value of G_{cc}. The KB integrals for the ion-water interactions slightly decrease, as implied by an increasingly more negative value for G_{cw}, when a potassium ion replaces a sodium ion in the solution. This could be due to the fact that an increase in the size of the cation can lead to a lower charge density and interaction with the water molecules. The water-water interactions, denoted by G_{ww}, are not affected by the change in cation in the aqueous salt solutions. The MD simulation density, KB integrals and excess coordination numbers are compared with the experimental values for KDHP solutions in Figure 4.7.

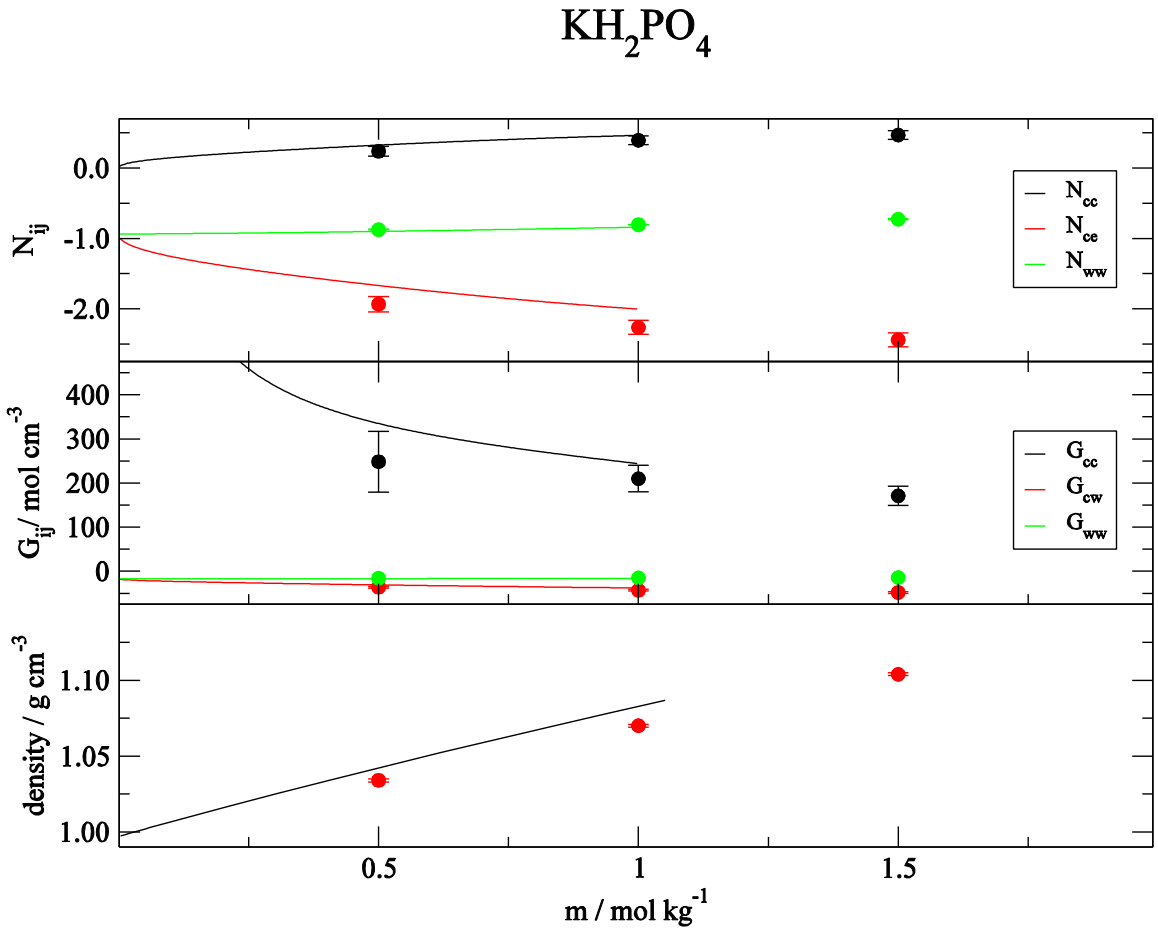

Figure 4.7. The density, KB integrals and the excess coordination number of aqueous KH_2PO_4 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

The comparison of the experimental and MD values for the partial molar volumes of ions and water, the activity derivatives, the diffusion constants, and dielectric decrement are displayed in Figure 4.8, Figure 4.9 and Figure 4.10, respectiviely. The results imply that the charge destribution parameterized for the sodium salt can reproduce most of the KB related experimental properties for the KDHP salt solutions as well. As displayed in Figure 4.10, the diffusion coefficient of water is higher than ions. The K^+ ion diffusion is higher than the DHP anion. The increasing salt concentration gradually decreases the diffusion of ions and water. Furthermore, as

discussed for NaDHP salt solutions the increase in salt concentration results in a decrease in dielectric constant in the MD simulations. However, experimentally the dielectric decrements starts to increase with increasing salt concentrations. The DHP ion lies on the borderline for the classification of kosmotrope and chaotrope, while the K^+ ion is a chaotrope.[39, 41-43] These solvation dynamics could be a reason for the trend in experimental dielectric decrements observed for the KDHP salt solutions.

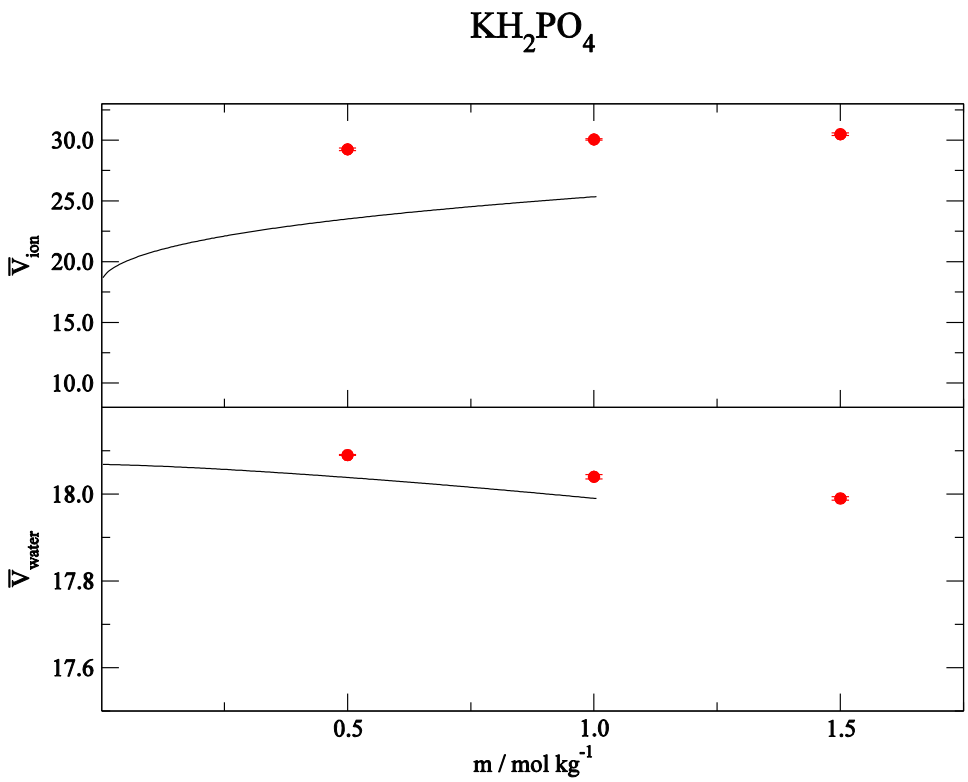

Figure 4.8. Partial molar volumes ($cm^3\ mol^{-1}$) of the ions and water for aqueous KH_2PO_4 salt solutions as a function of salt molality. Lines are fort the KB analysis of the experimental data and symbols are for the MD simulations.

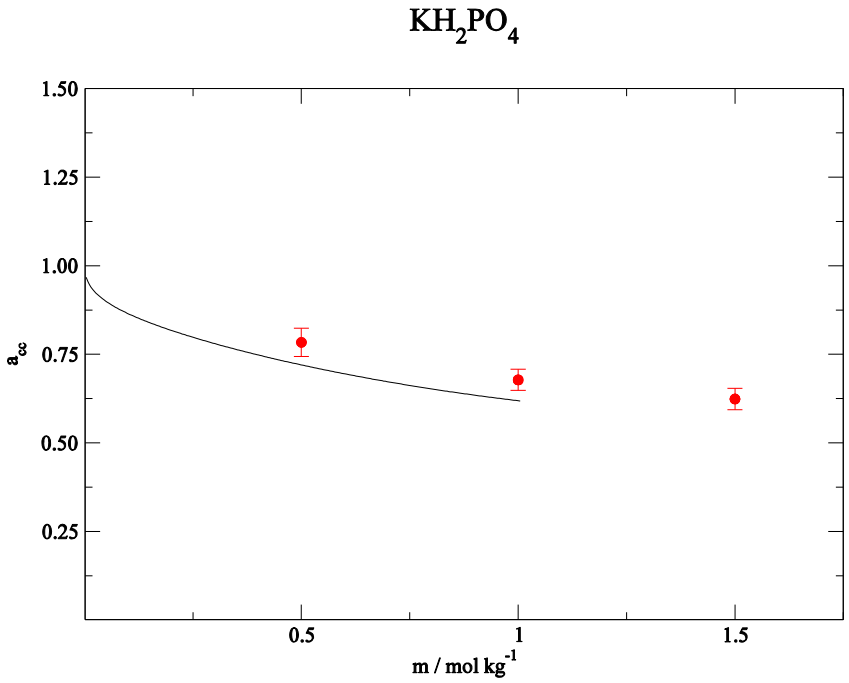

Figure 4.9. Activity derivatives of aqueous KH_2PO_4 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

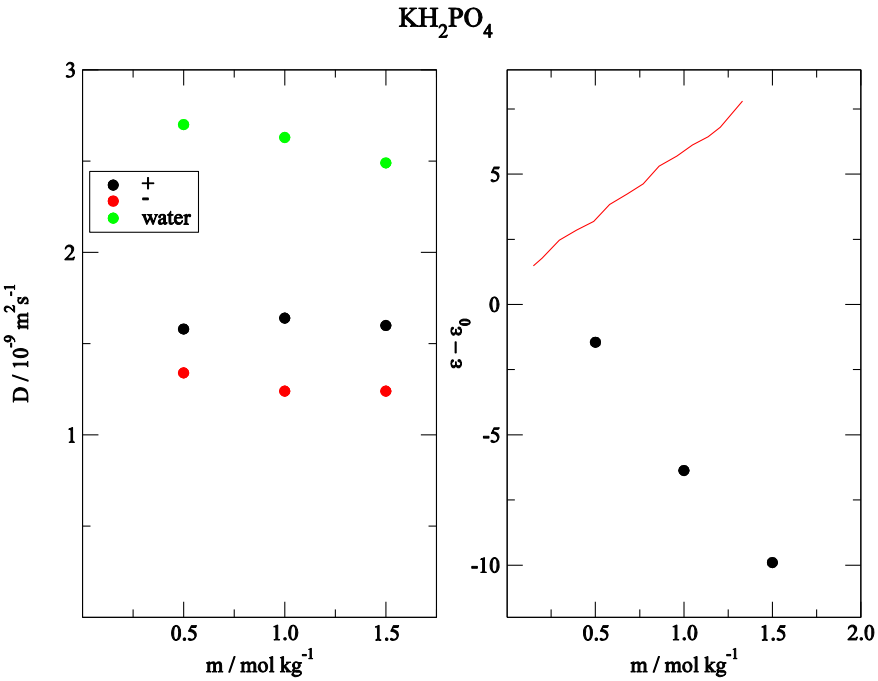

Figure 4.10. The diffusion coefficients and the dielectric decrements for aqueous KH_2PO_4 salt solutions as a function of salt molality. The lines are displayed for the experimental data[44] and symbols are for the MD simulations.

4.4.2. Dimethylphosphate (DMP) ion

For the parameterization of the dimethylphosphate (DMP) ion, instead of using the activity coefficients we had to use the osmotic coefficients because experimental activity data was not available for the LiDMP, NaDMP, and KDMP salt solutions. Even for the osmotic coefficients the solubility limited data to 1 m. Although, no unrealistic aggregation behavior was observed in the MD simulations, a major issue was that the KB integrals were not sensitive to many of the charge distribution attempted for the DMP ions. However, a comparison of experimental KB integrals for DHP and DMP in Figure 4.3 and Figure 4.11 reflects some structural changes. In the DHP molecule the terminal hydrogen molecules are replaced by methyl groups to give DMP. This lowers the DMP molecules ability to form hydrogen bonds with surrounding waters. As an example, in 1m aqueous salt solutions of NaDHP the G_{cc}, G_{cw} and G_{ww} values are 177, -24 and -16, respectively. For 1m aqueous salt solutions of NaDMP the corresponding G_{cc}, G_{cw} and G_{ww} values are -144, -15 and -17, respectively. In contrast to the replacement of the cation, from sodium to potassium DHP salts, changing the anion from DHP to DMP has clearly changed the G_{cc} values. The positive value for the NaDHP salt solution G_{cc} values has shifted to a negative value for the NaDMP implying that ion-ion interactions have changed and are less favorable in the solution.

Table 4.2. The charge distribution for the dimethylphosphate (DMP) ion that best reproduced the experimental KBIs in MD simulations of aqueous NaDMP salt solutions.

CH_3	O	P	O	O	O	CH_3
0.1	-0.4	0.4	-0.4	-0.4	-0.4	0.1

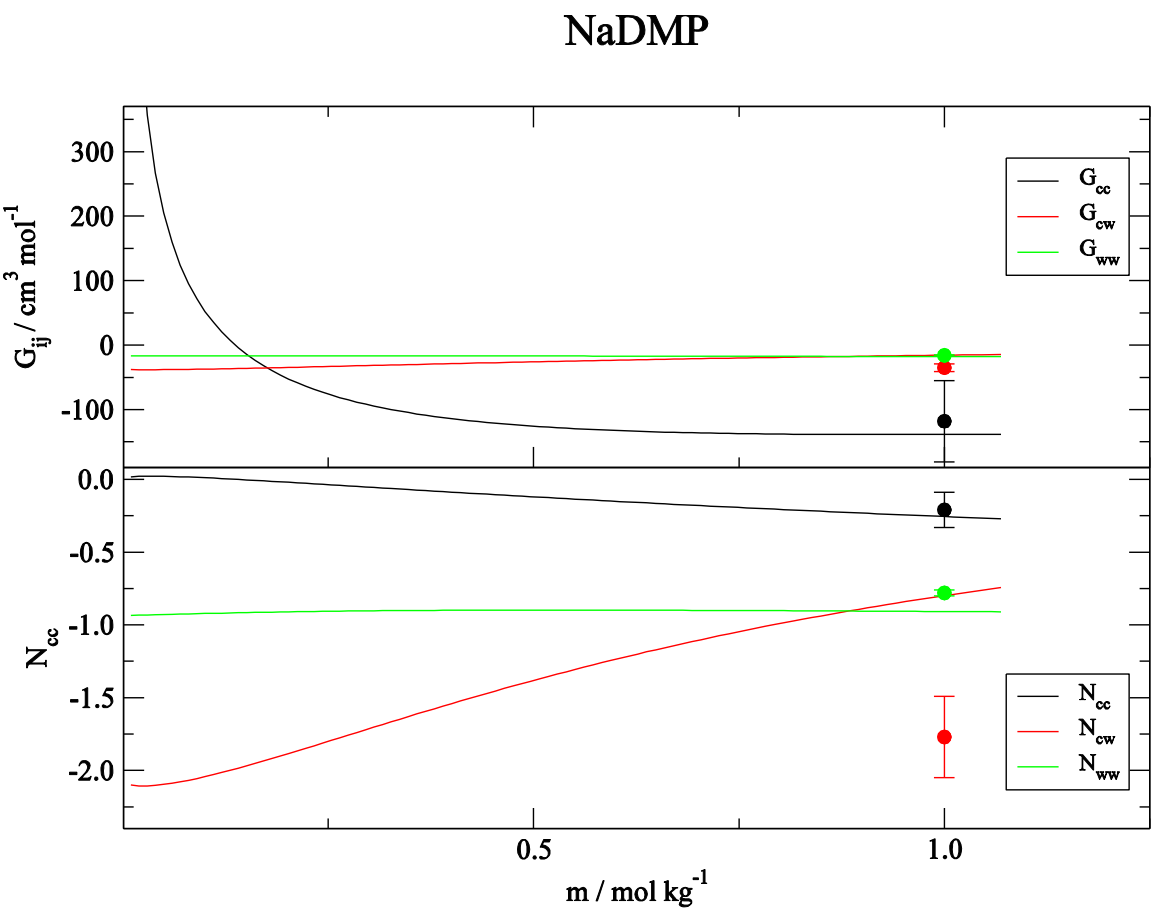

Figure 4.11. The KB integrals and excess coordination number for aqueous NaDMP salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

Also could not find any trends or preferences of charges on a specific atoms in the DMP salts of lithium, sodium and potassium. Table 4.2 displays the DMP charge distribution that most closely reproduced the values of the experimental KBIs for the sodium salt in the MD simulations.

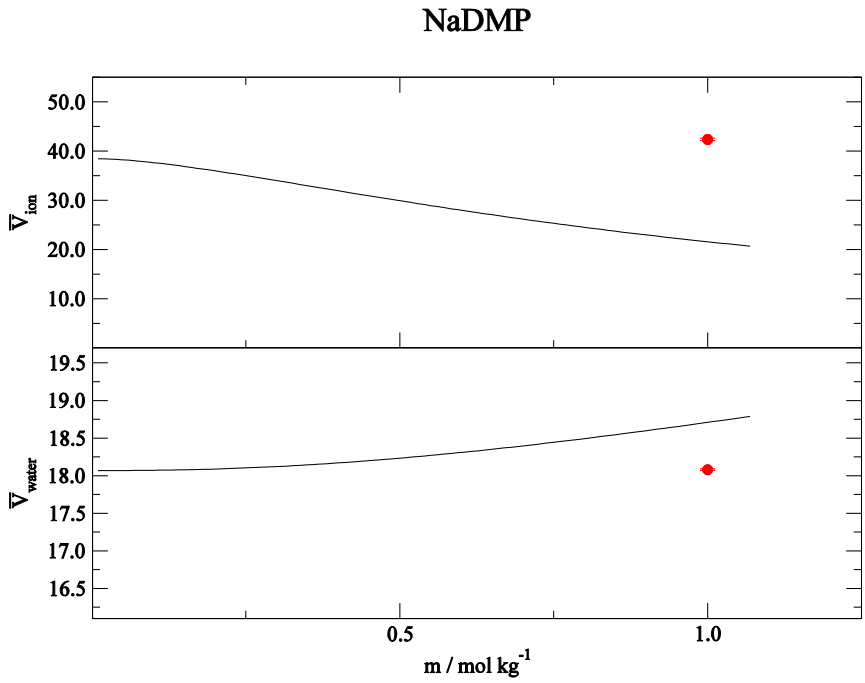

Figure 4.12. Partial molar volumes (cm^3 mol^{-1}) of the ions and water for aqueous NaDMP salt solutions as a function of salt molality. Lines are fort the KB analysis of the experimental data and symbols are for the MD simulations.

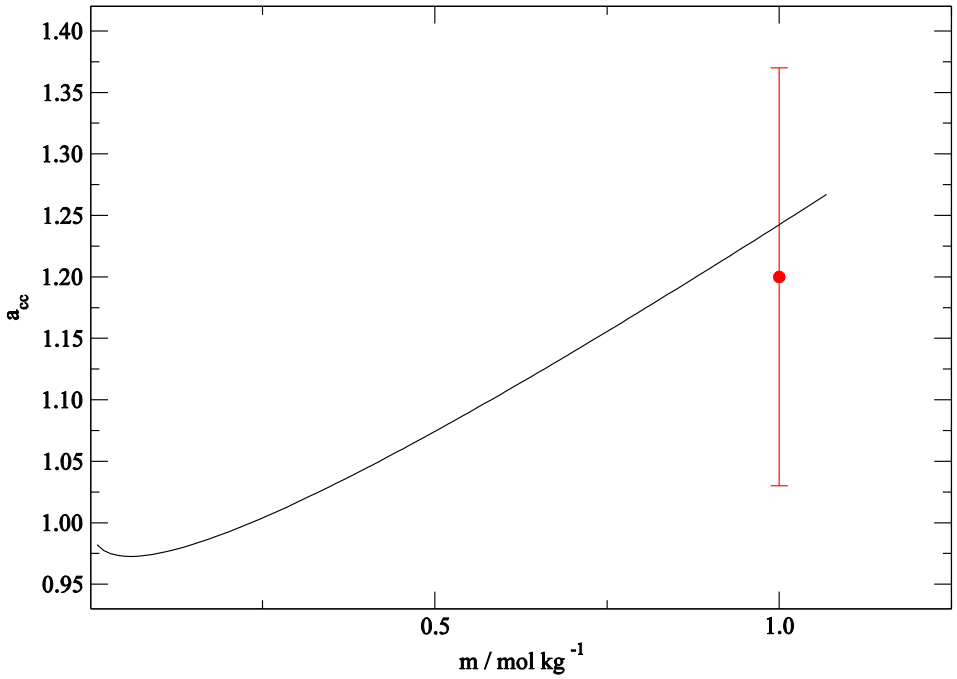

Figure 4.13. Activity derivatives fort aqueous NaDMP salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data and symbols are for the MD simulations.

The interested reader is referred to some of the trial charge distributions attempted for the DMP salts, as displayed in the supporting material of this chapter, see Table S 4.1, Table S 4.2 and Table S 4.3. The charge distributions in these tables also includes a comparison with the charge distribution adopted by the AMBER force field for the DNA backbone[45]. Also included are the Amber99[46], GROMOS G43a1 force field[28] and some QM charge distributions (using ESP, Lowdin and Mullikan analysis). The charge distribution that best reproduced the experimental KB integral for DHP salts was also attempted for DMP, but did not results in reasonable KBIs. The experimental trend of increasing negative values for the ion-ion KB integrals - from -117, -144 and -153 for the lithium, sodium and potassium salt solutions, respectively – were also not captured in the MD simulations.

As displayed in Table S 4.1, Table S 4.2 and Table S 4.3, using the AMBER force field DNA backbone[45] charges for the DMP molecule did not reproduce the experimental KBI values in the MD simulations for LiDMP, NaDMP and KDMP solutions. The G_{cw} values displayed a higher negative value than the expected experimental values. Also the G_{cc} values did not display an increasing negative value when the size of the cation is increased. Further attempts to use this same DMP charge distribution, together with the KBFF parameters was also attempted and did not help to reproduce experimental KBI. The same approach was attempted with Amber99, GROMOS G43a1, some QM charge distributions and the best charge distribution of DHP, yet non of them could reproduce the experimental KBI values in simulation of LiDMP, NaDMP, and KDMP salt solutions. However, the standard GROMOS G43a1 force field charge distribution for DMP did produce KBI closer to the experimental values, but only for NaDMP solutions. However, KDMP and LiDMP could not simulated because those cations are not included in the standard GROMOS G43a1 force field. Furthermore, the experimental KBIs could not be reproduced in the simulations

when the charge distribution from the standard GROMOS G43a1 force field for the DMP molecule was used together with the KBFF parameters.

The charge distribution for the NaDMP salt that best reproduced the experimental KBIs when used with the other KBFF parameters in the MD simulations is displayed in Table 4.2. The corresponding KB integrals and excess coordination numbers are displayed in Figure 4.11. However, the KB integral for the ion-water interactions shows a significant deviation from the experimental data. Figure 4.12 compares the partial molar volumes in the MD simulations with the experimental values. The partial molar volume of the ions is slightly overestimated in the MD simulations and the partial molar volumes of the water molecules are underestimated compared to the experimental values. The activity coefficient derivative with respect to molality is displayed in Figure 4.13 and MD simulation values are within the range of the experimental values. The diffusion coefficients for the sodium ion, DMP ion, and water are 1.06 $x10^{-9}$ m^2s^{-1}, 1.03 $x10^{-9}$ m^2s^{-1} and 2.35 $x10^{-9}$ m^2s^{-1}, respectively. The dielectric decrement of the 1m aqueous NaDMP salt solution is -7.25. However, the charge distribution displayed in the Table 4.2 did not reproduce the experimental KBI values for lithium and potassium salt solutions in the MD simulations. The charge distribution displayed in Table 4.2 when used for the 1m LiDMP salt solutions gave G_{cc}, G_{cw} and G_{ww} values of -44 ±13, -46 ±3,-16 ±0, whereas the experimental KBI values are -117, -16, 17, respectively. The same charge distribution gave values -67 ±5, -48 ±1,-16 ±0 for the 1m KDMP salt solution in the MD simulations whereas -153, -17, -17 are the experimental G_{cc}, G_{cw} and G_{ww}, respectively.

4.4.3. Monohydrogenphosphate (MHP) ion

Sodium monohydrogenphosphate (MHP) salt solutions displayed unrealistic aggregation behavior at experimentally soluble concentrations in the MD simulations as displayed in Figure 4.14. The experimental KB analysis reflects a positive value for Gcc implying ion-ion interactions are favorable in the aqueous salt solution. However, the values of Gcc, Gcw are more positive compared to the DHP molecule, presumably due to the high charge density of the MHP molecule, as displayed in Figure 4.15 and Table 4.3 .

All the charge distributions attempted for MHP displayed aggregation in the MD simulations, as denoted by the high positive values for the ion-ion interactions provided in Table 4.3. Since the MHP ion has negative two charge, the incompatibility of high charge ions with the standard water models could lead to excessive ion clustering in the MD simulations. Furthermore, attempts to decrease the excessive ion clustering by scaling the LJ interaction strength of the oxygen atoms between the MHP and water was not productive.

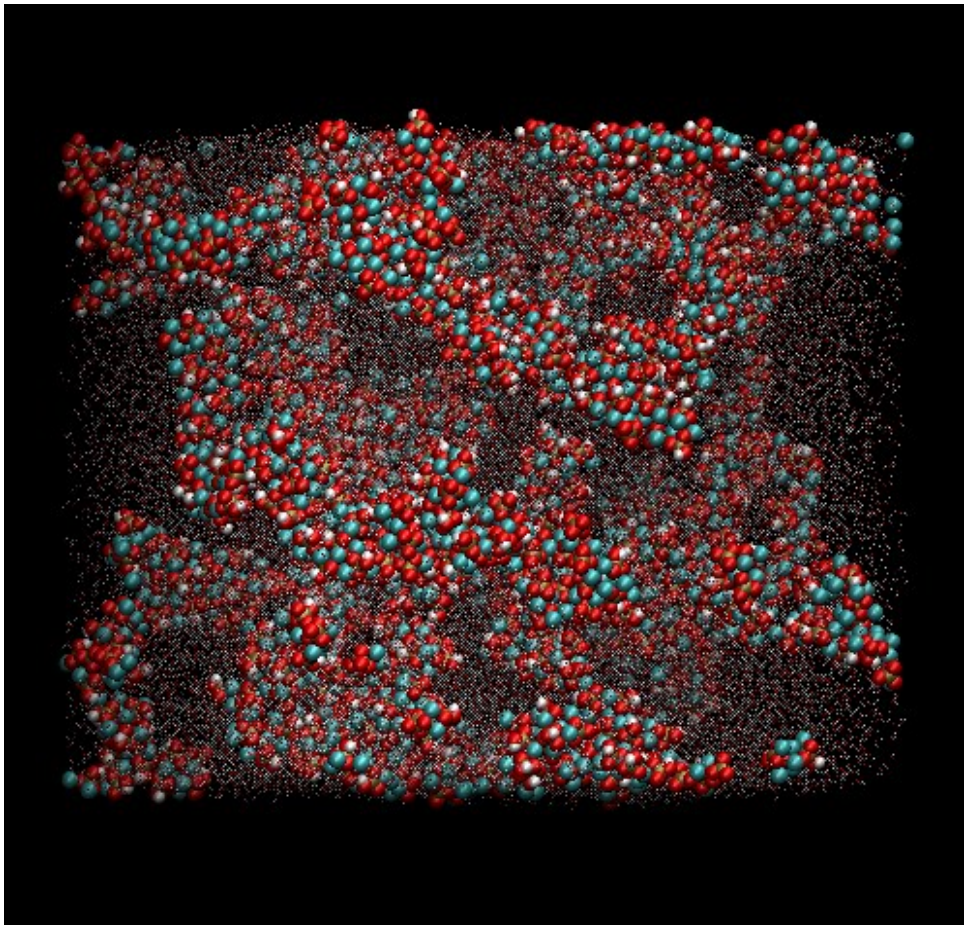

Figure 4.14. Ion aggregation in simulations of 2 m NaMHP salt solutions with the standard SPC/E water model. Sodium ions are colored cyan and oxygen atoms of the MHP ions are displayed in red. The charges on the MHP ion atoms are H = 0.2, O connected to H= 0, P = 0.25, O connected to P = -0.817.

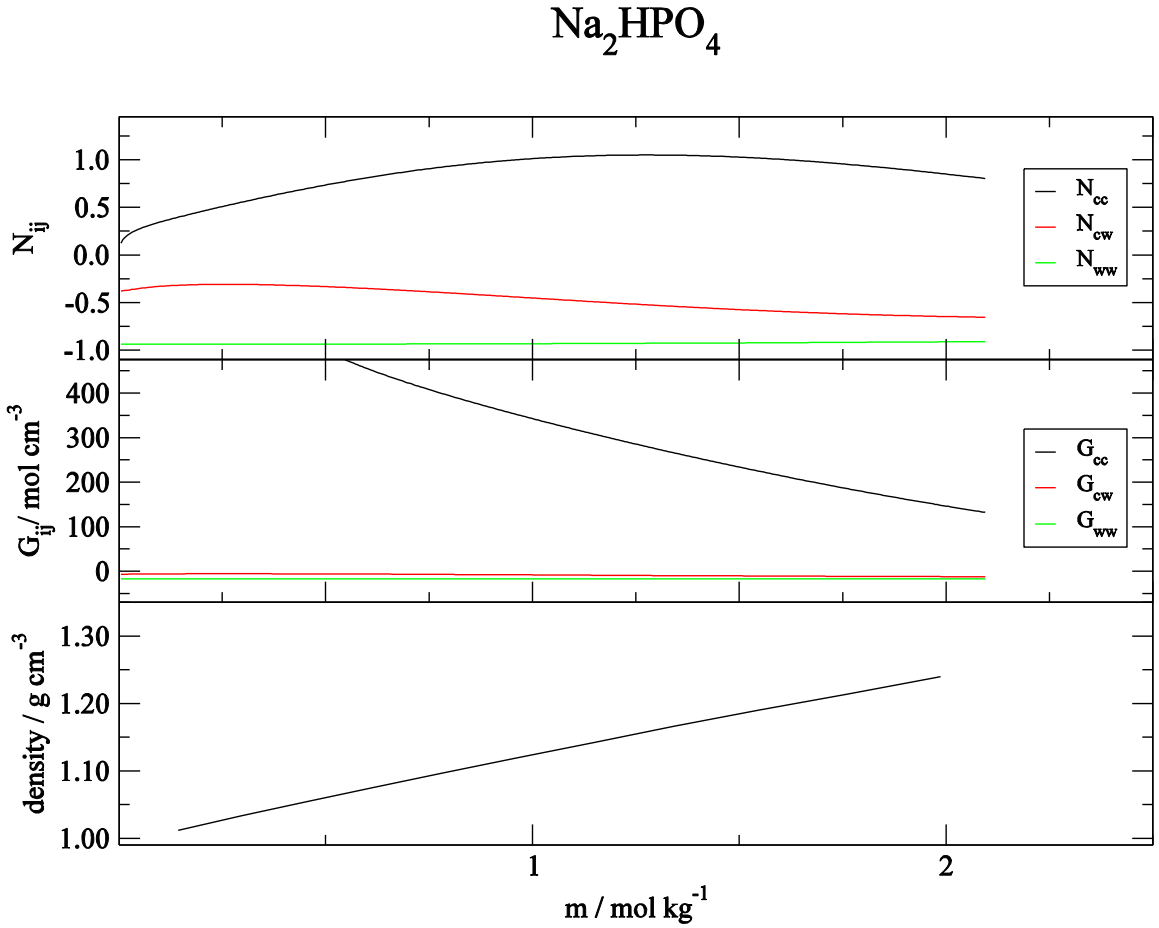

Figure 4.15. The density, KB integrals and the excess coordination number for aqueous NaMHP salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data. MD simulation results are not displayed due to an unrealistic aggregation behavior at all concentrations.

Table 4.3. A few examples of the charge distributions tried for the MHP ion in simulations of aqueous 2 m NaMHP salt solutions. All the charge distributions gave high positive values for the ion-ion interactions (G_{cc}), implying excessive aggregation. The last charge distribution reproduced the lowest ion-ion interactions (G_{cc}) for the aqueous 2 m sodium MHP salt solutions, but are still orders of magnitude higher than the experimental numbers. The units for the KBIs are cm^3mol^{-1}.

H	O	P	O		SPC/E		
					G_{cc}	G_{cw}	G_{ww}
				Exp	147	-12	-17
0.4	-0.6	0.75	-0.85	MD	11503	-1154	96
0.4	-0.2	0.25	-0.817	MD	5028	-495	29
0.4	-0.2	0	-0.73	MD	5966	-584	37
0.4	-0.2	-0.25	-0.65	MD	6296	-619	41
0.4	-0.2	-0.75	-0.483	MD	8055	-783	56
0.2	0	0.25	-0.817	MD	3293	-393	13

4.4.4. Phosphate ion

Sodium phosphate salt solutions also displayed an unrealistic aggregation behavior in the MD simulations at, experimentally soluble, 0.5 m concentration as shown in Figure 4.16. This aggregation is evident in all the charge distributions attempted here. In Table 4.4 we display a few examples of the charge distributions for the phosphate ion where systemic changes in the charges on the P and O atoms are investigated. The experimental Gcc values are higher positive numbers compare to the DHP, MHP and DMP slats implying ion-ion interactions are more favorable due to the high charge density of the phosphate ion. However, instead of the experimentally expected range of positive numbers displayed in Figure 4.17 for the ion-ion KB integrals, the MD simulations resulted in very high positive numbers, as displayed in Table 4.4 and implying a high

degree of aggregation. The negative three charge on the phosphate ion could be a reason for the formation of a large number of ion clusters in the MD simulations. Scaling the LJ interaction strength between the oxygen atoms of water molecule and the phosphate ion, to increase the hydration of the ion, was also attempted. Unfortunately, this did not help to decrease the excessive ion clustering in the simulations. Furthermore, the incompatibility of the non-polarizable simple water models with high charge ions, discuss in the first chapter in this thesis, could be a major problem for this system.

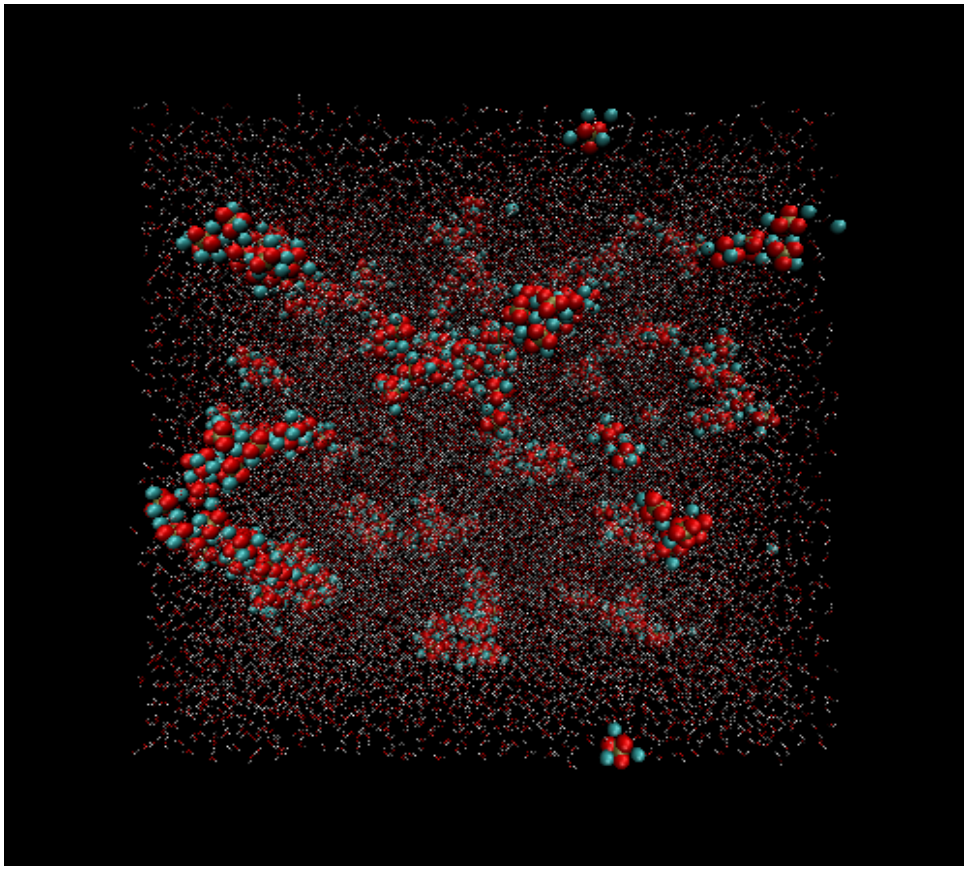

Figure 4.16. The 0.5 m Na_3PO_4 salt solution aggregation with the standard SPC/E water model. Sodium ions are colored cyan color and oxygen atoms of the phosphate ions are red. The charge distribution for the phosphate ion is P = 3 and O = -1.5

Table 4.4. A few examples of the charge distribution tried for the phosphate ion in simulations of aqueous 0.5 m sodium phosphate salt solutions. The phosphate atom and oxygen atom charges are systematically changed. The higher positive values for the ion-ion interactions (G_{cc}) implies an unrealistic aggregation. The units for the KB integrals are cm^3mol^{-1}.

P	O		SPC/E		
			G_{cc}	G_{cw}	G_{ww}
		Exp	697	-1	-17
4	-1.75	MD	6073	13	-16
3	-1.5	MD	11242	-8	-17
2	-1.25	MD	6171	-16	-17
1.5	-1.125	MD	9478	-73	-17
1	-1	MD	6741	-57	-17
-1.5	-0.375	MD	11786	-158	-15

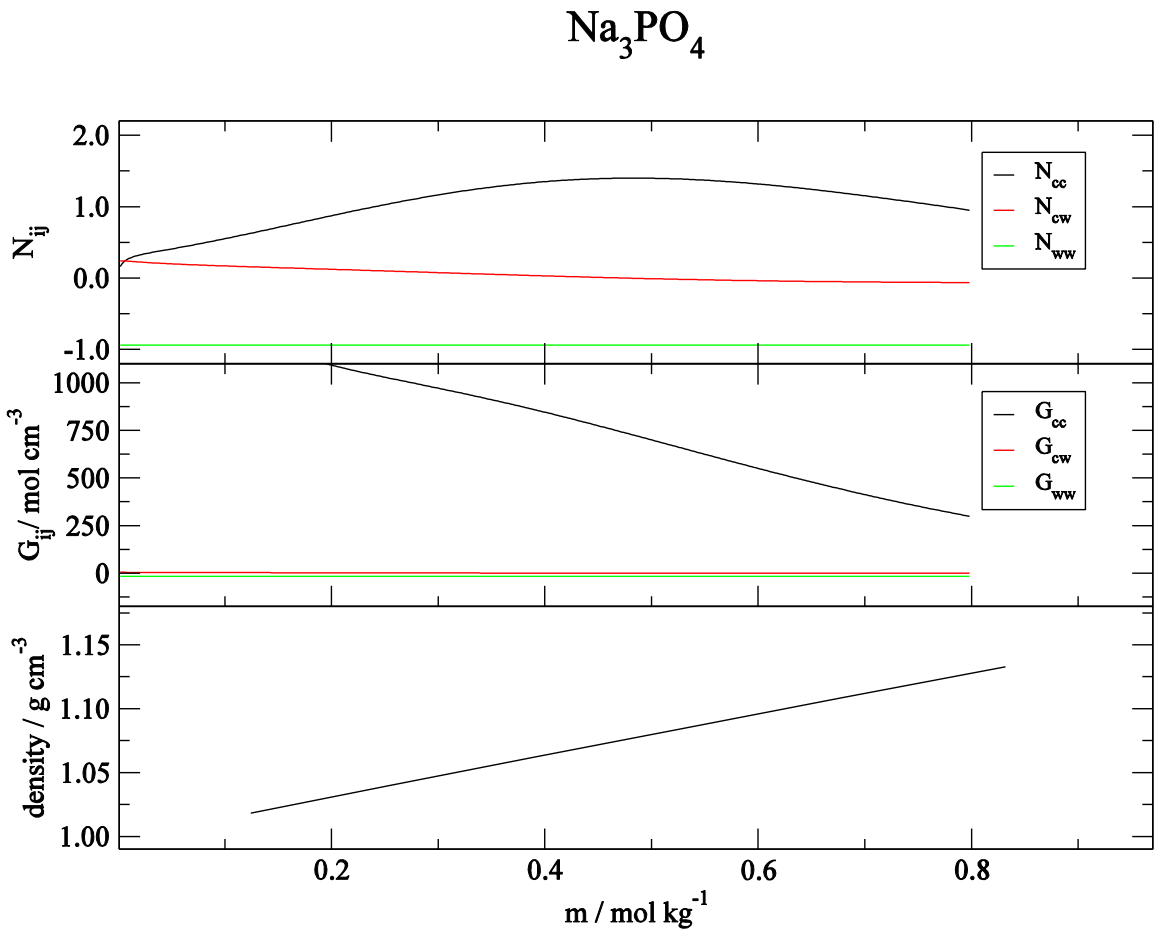

Figure 4.17. The density, KB integrals and the excess coordination numbers for aqueous Na_3PO_4 salt solutions as a function of salt molality. Lines are for the KB analysis of the experimental data. Simulation results are not displayed due to an unrealistic aggregation behavior at all concentrations.

4.5. Conclusions

The parametrization of the PO_4^{3-} and MHP ions was unsuccessful due to the excessive unrealistic aggregation in the MD simulations. However, the DHP parameterization for the KBFF force field was successful. Indeed, the same charge distribution parametrized for the aqueous sodium salt solutions was successfully transferred for the aqueous potassium salt solutions. The DMP ion parameterization using sodium, potassium and lithium salt solutions could not identfy a reasonable single charge distribution effective for all three salt solutions. Therefore, a charge distribution that best reproduced the KBI values for the sodium salt was adopted.

4.6. Supporting Information

Table S 4.1. A few examples of the charge distribution tried for the DMP ion in simulations of aqueous 1 m LiDMP salt solutions. The units for the KB integrals are cm^3mol^{-1}. The standard G43a1 force field do not include lithium ion parameters to perform MD simulation.

Comment	CH3	O	P	OT		LiDMP		
						G_{cc}	G_{cw}	G_{ww}
					Exp	-117	-16	-17
Amber DNA Charges[45] used in the KBFF	0.2	-0.4	1.0	-0.8	MD	33	-47	-16
Standard GROMOS G43a1[28]	0.0	-0.36	0.99	-0.63	MD			
G43a1 charges with KBFF					MD	-58	-49	-14
Standard Amber99[46]	0.20	-0.51	1.16	-0.78	MD	732	-92	-12
Amber99 Charges with KBFF					MD	21	-48	-14
QM charges - ESP	0.14	-0.46	1.15	-0.76	MD	8	-47	-14
QM charges - Lowdin	-0.17	0.05	-0.12	-0.33	MD	7	-49	-15
QM charges - Mullikan	0.08	-0.40	0.90	-0.64	MD	75	-55	-12
From DHYP distribution	0.40	-0.63	0.75	-0.65	MD	61	-44	-15

Table S 4.2. A few examples of the charge distribution tried for the DMP ion in simulations of aqueous 1 m NaDMP salt solutions. The units for the KB integrals are cm^3mol^{-1}.

Comment	CH3	O	P	OT		NaDMP		
						G_{cc}	G_{cw}	G_{ww}
					Exp	-144	-15	-17
Amber DNA Charges[45] used in the KBFF	0.2	-0.4	1.0	-0.8	MD	-16	-40	-15
Standard GROMOS G43a1[28]	0.0	-0.36	0.99	-0.63	MD	-141	-30	-17
G43a1 charges with KBFF					MD	-26	-44	-15
Standard Amber99[46]	0.20	-0.51	1.16	-0.78	MD	-108	-28	-16
Amber99 Charges with KBFF					MD	-15	-40	-15
QM charges - ESP	0.14	-0.46	1.15	-0.76	MD	-40	-43	-16
QM charges - Lowdin	-0.17	0.05	-0.12	-0.33	MD	-104	-38	-15
QM charges - Mullikan	0.08	-0.39	0.90	-0.64	MD	-65	-39	-17
From DHYP distribution	0.40	-0.63	0.75	-0.64	MD	92	-46	-15

Table S 4.3. A few examples of the charge distribution tried for the DMP ion in simulations of aqueous 1 m KDMP salt solutions. The units for the KB integrals are cm^3mol^{-1}. The standard G43a1 force field do not include potassium ion parameters to perform MD simulation.

Comment	CH3	O	P	OT		KDMP		
						G_{cc}	G_{cw}	G_{ww}
					Exp	-153	-17	-17
Amber DNA Charges[45] used in the KBFF	0.2	-0.4	1.0	-0.8	MD	23	-50	-16
Standard GROMOS G43a1[28]	0.0	-0.36	0.99	-0.63	MD			
G43a1 charges with KBFF					MD	40	-55	-14
Standard Amber99[46]	0.20	-0.51	1.16	-0.78	MD	-154	-28	-14
Amber99 Charges with KBFF					MD	36	-51	-15
QM charges - ESP	0.14	-0.46	1.15	-0.76	MD	-23	-51	-14
QM charges - Lowdin	-0.17	0.05	-0.12	-0.33	MD	-59	-49	-14
QM charges - Mullikan	0.08	-0.39	0.90	-0.64	MD	-44	-50	-14
From DHYP distribution	0.40	-0.63	0.75	-0.64	MD	27	-45	-15

References

1. Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P., *Molecular biology of the cell*. 4 ed.; Garland Science: 2002.

2. Costard, R.; Tyborski, T.; Fingerhut, B. P., *Phys. Chem. Chem. Phys.* **2015,** *17*, 29906-29917.

3. Westheimer, F., *Science* **1987,** *235* (4793), 1173-1178.

4. Yoo, J.; Aksimentiev, A., *The Journal of Physical Chemistry Letters* **2012,** *3*, 45-50.

5. Margreitter, C.; Reif, M. M.; Oostenbrink, C., *Journal of Computational Chemistry* **2017,** *38* (10), 714-720.

6. Yoo, J.; Wilson, J.; Aksimentiev, A., *Biopolymers* **2016,** *105*, 752-763.

7. Mahamid, J.; Sharir, A.; Addadi, L.; Weiner, S., *Proceedings of the National Academy of Sciences of the United States of America* **2008,** *105*, 12748-53.

8. Koleva, V.; Stefov, V., *Vibrational Spectroscopy* **2013,** *64*, 89-100.

9. Jiang, L.; Sun, S.-T.; Heine, N.; Liu, J.-W.; Yacovitch, T. I.; Wende, T.; Liu, Z.-F.; Neumark, D. M.; Asmis, K. R., *Phys. Chem. Chem. Phys.* **2014,** *16*, 1314-1318.

10. Costard, R.; Tyborski, T.; Fingerhut, B. P.; Elsaesser, T., *The Journal of Chemical Physics* **2015,** *142*, 212406.

11. Marco Klähn; Gerald Mathias; Carsten Kötting; Marco Nonella; Jürgen Schlitter; Klaus Gerwert, a.; Paul Tavan*, *J. Phys. Chem. A* **2004,** *108* (29), 6186-6194.

12. VandeVondele, J.; Tröster, P.; Tavan, P.; Mathias, G., *The Journal of Physical Chemistry A* **2012,** *116*, 2466-2474.

13. Lee, W. K.; Prohofsky, E. W., *Chemical Physics Letters* **1982,** *85*, 98-102.

14. Florián, J.; Štrajbl, M.; Warshel, A., *Journal of the American Chemical Society* **1998,** *120*, 7959-7966.

15. Kuo, I. F.; Tobias, D. J., *Journal of Physical Chemistry B* **2001,** *105*, 5827-5832.

16. Alagona, G.; Ghio, C.; Kollman, P., *Journal of the American Chemical Society* **1985,** *107*, 2229-2239.

17. Jayaram, B.; Ravishanker, G.; Beveridge, D. L., *The Journal of Physical Chemistry* **1988,** *92*, 1032-1034.

18. Jayaram, B.; Mezei, M.; Beveridge, D. L., *Journal of the American Chemical Society* **1988,** *110*, 1691-1694.

19. Feller, S. E.; Mackerell Jr, A. D., *J. Phys. Chem. B* **2000,** *104*, 7510-7515.

20. Sook Kang, N.; Hyun Jung, D.; Tai No, K.; Shik Jhon, M., Molecular dynamics simulation of Na+-DMP− and Na+-MP2− ion pair in aqueous solution. In *Chemical Physics Letters*, 2002; Vol. 364, pp 580-585.

21. Ben-Naim, A., *The Journal of Chemical Physics* **1977,** *67*, 4884.

22. Smith, P. E., *The Journal of chemical physics* **2008,** *129*, 124509.

23. Pitzer, K. S., *The Journal of Physical Chemistry* **1973,** *77*, 268-277.

24. Robinson, R. A.; Stokes, R. H., *Electrolyte Solutions*. Butterworth and Company: London, 1970.

25. Tamaki, K.; Suga, K.; Tanihara, E., *Bulletin of the Chemical Society of Japan* **1987,** *60*, 1225-1229.

26. Kirkwood, J. G.; Buff, F. P., *The Journal of Chemical Physics* **1951,** *19*, 774.

27. Bratko, D.; Blum, L.; Luzar, A., *The Journal of Chemical Physics* **1985,** *83*, 6367.

28. Van Gunsteren, W. F.; Berendsen, H. J. C., *Library manual (Nijenborgh, Groningen)* **1987**, 1-221.

29. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

30. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81*, 3684.

31. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

32. Darden, T.; York, D.; Pedersen, L., *The Journal of Chemical Physics* **1993,** *98*, 10089.

33. Chandrasekhar, S., *Reviews of Modern Physics* **1943,** *15* (1), 1-89.

34. Chitra, R.; Smith, P. E., *The Journal of Physical Chemistry B* **2000,** *104* (24), 5854-5864.

35. Allen, M. P.; Tildesley, D. J., *Computer Simulation of Liquids*. Oxford University Press: USA, 1987.

36. Smith, P. E.; Gunsteren, W. F. v., *The Journal of Chemical Physics* **1994,** *100* (1), 577-585.

37. Eiberweiser, A. Hydration and Ion Pairing of Aqueous Phosphate Solutions as Observed by Dielectric Spectroscopy. University of Regensburg, 2013.

38. Perez-Jimenez, R.; Godoy-Ruiz, R.; Ibarra-Molero, B.; Sanchez-Ruiz, J. M., *Biophysical Journal* **2004,** *86*, 2414-2429.

39. Collins, K. D., *Methods* **2004,** *34* (3), 300-311.

40. Zhang, Y.; Cremer, P. S., *Current Opinion in Chemical Biology* **2006,** *10*, 658-663.

41. Lo Nostro, P.; Ninham, B. W., *Chemical Reviews* **2012,** *112*, 2286-2322.

42. Record, M. T.; Zhang, W.; Anderson, C. F., *Advances in protein chemistry* **1998,** *51*, 281-353.

43. Ganguly, P.; Hajari, T.; van der Vegt, N. F. A., *The Journal of Physical Chemistry B* **2014,** *118*, 5331-5339.

44. Eiberweiser, A.; Nazet, A.; Hefter, G.; Buchner, R., *The Journal of Physical Chemistry B* **2015,** *119* (16), 5270-5281.

45. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., *J. Am. Chem. SOC* **1995,** *117*, 5179-5197.

46. Wang, J.; Cieplak, P.; Kollman, P. A., *Journal of Computational Chemistry* **2000,** *21* (12), 1049-1074.

Chapter 5 - The Impact of Including Electronic Screening Effects in Ion Simulations

5.1. Abstract

The incompatibility of high charged ions and simple non-polarizable water models is a major concern for MD simulations. This results in excessive ion clustering and an unrealistic aggregation behavior at finite concentrations of salts for ions with a high charge density. MD simulations of the Na_2CO_3, Na_2HPO_4 and Na_3PO_4 salt systems with the SPC/E water model displayed such aggregation behavior. Here, we investigate the possibility of including electronic screening effects in the simulations and the consequences for the aggregation of ions in these salt solutions. Preliminary studies using a water model with a dipole moment close to the experimental value, after including electronic screening, display promising results for the simulation of ions with high charge densities.

5.2. Introduction

In classical MD simulations one uses an average charge distribution for a molecule, in terms of "effective" partial atomic charges, to account for polarization effects in solution. In a real system the charge distribution will adjust to the local microscopic environment, and also include screening effects from the electron cloud of the surrounding molecules. Since classical simulations do not define electrons explicitly, it is a tedious task to model these adjustments to the local microscopic environments. For a neutral molecule the impact of the electronic screening effect is implicitly handed by the use of "effective" charges in classical non polarizable force fields. However, for ions this strategy is not capable of modeling electronic screening in MD simulations

as most force fields use integer charges for ions. This incompatibility appears to be more prominent with high charge density ions in water resulting unrealistic excessive ion clusters. This aggregation behavior in MD simulations prevails irrespective of the water model used with the high charge ion species. We believe this is a major factor that prevents the development of simple models for many of the ono anions discussed in Chapters Three and Four of this thesis.

Indeed, Leontyev et al. have investigated the possibility of including the charge fluctuations around the average value by using an additional charge renormalization factor which is shown to produce promising results.[1-4] The adjustment to the charge distribution changes the effective dipole moment of the molecule. Furthermore, other research groups have reported that the scaling of ion charges is an effective strategy to model electronic screening in classical MD simulations for ions with high charge density.[5] However, the renormalization of the charges results in non integer values which could lead to inconsistencies in the MD simulations.

Therefore, the possibility of including the renormalization factor implicitly in the solvent to account the electron screening effects is explored in this project. This makes the solvent and charges of ions more compatible in MD simulations. Water is the most abundant, biologically most important, and a highly polarizable solvent with unique characteristics. This makes it very challenging to model in computer simulations. Classical molecular dynamics (MD) are widely used as the most common type of computer simulations to study biological systems due to limitations in applications of quantum mechanical simulations. Despite the great number of water models available for use in molecular dynamics simulations, the majority of them involve classical, non-polarizable, simple rigid water models that were developed decades ago.[6-8] These simple water models do not include electronic polarization and screening effects due to the fact the electrons are not explicitly included in classical MD simulations.[9-11] However, there are

polarizable water models available with flexible bonds and multiple sites with complex empirical potential to describe the interactions and specifically developed to include many body effects such as polarization, charge transfer, etc. These polarizable force fields have been shown to perform well compared to non-polarizable force fields, especially in low dielectric environment such as a bio membrane.[12-16] However, polarizable force fields also suffer from some unresolved issues such as computational efficiency, consistency, balancing inter-intra molecular interactions, sampling of the relevant configurations, etc.[12, 14, 17-18] Therefore, using non polarizable force fields is preferred by most researchers over polarizable force fields models, if the non-polarizable force fields models can reproduce reasonably accurate experimental properties in the simulations.

However, there is a growing interest in the scientific community to find strategies to incorporate electronic screening effects with reasonable efficiency in classical MD simulations. One strategy for non-polarizable water models is to use the approach of modeling the charge distribution by multipoles.[19] The monopole, dipole, quadrupole, octupole and hexadecapole moments are the zeroth, first, second, third and fourth moments of the charge distribution. These water models help to capture effects that are hard to model in simple non polarizable water models. As an example, it is claimed that the quadrupoles moments are more sensitive than just dipoles and this influences the orientation of molecules leading to a dense packing of the system.[20] Also including an octupole is argued to be necessary to distinguish anion and cation solvation dynamics.[21] Although including higher order multipole moments can increase the accuracy of the water model, this can increase the computational cost as well. Consequently, most of the water models neglect quadrupole or higher moments because the liquid properties, such as density or energy, used to fit the potential are not very sensitive to the dipole or higher moments.

Furthermore, as an alternative strategy some water models attempt to use an increased dipole moment to model the contributions by multipolar interactions. Leontyev et al. describe how the electron screening of the system can be treated as macroscopically homogenous. [1-4] As an example, the electronic contribution to the dielectric constant for organic materials is close to 2 and for water is 1.78.[3] Therefore, the average effect of electronic screening can be accounted for in most non polarizable force fields by decreasing the partial atomic charges, while the average effect of polarization in the condensed phase can be accounted for by increasing the partial atomic charges.[1, 22] The gas phase experimental electrostatic dipole moment of water is about 1.85 Debye.[23] For the water dimer the electrostatic dipole moment is about 2.1 Debye, and this value increases in larger water clusters and reaches values close to 3 Debye in the liquid phase.[1, 23] However, most of the water models used in MD simulations use effective dipole moments (μ^{eff}) with values around 2.3.[1] Leontyev et al. explain this value for the effective dipole moment as the value of the ratio between of the real dipole moment of the medium and square root of the electronic dielectric for water (ε_{el}), which is 1.78, to account the electron screening effects as displayed in Eq. 5.1.[1]

$$\mu_{eff} = \frac{\mu_{medium}}{\sqrt{\varepsilon_{el}}} \qquad \textbf{Eq. 5.1}$$

The fixed charge water models that produce a permanent dipole of 2.2 - 2.4 Debye claim to incorporate many body polarization effects. However, this dipole moment is still lower than the experimental liquid water value due to implicit screening effects. A solution to this, called the MDEC (Molecular dynamics in electronic continuum) model, has been proposed by Leontyev et al.[4] In this model, one considers the point charges are moving in a homogenous electronic continuum of known dielectric constant, ε_{el}. Consequently, the new dielectric constant (ε)

obtained from the MD simulations needs to be corrected to include contributions from electronic screening effects as displayed in Eq. 5.2.

$$\varepsilon = \varepsilon_{MD}\varepsilon_{el} \qquad \textbf{Eq. 5.2}$$

Therefore, they argue that current simulations implicitly include this electronic screening effect. This has no real effect on the water models developed previously, but it implies that one should use charges for ions that also include this effect, i.e. the ion charges should be scaled by a factor of $1/\sqrt{\varepsilon_{el}}$ (about 0.75 in an aqueous solution). However, this strategy results in non integer charges for ions. Nevertheless, this would reduce the magnitude of the ion-ion interactions and may help to prevent the aggregation problems observed for many high charge density ions.

Hence, this incompatibility of many ions with standard water models could be the reason for observed unrealistic aggregation behavior of Na_2CO_3, Na_2HPO_4 and Na_3PO_4 salt solutions in the MD simulations as discussed previously. Here, we attempt a similar approach to Leontyev et al. However, instead of using non integer in charges we keep the original chares and increase the dielectric constant used for the electrostatic interactions to account for electronic screening. Furthermore, the possibility of using KB integrals to quantify the ion and water interactions in the modified new water model is also explored.

5.3. Experimental Analysis and MD simulations

A modified water model with a dipole moment closer to the experimental value, after including electronic screening effects, could resolve the excessive aggregation behavior for ions with high charge density observed in our previous MD simulations. Here, we have explored ions with high charge densities and their interactions with the standard SPC/E[7] water model, with dipole moment of 2.351 Debye, and with a modified version of the SPC/E water model which includes

electronic screening effects. In this modified version the dipole moment is increased to 3.16 Debye, which is close to the experimental number. This is achieved by increasing the dielectric constant for electrostatic interactions by an equivalent amount such that the final electrostatic interactions remain unchanged. The parameters are listed in the Table 5.1.

Table 5.1. The parameters used for the SPC/E water model and the modified SPC/E water model with electronic screening effects.

	SPC/E[7]	Modified SPC/E
r(OH), Å	1.0	1.0
HOH Angle	109.47	109.47
q(O)	−0.8476	-1.1306
q(H)	+0.4238	+0.5653
Dipole Moment/ Debye	2.351	3.16

5.3.1. Reparametrization of the cation for simulations using the modified SPC/E water model

All the MD systems studies in this project include Na^+ as the cation which was parameterized in an earlier publication.[24] These ion parameters are compatible with the standard SPC/E water model. Therefore, we need to develop cation parameters that are compatible with the new modified SPC/E water model. Hence, the properties of NaCl solutions was chosen, as in the

original parameterization process, to evaluate the performance of cation and anions in the new water model that includes electronic screening effects. MD simulations of 4 m salt solutions of NaCl, and the NaCl salt crystal, were performed and the results indicated that the existing Na^+ and Cl^- ion parameters with the modified water model could not reproduce the experimental aqueous salt solution density, KB integrals, ion-water contact distances, and NaCl crystal dimensions. Therefore, reparameterization of cation and anion was needed.

For a monatomic ion there are two adjustable parameters, which are the LJ interaction length (σ) and the interaction strength (ε). Therefore, one adjustable parameter was kept constant and the other adjustable parameter was scaled by a known factor for both Na^+ and Cl^- ions and MD simulations performed and the corresponding KB integrals calculated. However, the cation to water oxygen interaction strength (ε) scaling was still required to reproduce experimental KB integrals. Further efforts to incorporate a water to cation interaction strength (ε) scaling implicitly in to the cation interaction strength (ε) was tried but failed to reproduce the experimental KB integrals.

Once the parameters for the sodium cation were established then they were used to perform MD simulations for the 0.5 m salt solutions of Na_3PO_4, 2 m Na_2HPO_4 , and 1 m salt solution of Na_2CO_3 in the modified SCP/E water model systems.

5.3.2. MD simulations details

All the computer simulations for NaCl, Na_2CO_3, Na_2HPO_4 and Na_3PO_4 were performed by using the GROMACS[25] version 4.6 software. The SPC/E water[7] model and a modified version of SPC/E water used in the MD simulations. Most importantly, the dielectric constant in the MD simulation was set to 1.78 to account for contributions from electronic screening effects in the

modified SPC/E water and ion simulations.[1] All simulations in this study were maintained in the isothermal isobaric ensemble at 300 K temperature and 1 atm pressure. Berendsen pressure and temperature coupling was used in this study.[26] Here, temperature and pressure is weakly coupled to a bath with relaxation time of 0.1 ps and 0.5 ps, respectively. A 2 fs time step was used to integrate the equation of motion. All the bonds in the simulations were constrained with the LINCS algorithm.[27] The particle mesh ewald (PME) was used to model electrostatic interactions with 1 nm cut off distance for real space calculations,[28] and 1.5 nm cutoff distance for the van der Waals interactions. All the simulations were performed in a 100 Å size cubic box. A random initial configuration of ions and water in a cubic box with length of 10 nm was generated to provide a known concentration of aqueous salt solution using a custom written Fortran code. An energy minimization using the steepest descend method followed by 2 ns of equilibration was performed before the production run. The production run of 15 ns was used to calculate the ensemble averages. In the production run configurations were saved every 0.1 ps. A custom written Fortran program used to calculate center of mass radial distribution function and to calculate the simulated KB integrals.

For the NaCl salt crystal an anisotropic MD simulation was performed at 300 K and 1 atm. Here, a roughly 5 nm size cubic crystal was generated using a custom written Fortran code. The unit cell dimensions and the symmetry group operations need to generate salt crystals were obtained from the crystallography open database.[29] The additional correction factor 1.78 for the dielectric constant was also used for the crystal simulations.

5.4. Results and Discussion

As displayed at the top of Table 5.2, the standard KBFF parameters for the sodium and chloride ions could not reproduce the experimental KBI values when using the modified SPC/E water model. The salt solution density deviated from the experimental values. Also, the NaCl crystal simulation did not reproduce the ion-ion distances.. In the standard KBFF model, the sodium ion to water interaction strength was scaled by 0.75 to help reproduce the experimental KBI.[30-31] As displayed in Table 5.2, removing this scaling factor resulted in significant deviation from the experimental KBI values in the simulations with modified SPC/E. The cation to water interaction strength scaling is used as a way to include polarization effects implicitly in the classical non polarizable force fields. Therefore, scaling the cation to water interaction strength was also required for the modified water model.

For a given ion there are two adjustable parameters. To determine which parameter to adjust in this present study the NaCl crystal simulation results were used. As shown in Figure 5.1 and Figure 5.2 the most feasible approach is to adjust the LJ interaction length (σ) without modifying the interaction strength (ε) to obtain the experimental crystal dimension and potential energies. Therefore, the LJ interaction length (σ) was scaled by different scaling factors for the both cation and anion without changing the interaction strength (ε) and MD simulations performed to compare with the experimental data. On scaling the LJ interaction length (σ) by 0.95 the crystal cation and anion distance and the salt solution density was not closer to the experimental results as displayed in Table 5.2. On the scaling of LJ interaction length (σ) by 0.9 then the KBI values, the density of salt solution and also the crystal dimensions are close to the experimental numbers.

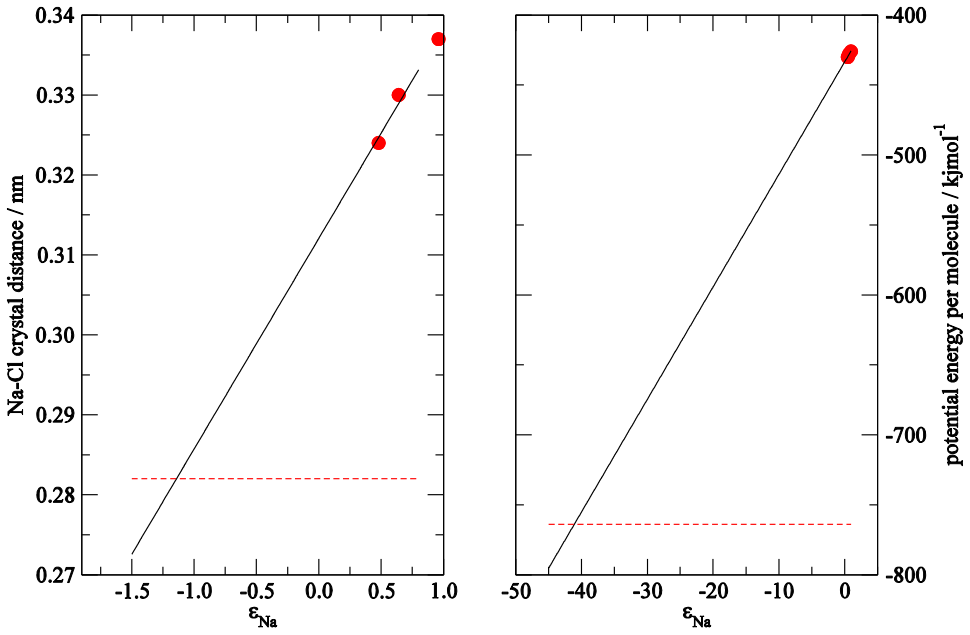

Figure 5.1. Simulations of a NaCl crystal with the interaction length (σ) values kept fixed to standard KBFF values and the interaction strength (ε) value of the sodium ions changed to estimate the experimental ion distances and potential energy of the NaCl crystal. The dash lines are for the experimental number and symbols are for the MD simulations.

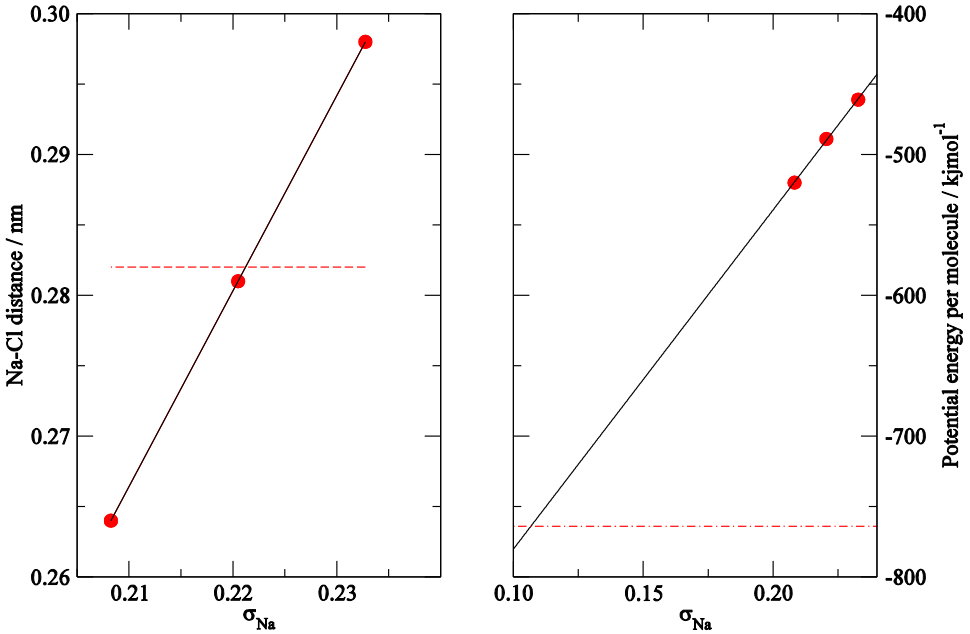

Figure 5.2. simulations of a NaCl crystal with the interaction strength (ε) values kept fixed to standard KBFF values and the interaction length (σ) value of the sodium ions changed to estimate the experimental ion distances and potential energy of the NaCl crystal. The dash lines are for the experimental number and symbols are for the MD simulations.

Table 5.2. comparison of cation and anion sigma, epsilon, cation-water epsilon scaling affect the KB values. The units for the KB integrals are cm^3mol^{-1}.

						Modified SPC/E						Crystal	
	Na		Cl		Scaled	Gcc	Gcw	Gww	Density	Contact Distance Na_OW	Contact Distance Cl_OW	Na-Cl	Pot E/
	σ	ε	σ	ε	ε_{Na-OW}				g/cm^3	/ nm	/ nm	/ nm	/ kj/mol
Exp						-54	-6	-18	1.137	0.24	0.32	0.282	-764
KBFF	0.245	0.32	0.44	0.47	Yes 0.75	-15	-18	-17	1.064	0.235	0.325	0.315	-437
KBFF without $\boldsymbol{\varepsilon_{Na-OW}}$	0.245	0.32	0.44	0.47	No	59	-30	-15	1.062	0.244	0.325	0.315	-437
Cation $\boldsymbol{\varepsilon}$ include Scaling	0.245	0.18	0.44	0.47	No	56	-29	-15	1.061	0.235	0.325	0.311	-442
$\boldsymbol{\sigma}$ Scaled by 0.95	0.233	0.32	0.418	0.47	Yes 0.75	-17	-15	-7	1.087	0.225	0.315	0.298	-462
$\boldsymbol{\sigma}$ Scaled by 0.9	0.221	0.32	0.396	0.47	No	59	-20	-17	1.111	0.225	0.305	0.281	-489
	0.221	0.32	0.396	0.47	Yes 0.75	-26	-11	-18	1.112	0.225	0.305	0.281	-489
					Yes 0.5	-60	-7	-18	1.112	0.225	0.305	0.281	-489
$\boldsymbol{\sigma}$ Scaled by 0.9 and Cation $\boldsymbol{\varepsilon}$ include Scaling	0.221	0.18	0.396	0.47	No	64	-21	-16	1.112	0.225	0.306	0.277	-494

Further trying to incorporate cation to water interaction strength (ε) scaling by 0.75 in to the cation alone did not reproduce the experimental KB integral. Hence, to model polarization effects in classical force fields the modification to the standard combination rule is necessary. Therefore,

while keep with the LJ interaction length (σ) scaling by 0.9, different scaling factors instead of 0.75 for the interaction strength (ε) between cation and water oxygen was attempted. A scaling factor of 0.5 for the interaction strength (ε) between cation and water oxygen resulted reproducing closer results for the experimental KBI and other properties as shown in the Table 5.2.

Figure 5.3 displays a comparison of the radial distribution functions of the 4 m NaCl salt solution with the standard SPC/E water and the modified SPC/E which includes electronic screening effects. The ion-ion interactions and the ion-water interactions are affected in the modified SPC/E water simulations. The rdf values indicate there is no unrealistic aggregation behavior in the modified SPC/E water model simulations. A comparison of the NaCl salt solution density and the KBI and excess coordination number with the experimental data for the new ion parameters and the modified SPC/E model is displayed in Figure 5.4. Therefore, all simulations of systems having sodium as the cation with a modified SPC/E model used a 0.9 scaling of the LJ interaction length (σ) from the standard KBFF, and a 0.5 scaling of the cation to water oxygen interaction strength (ε).

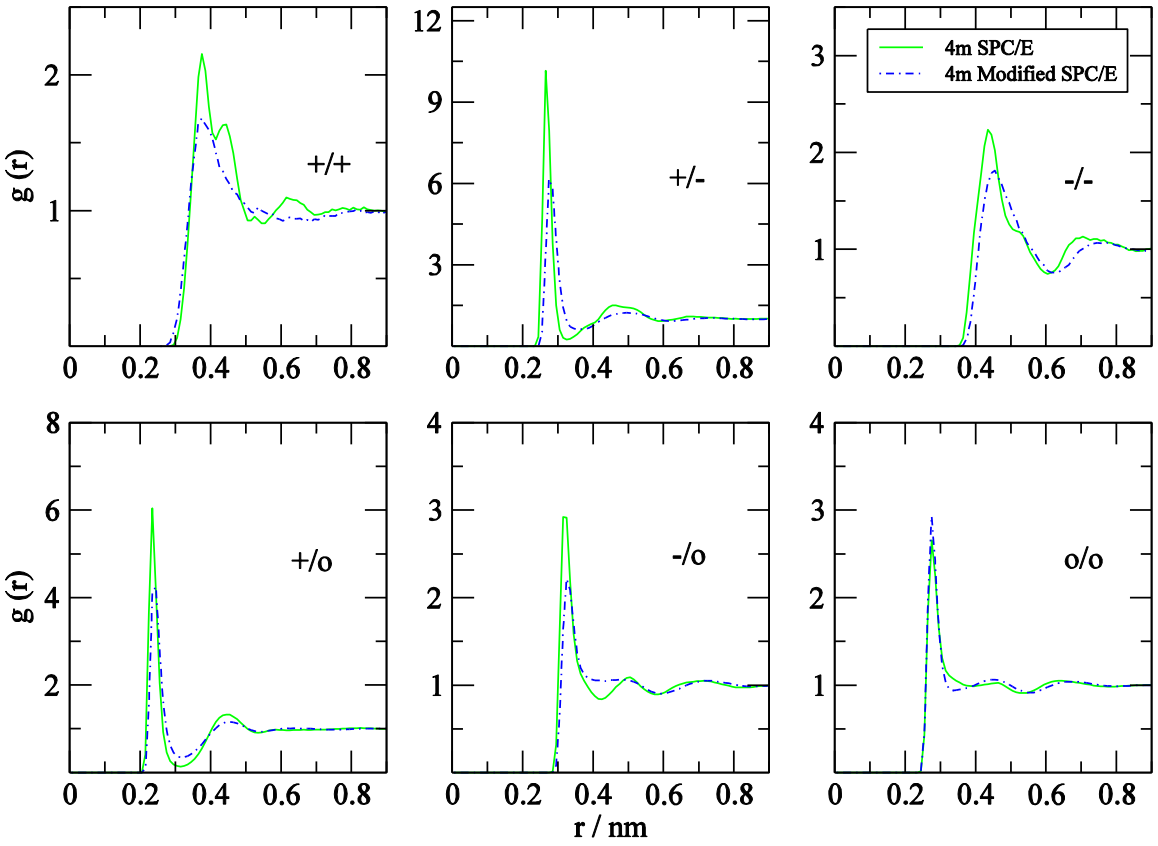

Figure 5.3. The radial distribution functions g (r) from MD simulations of 4 m sodium chloride salt solution in SPC/E water and the modified SPC/E water models. The symbols of +, -, O are for the cation, anion, and water oxygen, respectively.

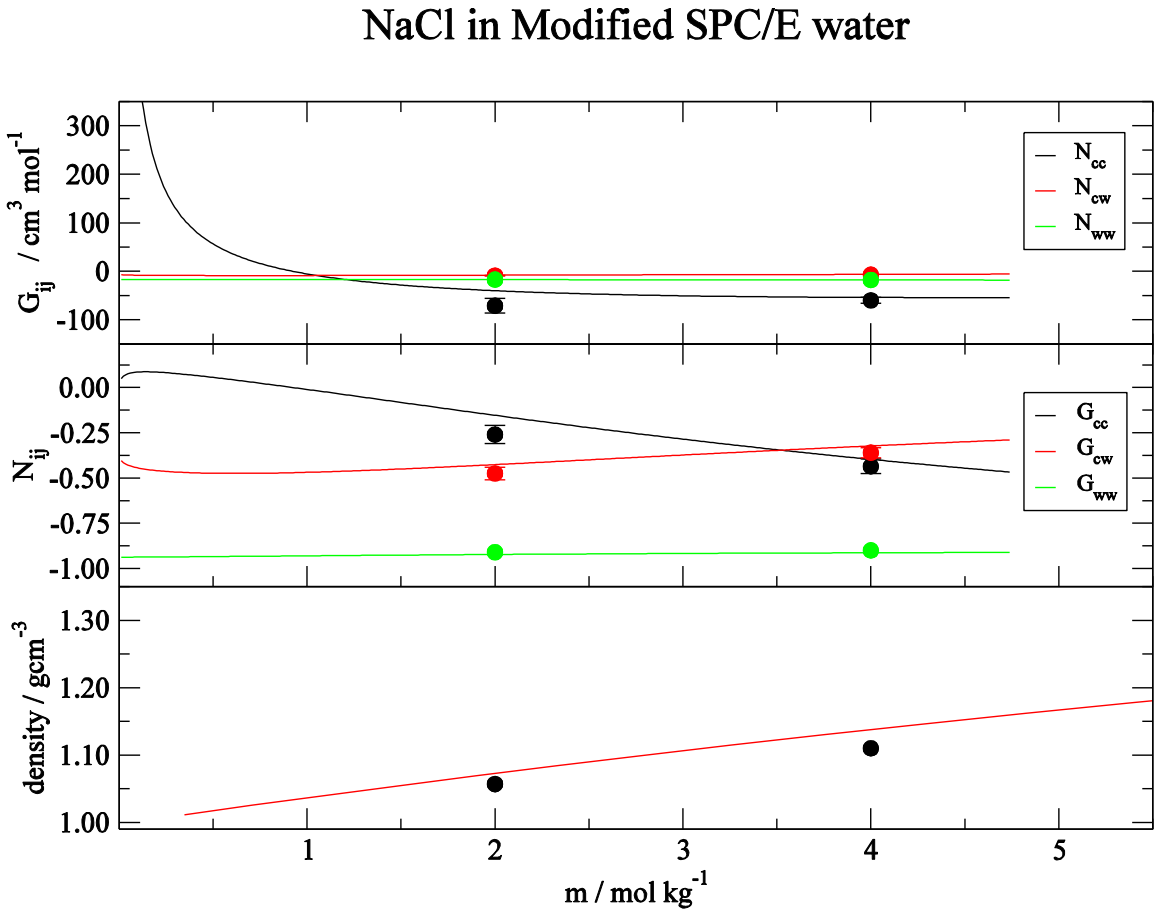

Figure 5.4. The density, KB integrals and the excess coordination numbers for NaCl salt solutions. The symbols are for MD simulations and lines are from the experimental KB analysis.

Table 5.3. Comparison of the 1 m Na_2CO_3 salt solutions using the SPC/E water and the modified SPC/E water models. The units for the KB integrals are cm^3mol^{-1}.

C	O		SPC/E			Modified SPC/E		
			G_{cc}	Gc_w	G_{ww}	G_{cc}	G_{cw}	G_{ww}
		Exp	114	-4	-17	114	-4	-17
1	-1	MD	34654	-973	12	53	-7	-17
0.4	-0.8	MD	31113	-1035	17	111	-8	-17

As mentioned elsewhere in this thesis, simulations of 1m Na_2CO_3 salts with the standard SPCE water models display unrealistic aggregation behavior at finite, experimentally soluble, concentrations. However, once the water model is changed to the modified SPC/E model, which includes electronic screening effects, the aggregation problem disappears. Also, the modified SPC/E water salt solution can then be developed to produce KB integrals close to the experimental values. Furthermore, the KBI values are now sensitive to the charge distributions as displayed in Table 5.3 and Figure 5.5.

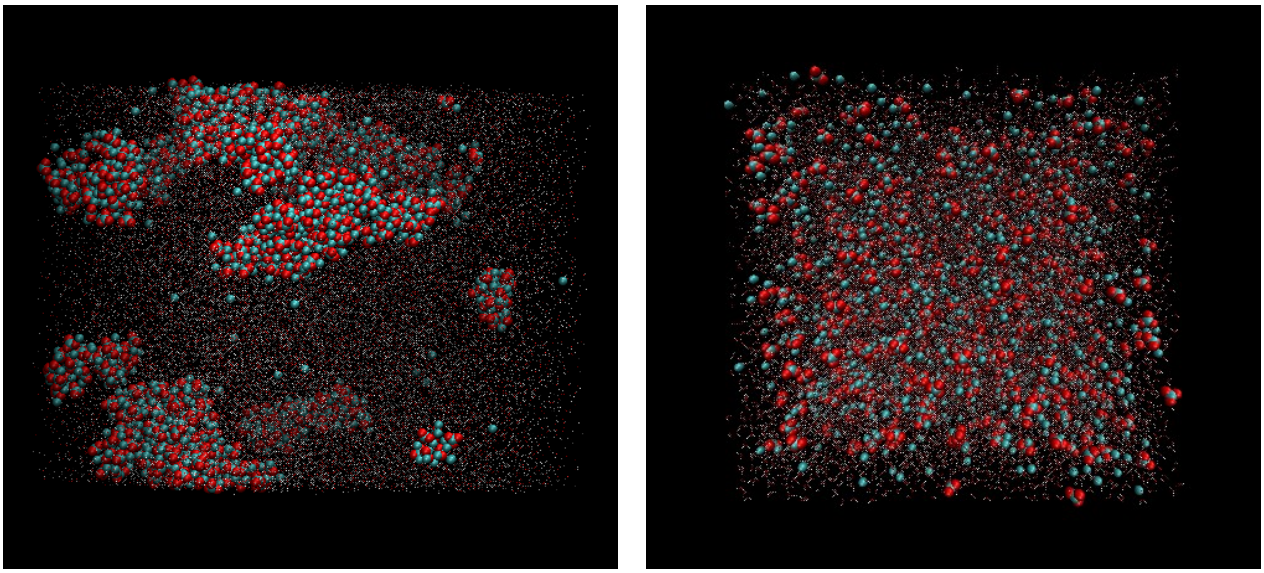

Figure 5.5. The 1m Na_2CO_3 salt solution aggregation as observed with the standard SPC/E water model (Left) and with the modified SPC/E water model (Right). Sodium ions are colored cyan and oxygen atoms of the carbonate ions are displayed in red. The charge distribution for the carbonate ion is C=0.4 and O = -0.8.

Table 5.4. A comparison of 2 m NaMHP salt solutions using the SPC/E water and the modified SPC/E water models. OT is the terminal oxygen atoms in the MHP molecule. The units for the KB integrals are cm^3mol^{-1}.

H	O	P	OT		SPC/E			Modified SPC/E		
					Gcc	Gcw	Gww	Gcc	Gcw	Gww
				Exp	147	-12	-17	147	-12	-17
0.2	0	0.25	-0.817	MD	2657	-282	10	304	-46	-14

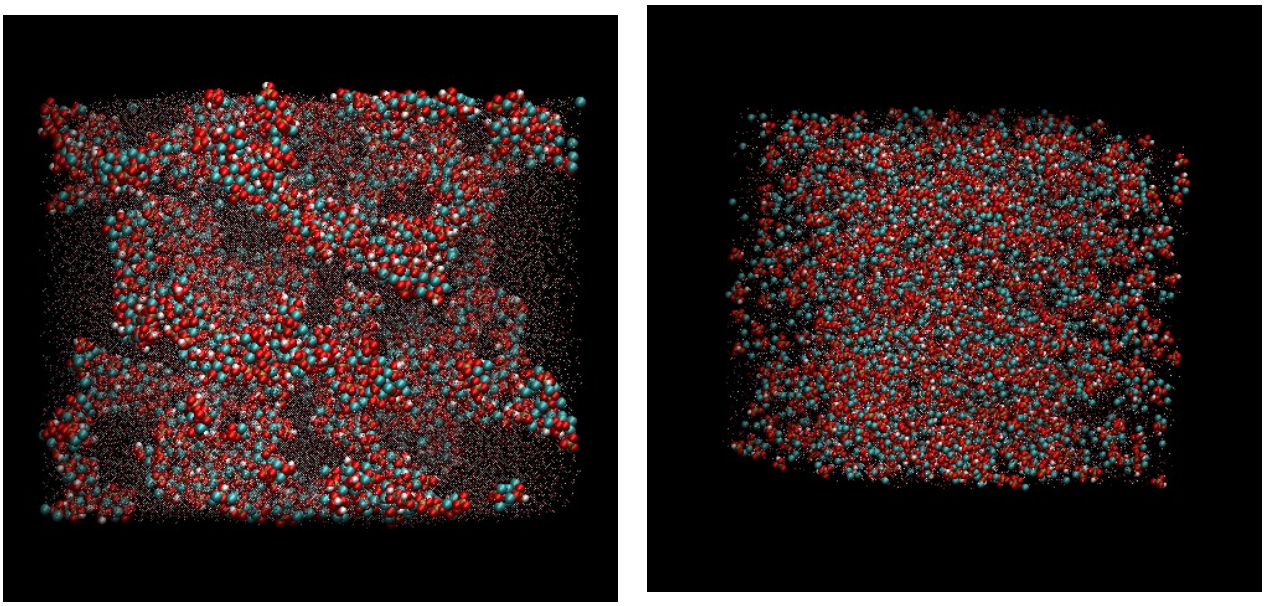

Figure 5.6. Simulations of 2 m NaMHP salt solution aggregation with the standard SPC/E water model (Left) and with the modified SPC/E water model (Right). Sodium ions are colored cyan and oxygen atoms of the MHP ions oxygen are displayed in red. The charge distribution for the MHP ion is as displayed in **Table 5.4**.

As mentioned elsewhere in this thesis, the MD simulations of 1m Na_2HPO_4 salt solutions with the standard SPC/E water model displays unrealistic aggregation behavior at finite, experimentally soluble, concentrations. However, once the water model is changed to the modified SPC/E model, which includes electronic screening effects, this problem appears to be solved. Also, the modified SPC/E water salt solution can be used to reproduce the KB integrals close to the experimental values. The KBI values are then sensitive to the charge distributions as displayed in Table 5.4 and

Figure 5.6.

Table 5.5. Comparison of 0.5 m Na_3PO_4 salt solutions using the SPC/E water and the modified SPC/E water model. The P atom charge systemically changed from positive value to a negative value. The charge distribution with +3.5 on the P atom provides the lowest KBI (G_{cc}) value for ion-ion interactions in the modified SPC/E water models. The units for the KB integrals are cm^3mol^{-1}.

P	O		SPC/E			Modified SPC/E		
			G_{cc}	G_{cw}	G_{ww}	G_{cc}	G_{cw}	G_{ww}
		Exp	697	-1	-17	697	-1	-17
4	-1.75	MD	6073	13	-16	1899	4	-17
3.5	-1.625	MD	4547	15	-17	995	8	-17
3.25	-1.5625	MD	6625	2	-17	1924	4	-17
3	-1.5	MD	11242	-8	-17	1437	4	-17
2	-1.25	MD	6171	-16	-17	3097	-3	-17
1.5	-1.125	MD	9478	-73	-17	4445	-24	-17
1	-1	MD	6741	-57	-17	6388	-72	-16
-1.5	-0.375	MD	11786	-158	-15	12186	-240	-13

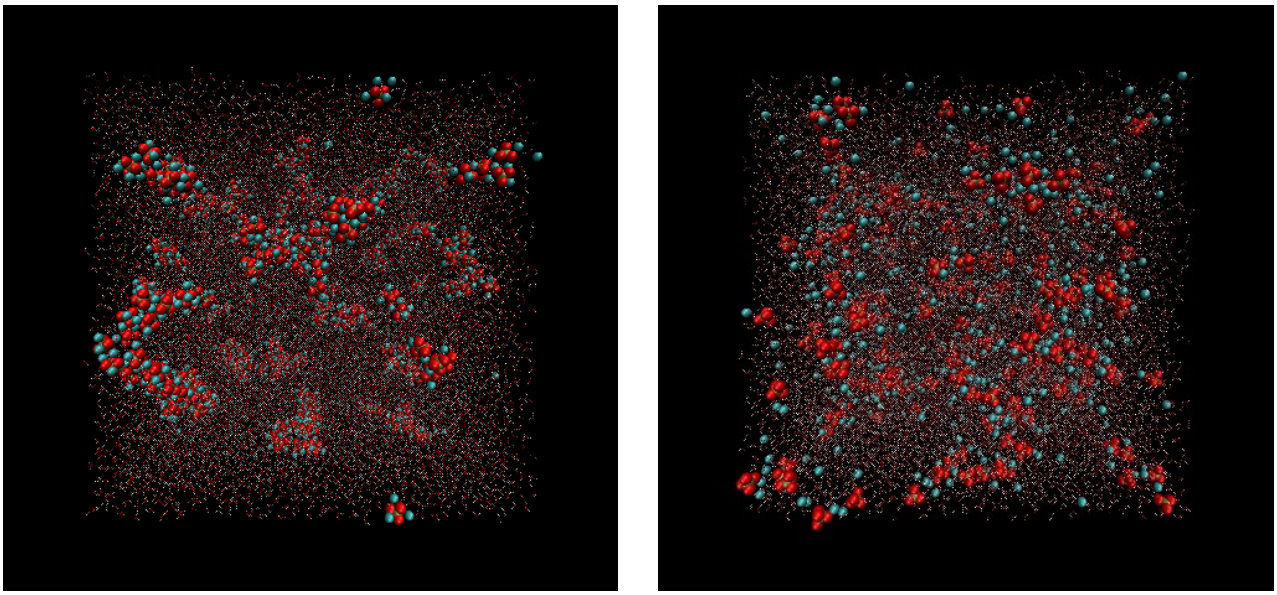

Figure 5.7. simulations of 0.5m Na_3PO_4 salt solution aggregation with the standard SPC/E water model (Left) and with the modified SPC/E water model (Right). Sodium ions are colored cyan and oxygen atoms of the phosphate ions are displayed in red. The charge distribution for the phosphate ion is P = 3.5 and O = -1.625.

For the Na_3PO_4 salts solutions using the modified SPC/E water model in the MD simulations still showed a slight excessive aggregation behavior. Nevertheless, the ion-ion KB integrals from the MD simulations clearly indicates that the modified SPC/E model, which includes electronic screening effects, significantly lowers the tendency to aggregate as displayed in Table 5.5 and Figure 5.7. The higher charge density could be a reason that the phosphate ion simulations did not to totally resolve the aggregation behavior completely in the standard MD simulations. The charge distributions attempted for the phosphate ion show that high negative charges on the oxygen atom help to lower the aggregation of the ions in the MD simulations. This is indicated by a lower positive value for the ion-ion KB integrals for the modified SPC/E water model simulations as displayed in Table 5.5. In comparison, the standard SPC/E water model

Na_3PO_4 salt simulations did not display any ability to reproduce the experimental KBI for the different charge distributions of the phosphate ions. Here, the ion aggregation is so severe that changes in the charge distribution have little effect on the resulting integrals.

The diffusion coefficients of ions observed in the MD simulations of the SPC/E water model and the modified SPC/E water systems are displayed in Table 5.6, Table 5.7 and Table 5.8, for aqueous salt solutions of 1 m Na_2CO_3, 2 m Na_2HPO4 and 0.5 m Na_3PO_4, respectively. The modified SPC/E water model clearly increases the diffusion rates of the ions in the MD simulations. As expected, the water molecules diffusion is not affected by modifications to the SPC/E water model. Hence, the water diffusion rate do not display any significant changes in the SPC/E model and modified SPC/E water model for a given salt solution. The sodium cation displays a roughly fivefold increase in diffusion rate when using the modified SPC/E water model with electronic screening. Furthermore, the anion diffusion is systematically affected by the charge density. Although the CO_3^{2-} ion and the HPO_4^{2-} ion have the negative two net charge, the planer flat CO_3^{2-} ion with a high charge density resulted in a 4.93 fold increase, whereas the larger HPO_4^{2-} ion with a lower charge density displayed a 10.3 fold increase when using the modified SPC/E water model. Furthermore, the PO_4^{3-} ion with the highest charge density, compare to the other two anions, displayed a lower 2.65 fold increase when using the modified SPC/E water model.

Table 5.6. The diffusion coefficient of cation, anion and water in aqueous 1 m Na_2CO_3 salt solutions using the SPC/E and modified SPC/E water systems, together with the ratio between the diffusion coefficients of ions in the modified SPC/E and standard SPC/E water systems

	Diffusion Coefficient / 10^{-9} m^2s^{-1}		Ratio
1 m Na_2CO_3 salts solutions	SPC/E	Modified SPC/E	
Na+	0.24	1.36	5.53
CO_3^{2-}	0.23	1.10	4.93
Water	2.27	2.32	1.03

Table 5.7. The diffusion coefficient of cation, anion and water in aqueous 2 m Na_2HPO_4 salt solutions using the SPC/E and modified SPC/E water systems, together with the ratio between the diffusion coefficients of ions in the modified SPC/E and standard SPC/E water systems

	Diffusion Coefficient / 10^{-9} m^2s^{-1}		Ratio
2 m Na_2HPO4 salts solutions	SPC/E	Modified SPC/E	
Na+	0.096	0.57	5.9
HPO_4^{2-}	0.081	0.84	10.3
Water	1.6	1.7	1.1

Table 5.8. The diffusion coefficient of cation, anion and water in aqueous 0.5 m Na_3PO_4 salt solutions using the SPC/E and modified SPC/E water systems, together with the ratio between the diffusion coefficients of ions in the modified SPC/E and standard SPC/E water systems

	Diffusion Coefficient / 10^{-9} m^2s^{-1}		Ratio
0.5 m Na_3PO_4 salts solutions	SPC/E	Modified SPC/E	
Na+	0.027	0.13	4.81
PO_4^{3-}	0.023	0.061	2.65
Water	0.32	0.29	0.91

5.5. Conclusions

A modified SPC/E water model employing a dipole moment closer to the experimental value leads to reduced ion clustering and aggregation behavior in MD simulations. For aqueous salt solutions of Na_2CO_3 and Na_2HPO_4 the modified SPC/E water completely solved the aggregation problem observed in the standard SPC/E simulations. However, for the salt solutions of Na_3PO_4, although modified SPC/E water lowered the aggregation behavior, it did not completely solve the issue. Furthermore, for the modified water model one regained the sensitivity of the KBI values to the charge distribution. Therefore, future attempts to quantify the ion and water interactions using KB integrals might be successful using such a modified water model. Unfortunately, although using a modified water model is a promising approach to resolve the unrealistic aggregation behavior of high charge ions in the MD simulations, this would require a reparameterization of all ion force fields currently compatible with standard water models. Furthermore, even using a water model which accounts for electronic screening effects, the scaling the cation to water interactions strength was still needed to (crudely) model polarization effects of ions in classical MD simulations.

References

1. Leontyev, I.; Stuchebrukhov, A., *Physical Chemistry Chemical Physics* **2011,** *13* (7), 2613.

2. Leontyev, I. V.; Stuchebrukhov, A. A., *Journal of Chemical Theory and Computation* **2012,** *8*, 3207-3216.

3. Leontyev, I. V.; Stuchebrukhov, A. A., *The Journal of Chemical Physics* **2014,** *141*, 014103.

4. Leontyev, I. V.; Stuchebrukhov, A. A., *The Journal of Chemical Physics* **2009,** *130*, 085102.

5. Fuentes-Azcatl, R.; Barbosa, M. C., *The Journal of Physical Chemistry B* **2016,** *120*, 2460-2470.

6. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., *The Journal of Chemical Physics* **1983,** *79*, 926.

7. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P., *The Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

8. Mahoney, M. W.; Jorgensen, W. L., *The Journal of Chemical Physics* **2000,** *112*, 8910.

9. Alder, B. J.; Wainwright, T. E., *The Journal of Chemical Physics* **1959,** *31* (2), 459-466.

10. Stillinger, F. H.; Rahman, A., *The Journal of Chemical Physics* **1974,** *60*, 1545-1557

11. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Clarendon Press: Oxford [England] :, 1989; p 385.

12. Grossfield, A.; Ren, P.; Ponder, J. W., *Journal of the American Chemical Society* **2003,** *125* (50), 15671-15682.

13. Patel, S.; Mackerell, A. D.; Brooks, C. L., *Journal of Computational Chemistry* **2004,** *25*, 1504-1514.

14. Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T., *The Journal of Physical Chemistry B* **2010,** *114*, 2549-2564.

15. Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P., *Journal of Chemical Theory and Computation* **2013,** *9*, 4046-4063.

16. Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., *Journal of Chemical Theory and Computation* **2013,** *9*, 5430-5449.

17. Ponder, J. W.; Case, D. A., *Advances in Protein Chemistry* **2003,** *66*, 27-85.

18. Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B., *Journal of Chemical Theory and Computation* **2010,** *6*, 774-786.

19. Spackman, M. A., *Chemical Reviews* **1992,** *92*, 1769-1797.

20. Vega, C.; Monson, P. A., *Molecular Physics* **1995,** *85*, 413-421.

21. Kusalik, P. G.; Patey, G. N., *The Journal of Chemical Physics* **1988,** *88*, 7715.

22. Ichiye, T., Water in the Liquid State: A Computational Viewpoint. In *Advances in Chemical Physics*, John Wiley & Sons, Inc.: 2014; pp 161-200.

23. Schropp, B.; Tavan, P., *Journal of Physical Chemistry B* **2008,** *112*, 6233-6240.

24. Weerasinghe, S.; Smith, P. E., *The Journal of Physical Chemistry B* **2003,** *107*, 3891-3898.

25. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R., *Computer Physics Communications* **1995,** *91*, 43-56.

26. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., *The Journal of Chemical Physics* **1984,** *81*, 3684.

27. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

28. Darden, T.; York, D.; Pedersen, L., *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

29. Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A., *Journal of applied crystallography* **2009,** *42*, 726-729.

30. Weerasinghe, S.; Smith, P. E., *The Journal of Chemical Physics* **2003,** *119*, 11342.

31. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E., *Journal of Chemical Theory and Computation* **2011,** *7* (5), 1369-1380.

Chapter 6 - Conclusions and Future Directions

The parameterization of alkaline earth ions for the KBFF was considered successful. The parameters developed in this model can reproduce the experimental KBIs and corresponding experimental thermodynamic properties in MD simulations. The parameterization of nitrate, perchlorate, dihydrogenphosphate, and sulfate anions provided reasonable models for biomolecular simulations. Furthermore, all these ion parameters result in acceptable diffusion and dielectric properties too. The dihydrogenphosphate ion parameters can be used to model structural components of nucleic acids and lipids using the KBFF approach. However, parametrization of carbonate, monohydrogenphosphate and phosphate was unsuccessful due to excessive ion clustering in MD simulations. Using a higher permittivity, and a water model with a dipole moment closer to the experimental value, to include electron screening effects helped to decrease excessive ion clustering and might be a step in right direction.

As for future directions, a set of MD simulations could be performed with other commonly used biomolecular force field models to determine what are the unique properties captured by the KBFF models, compared to other models of aqueous salt solutions. Furthermore, MD simulations can now be performed to investigate the interactions between ions and biomolecules such as proteins, DNA, RNA and lipids. This can help to determine whether the new ion models developed for the KBFF models can resolve issues observed with other force fields, such as excessive ion clustering, cation to phosphate backbone structural artifacts, etc. One can also attempt to understand the mechanism of ion paring and the role of the counter ion in the vicinity of biomolecules. Solvation free energies and entropies of hydration could be calculated for ions using MD simulations to compare with experimental values. A fine balance between the solvation free energy and the entropy of hydration might lead to an explanation of ion specific effects. The size

of the solvation shell can provide an effective size for an ion, that could then be correlated with binding sites of biomolecules to explain specificity. This might help us to understand whether the first solvation shell, or multiple tightly bound solvation shells, around an ion collectively interact with binding sites to produce specificity. Potential of mean force (PMF) studies between cations and anions could be used to quantify and compare stabilities of contact ion pairs and solvent separated ion pairs in different types of aqueous salt solutions. The impact of the presence of charged residues or biomolecules on the PMF also could be explored.

Furthermore, properties such as the residence time of water molecules in the first solvation shell of the ions can be calculated and compared with experimental values. This may help to provide insights into the binding affinity and exchange dynamics of water molecules in these solvation shells. Hence, the impact of changing ion size and negativity on the affinity of water and ion interactions could then be compared. MD simulations of N-methylacetamide with different types of ions could be used to evaluate peptide to ion interactions in aqueous environments. The rdfs could be used to study whether cations prefer to interact with counter ions or with the carbonyl oxygen of the amide bond. One could also attempt to correlate these interactions with ion properties such as the charge density or number of solvation shells. Negatively charged amino acids, such as Glu and Asp, interactions with divalent cations simulations could be used to investigate possible ion to side chain interactions. The same type of study can be performed with positively charged amino acids, such as Lys and Arg, with the oxo anions. Hence, the interactions of many different ion types with all the amino acids side chains could be explored. The preferential solvation, which can be calculated using KB integrals, can be used to quantify and correlate the role of ions in molecular crowding, cosolvent induced protein denaturation, pressure denaturation, Hofmeister effects etc.

Appendix A - Gaussian and non-Gaussian fluctuations in pure classical fluids

AIP PUBLISHING LLC LICENSE TERMS AND CONDITIONS

Mar 17, 2017

This Agreement between Nawavi Naleem ("You") and AIP Publishing LLC ("AIP Publishing LLC") consists of your license details and the terms and conditions provided by AIP Publishing LLC and Copyright Clearance Center.

License Number	4071401189831
License date	Mar 17, 2017
Licensed Content Publisher	AIP Publishing LLC
Licensed Content Publication	Journal of Chemical Physics
Licensed Content Title	Gaussian and non-Gaussian fluctuations in pure classical fluids
Licensed Content Author	
Licensed Content Date	Mar 7, 2017
Licensed Content Volume	146
Licensed Content Issue	9
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (< 800 words)
Will you be translating?	No
Title of your thesis / dissertation	Molecular dynamics simulations of aqueous ion solutions
Expected completion date	Apr 2017
Estimated size (number of pages)	200
Requestor Location	Nawavi Naleem 2020 Tunstall Cir Apt 28 MANHATTAN, KS 66502 United States Attn: Nawavi Naleem
Billing Type	Invoice
Billing Address	Nawavi Naleem 2020 Tunstall Cir Apt 28 MANHATTAN, KS 66502 United States Attn: Nawavi Naleem
Total	0.00 USD
Terms and Conditions	

AIP Publishing LLC -- Terms and Conditions: Permissions Uses

AIP Publishing hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

1. You agree to include the following copyright and permission notice with the reproduction of the Material:"Reprinted from [FULL CITATION], with the permission of AIP Publishing." For an article, the credit line and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your responsibility to ensure that the material is original to AIP Publishing and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP Publishing, with permission, from another source. Under no circumstances does AIP Publishing purport or intend to grant permission to reuse material to which it does not hold appropriate rights.
   You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes.
3. The foregoing license shall not take effect unless and until AIP Publishing or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
4. AIP Publishing or Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP Publishing, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
5. AIP Publishing makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP Publishing, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
6. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP Publishing. This license may not be amended except in a writing signed by the party to be charged.
7. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

This Agreement shall be governed by and construed in accordance with the laws of the State of New York. Both parties hereby submit to the jurisdiction of the courts of New York County for purposes of resolving any disputes that may arise hereunder.

V1.1

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.

THE JOURNAL OF CHEMICAL PHYSICS **146**, 094509 (2017)

Gaussian and non-Gaussian fluctuations in pure classical fluids

Nawavi Naleem, Elizabeth A. Ploetz,[a] and Paul E. Smith
Department of Chemistry, Kansas State University, 213 CBC Building, 1212 Mid Campus Dr. North, Manhattan, Kansas 66506-0401, USA

(Received 21 November 2016; accepted 12 February 2017; published online 6 March 2017)

The particle number, energy, and volume probability distributions in the canonical, isothermal-isobaric, grand canonical, and isobaric-isenthalpic ensembles are investigated. In particular, we consider Gaussian and non-Gaussian behavior and formulate the results in terms of a single expression valid for all the ensembles employing common, experimentally accessible, thermodynamic derivatives. This is achieved using Fluctuation Solution Theory to help manipulate derivatives of the entropy. The properties of the distributions are then investigated using available equations of state for fluid water and argon. Purely Gaussian behavior is not observed for any of the state points considered here. A set of simple measures, involving thermodynamic derivatives, indicating non-Gaussian behavior is proposed. A general expression, valid in the high temperature limit, for small energy fluctuations in the canonical ensemble is provided. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4977455]

I. INTRODUCTION

Most applications of equilibrium statistical mechanics focus on the determination of ensemble averages, e.g., pressure, density, and chemical potential, that correspond to first derivatives of the thermodynamic potentials.[1] In the thermodynamic limit, the same results are then obtained for all ensembles. In contrast, fluctuations are often only used to illustrate that the width of the energy distribution, for a system in the canonical ensemble, for example, is negligible when compared to the mean energy of the system.[1,2] However, this implies that fluctuations are relatively unimportant. This is not true and many thermodynamic properties can be related to the fluctuations, although the fluctuations then depend on the ensemble of interest, i.e., the thermodynamic constraints placed on the system. In particular, the second moments of the corresponding probability distributions can be directly related to second derivatives of the thermodynamic potentials—and include the important response functions.[2,3]

The study of equilibrium fluctuations has a long history and the interested reader is referred to a comprehensive review by Ruppeiner for more details.[4] Equilibrium fluctuations can be related to a series of thermodynamic properties (derivatives), and play an integral role in a variety of fluctuation-dissipation theorem based relationships,[5,6] and other fluid behaviors.[7,8] The majority of approaches involve an expansion of the entropy. The most common treatment of fluctuations is provided by Landau and Lifshitz,[9] while a more extensive discussion was recently provided by Mishin.[10] However, it is rare to expand beyond second order, which limits the resulting probability distribution to be Gaussian in nature. This is the distribution often expected in the thermodynamic limit.[4] Ruppeiner and others have indicated that there are significant problems with (energy) conservation and non-covariant behavior of entropy expansions beyond the Gaussian approximation.[4,11,12] Nevertheless, there is interest in the properties of the resulting non-Gaussian distributions.[4,12–14] The third and higher moments of these distributions are finite and have been developed by Greene and Callen.[15] They are important for understanding higher thermodynamic derivatives, especially close to the critical point, which characterize the behavior of fluids and their mixtures. Unfortunately, knowing the moments does not necessarily lead to the underlying distribution.

In principle, one can directly measure the fluctuations experimentally. For instance, particle number (density) fluctuations in fluids can be measured using light scattering.[16,17] It is unlikely, however, that these methods can be easily extended to provide higher moments of the distributions characterizing any non-Gaussian behavior. Alternatively, the fluctuations can be directly related to thermodynamic derivatives and properties of fluids that are relatively easy to determine. In the Gaussian approximation, the distributions are characterized by their covariance, corresponding to fluctuations and cross fluctuations between pairs of variables, and the covariance matrix is then directly related to second derivatives of the thermodynamic potentials.[2,10] Non-Gaussian behavior is then manifested in the deviations from Gaussian behavior characterized by the coskewness of the distribution, corresponding to correlations between three variables, and can be directly related to third derivatives of the thermodynamic potentials.[15] This is the approach taken here.

Recently, we have been using Fluctuation Solution Theory (FST) to investigate the properties of liquids and liquid mixtures.[14,18–20] FST provides an alternative view of the commonly studied closed system thermodynamic properties in terms of fluctuating quantities occurring in an equivalent Grand Canonical Ensemble (GCE) system. FST represents a generalization of the Kirkwood-Buff theory of solutions,[21–23] to include energy fluctuations.[19] By examining the corresponding fluctuations, one can obtain further insight into the properties of mixtures and simple fluids.[14,24] It is well known that

[a]Present address: Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205, USA.

particle and energy pair fluctuations in the GCE can be related to the response functions. However, it is less common to use the particle-energy triplet (or higher) fluctuations to investigate higher derivatives of the response functions. Nevertheless, the third moments of the bivariate particle-energy joint distribution in the GCE can be useful and quantify the non-Gaussian behavior of the underlying probability distribution.[14]

Here, we use FST to help provide the second and third derivatives of the entropy for use in the usual entropy expansion to provide a general non-Gaussian particle, volume, and energy probability distribution that is applicable to any ensemble. The initial steps essentially follow previous entropy expansion approaches,[9,15] although here we start from the probability distributions for a series of ensembles. These general expressions are then developed, using FST, to provide the Gaussian and non-Gaussian distributions in terms of common thermodynamic derivatives available experimentally. The nature of the resulting distribution is then investigated.

II. THEORY

A. Fluctuations

The thermodynamic potentials associated with the four main ensembles of statistical mechanics can be written as[1,2]

$$\begin{aligned} S/R &= \ln \Omega(N_1, V, E), \\ -\beta A &= -\beta \langle E \rangle + \langle S \rangle / R = \ln Q(N_1, V, T), \\ -\beta G &= -\beta \langle E \rangle - \beta p \langle V \rangle + \langle S \rangle / R = \ln \Delta(N_1, p, T), \\ \beta pV &= -\beta \langle E \rangle + \beta \mu_1 \langle N_1 \rangle + \langle S \rangle / R = \ln \Xi(\mu_1, V, T), \end{aligned} \tag{1}$$

where $\beta = 1/(RT)$ and R is the gas constant. The symbols Ω, Q, Δ, and Ξ represent the partition functions for the microcanonical (NVE), canonical (NVT), isothermal-isobaric/Gibbs (NpT), and grand canonical (μVT) ensembles, respectively. These are then related to the number of particles (N_1), the volume (V), the internal energy (E), the entropy (S), the pressure (p), the absolute temperature (T), the chemical potential (μ_1), the Helmholtz free energy (A), and the Gibbs free energy (G) of the system. The angular brackets denote the appropriate ensemble averages, and it is understood that for the entropy $\langle S \rangle = S(\langle N_1 \rangle, \langle V \rangle, \langle E \rangle)$. Formally, the entropy described above corresponds to an entropy maximum and therefore the average internal energy, $\langle E \rangle$, should really be the most probable internal energy, E^*. However, we can assume these to be equivalent in the thermodynamic limit (see later discussion).

The conditional probability that a system has a particular energy, volume, and/or particle number between E and $E + dE$, etc., is then given by

$$\begin{aligned} P(E|N_1, V)dE &= \frac{e^{-\beta E} e^{S/R}}{Q} dE = e^{-\beta\delta E + \delta S/R} dE, \\ P(V, E|N_1)dVdE &= \frac{e^{-\beta E} e^{-\beta pV} e^{S/R}}{\Delta} dVdE \\ &= e^{-\beta\delta E - \beta p\delta V + \delta S/R} dVdE, \\ P(N_1, E|V)dN_1dE &= \frac{e^{-\beta E} e^{\beta\mu_1 N_1} e^{S/R}}{\Xi} dN_1dE \\ &= e^{-\beta\delta E + \beta\mu_1\delta N_1 + \delta S/R} dN_1dE \end{aligned} \tag{2}$$

for the NVT, NpT, and μVT ensembles, respectively. Here, $\delta X = X - \langle X \rangle$ represents a fluctuation in the property $X = N_1$, V, E, or S. Clearly, a general probability distribution function (pdf) can be written as

$$\ln P(N_1, V, E) = \beta\mu_1\delta N_1 - Z\rho_1\delta V - \beta\delta E + \delta S/R, \tag{3}$$

where $\rho_1 = N_1/V$ is the number density, $Z = \beta pV_1$ is the compressibility factor of the pure fluid, and V_1 is the molar volume. This expression is applicable to all the ensembles, with the understanding that the ensembles provide the same ensemble average thermodynamic properties, and that δN_1 is zero in a closed system, etc.

The presence of the entropy in Equation (3) is problematic. In order to develop this expression further, knowledge of the variation in entropy (density of states) with the energy, volume, and/or particle number is required.[15] The entropy variations can be evaluated by noting that a Taylor expansion of the entropy around $\langle E \rangle = E^*$, etc., provides

$$\delta S/R = S^{(1)} + \frac{1}{2}S^{(2)}N_1^{-1} + \frac{1}{6}S^{(3)}N_1^{-2} + \cdots, \tag{4}$$

where the N_1's have been included to make the corresponding entropy derivatives that appear in the $S^{(n)}$ terms intensive. We will assume that these derivatives exist and that the expansion is valid, which is generally true if we stay away from phase boundaries and the critical point. The $S^{(n)}$'s describe a series of increasingly more complicated terms given by

$$\begin{aligned} S^{(1)} &= \sum_x S_x \delta x, \\ S^{(2)} &= \sum_{x,y} S_{xy} \delta x \delta y, \\ S^{(3)} &= \sum_{x,y,z} S_{xyz} \delta x \delta y \delta z, \end{aligned} \tag{5}$$

where x, y, or z refer to each of the three dimensionless fluctuations given by βE, $\rho_1 V$, or N_1, corresponding to derivatives of the entropy with respect to E, V, or N_1. Here we will only consider the terms up to and including third derivatives of the entropy. The sums in $S^{(1)}$, $S^{(2)}$, and $S^{(3)}$ involve 3, 9, and 27 terms with 3, 6, and 10 unique entropy derivatives (S_x, S_{xy}, S_{xyz}), respectively. The exact form of the dimensionless intensive entropy derivatives is described in Appendix A.

The first and second law for an open system at equilibrium with only pressure-volume work can be written as[25]

$$dS/R = -\beta\mu_1 dN_1 + Z\rho_1 dV + \beta dE. \tag{6}$$

This provides the first derivatives of the entropy given in Appendix A. Using these in Equations (3) and (4) leads to the general expression

$$\ln P(N_1, V, E) = \frac{1}{2}S^{(2)}N_1^{-1} + \frac{1}{6}S^{(3)}N_1^{-2} + \cdots. \tag{7}$$

Truncation after the $S^{(2)}$ term results in a Gaussian distribution, and represents the stopping point for many studies of fluctuations, while inclusion of the $S^{(3)}$ (and higher) term will lead to non-Gaussian distributions. The resulting probabilities are not normalized but it is clear that $P(\langle N_1 \rangle, \langle V \rangle, \langle E \rangle) = 1$, irrespective of how many terms are included in Equation (7). It should be noted that, using an entropy expansion approach beyond Gaussian, it is problematic to assume that the mean values

of N_1, V, and E are also the most probable values. Indeed, previous studies have shown that the GCE distributions are not Gaussian (or symmetric) and therefore this assumption is not justified for the present purposes.[14,20] This point will also be discussed further in Sec. III.

In order to develop Equations (5) and (7), one requires higher derivatives of the entropy. To evaluate these second and third derivatives, we turn to FST. We will see below that not all these derivatives are available experimentally. Hence, further manipulation of the derivatives will also be required.

B. Fluctuation Solution Theory

The next step requires the conversion of the entropy derivatives to more common thermodynamic derivatives or combinations of derivatives. In our opinion, the second, and especially the third, derivatives can be evaluated most easily using a FST based approach utilizing the corresponding pair and triplet fluctuations for the equivalent GCE. To achieve this, we define a series of fluctuating pair and triplet densities

$$\begin{aligned} B_{XY} &\equiv \langle \delta X \delta Y \rangle / V, \\ C_{XYZ} &\equiv \langle \delta X \delta Y \delta Z \rangle / V, \end{aligned} \tag{8}$$

where X, Y, $Z = N_1$ (indicated by a subscript 1), E, ε, or E'. The angular brackets denote a GCE average. While other fluctuating quantities will appear later, the most useful energy quantities for our initial manipulations are

$$\begin{aligned} \varepsilon &\equiv E - N_1 H_1, \\ E' &= E - N_1 E_1, \end{aligned} \tag{9}$$

where ε and E' are the excess energies, E and N_1 are the instantaneous (in the time average sense) internal energy and particle number for the equivalent GCE system, H_1 is the average molar enthalpy, and E_1 is the average molar internal energy. The main reason for adopting the above combinations is to negate the fact that the energy fluctuations in the GCE are sensitive to the zero of energy.[24] The above pair fluctuation densities can be converted to a set of corresponding dimensionless counterparts (indicated by lower case letters) by

$$\begin{aligned} b_{11} &\equiv B_{11}/\rho_1, \\ b_{1\varepsilon} &\equiv \beta B_{1\varepsilon}/\rho_1, \\ b_{\varepsilon\varepsilon} &\equiv \beta^2 B_{\varepsilon\varepsilon}/\rho_1. \end{aligned} \tag{10}$$

Similar expressions can be written for the C's and then provide the corresponding dimensionless c's. The above fluctuating quantities are related to common thermodynamic properties of fluids via[18,19,26,27]

$$\begin{aligned} \frac{p\kappa_T}{Z} &\equiv \frac{p}{\rho_1 Z}\left(\frac{\partial \rho_1}{\partial p}\right)_T = b_{11}, \\ T\alpha_p &\equiv -\frac{T}{\rho_1}\left(\frac{\partial \rho_1}{\partial T}\right)_p = -b_{1\varepsilon}, \\ \frac{C_{p,m}}{R} &\equiv \frac{1}{R}\left(\frac{\partial H_1}{\partial T}\right)_p = b_{\varepsilon\varepsilon}, \\ \frac{C_{V,m}}{R} &\equiv \frac{1}{R}\left(\frac{\partial E_1}{\partial T}\right)_{\rho_1} = b_{\varepsilon\varepsilon} - \frac{b_{1\varepsilon}^2}{b_{11}}, \end{aligned} \tag{11}$$

where κ_T is the isothermal compressibility, α_p is the thermal expansion coefficient, $C_{p,m}$ is the isobaric molar heat capacity, and $C_{V,m}$ is the isochoric molar heat capacity. Hence, if one knows the b's for the GCE system, one also knows the response functions for the equivalent closed system. Likewise, if one knows the response functions, one also knows the pair fluctuations for an equivalent open GCE system. Finally, we note that the following combination of terms occurs repeatedly for the entropy derivatives and so we define

$$S_0 \equiv \frac{1}{b_{\varepsilon\varepsilon} b_{11} - b_{1\varepsilon}^2} = \frac{Z}{p\kappa_T}\frac{R}{C_{V,m}} \tag{12}$$

to help simplify many of the resulting equations. This particular combination of B's can be written in several ways

$$B_{\varepsilon\varepsilon} B_{11} - B_{1\varepsilon}^2 = B_{EE} B_{11} - B_{1E}^2 = B_{E'E'} B_{11} - B_{1E'}^2 \tag{13}$$

which also helps to simplify the process of obtaining many of the third derivatives.

Using the above relationships, the six unique dimensionless second derivatives of the entropy are given by (see Appendix B)

$$\begin{aligned} S_{NN} &= -S_0 b_{EE}, \quad S_{VV} = -S_0 b_{E'E'}, \quad S_{EE} = -S_0 b_{11}, \\ S_{NV} &= S_0 b_{EE'}, \quad S_{NE} = S_0 b_{1E}, \quad S_{VE} = -S_0 b_{1E'} \end{aligned} \tag{14}$$

in terms of the dimensionless particle and energy fluctuations. It should be noted that S_{NN}, S_{NV}, and S_{NE}, i.e., the particle number derivatives, are not available experimentally as the energy fluctuations in the GCE depend on the zero of energy. The fluctuations that involve just E' do not depend on the zero of energy and hence S_{VV}, S_{VE}, and S_{EE} are available.

The third derivatives of the entropy are more involved. They can be obtained from Equation (14) using the approach outlined in Appendix C. There are ten unique derivatives. In their most condensed form, they are given by

$$\begin{aligned} S_{NNN} &= -S_0^3 c_{nnn}, \\ S_{NNV} &= S_0^3 c_{nnv} + S_0 b_{EE}, \\ S_{NNE} &= S_0^3 c_{nne} \end{aligned} \tag{15}$$

with

$$\begin{aligned} S_{VVN} &= -S_0^3 c_{vvn} - 2S_0 b_{1E'}\beta E_1, \\ S_{VVV} &= S_0^3 c_{vvv} + 3S_0 b_{E'E'}, \\ S_{VVE} &= S_0^3 c_{vve} + 2S_0 b_{1E'} \end{aligned} \tag{16}$$

and

$$\begin{aligned} S_{EEN} &= -S_0^3 c_{een}, \\ S_{EEV} &= S_0^3 c_{eev} + S_0 b_{11}, \\ S_{EEE} &= S_0^3 c_{eee} \end{aligned} \tag{17}$$

and finally

$$S_{NVE} = -S_0^3 c_{nve} - S_0 b_{1E}, \tag{18}$$

where we have used the following definitions:

$$\begin{aligned} n &\equiv \beta E b_{1E} - N_1 b_{EE}, \\ v &\equiv \beta E' b_{1E'} - N_1 b_{E'E'}, \\ e &\equiv \beta E b_{11} - N_1 b_{1E}. \end{aligned} \tag{19}$$

While most of these derivatives are available experimentally, the particle number derivatives are not due to their dependence

on the zero of energy.[24] Note that, in the above volume derivatives of the entropy, there is an additional term involving pair fluctuations that will only contribute for isobaric ensembles.

III. RESULTS AND DISCUSSION

A. Gaussian distributions

Gaussian approximations for the particle, energy, and volume distributions are common in textbooks.[2,16] However, we have included them here for three main reasons. First, they illustrate the process of obtaining the distributions from the second entropy derivatives, which is then extended to include the third derivatives in Sec. III B. Second, the forms that most commonly appear in textbooks often treat multiple fluctuating variables independently, whereas in reality they are usually correlated. In many cases, this limitation is not a major problem as only the order of magnitude of the fluctuation is of interest. However, the appropriate multivariate Gaussian distribution can also be of importance. Third, an initial analysis of the *NVT*, *NpT*, and μVT results suggested that a single probability distribution valid for a variety of ensembles is available.

As mentioned previously, many of the entropy derivatives are not available experimentally. This indicates that some form of substitution is required to provide combinations of entropy derivatives that are experimentally available. Based on our previous applications of FST,[19] we will adopt the following dimensionless quantities:

$$\begin{aligned} V^E &\equiv \rho_1(V - N_1V_1), \\ H^E &\equiv \beta(E + \langle p\rangle V - N_1H_1), \\ S^E &\equiv (S - N_1S_1)/R \end{aligned} \tag{20}$$

representing an instantaneous excess volume, enthalpy, and entropy, respectively. Here, S_1 is the molar entropy and the values of N_1, V, E, and S can fluctuate depending on the ensemble. These excess quantities average to zero in all the main ensembles. The enthalpy involves the average pressure written as $\langle p\rangle$ to emphasize that pressure never fluctuates in these quantities regardless of the ensemble. Furthermore, the excess enthalpy definition avoids the zero of energy problem for open systems.[24] Before using the above quantities to develop the entropy expansion, we note that the corresponding quantities, $\rho_1\delta V = \delta V^E + \delta N_1$ and $\beta\delta E = \delta H^E - Z\delta V^E + \beta E_1\delta N_1$, when used in Equation (6) give

$$\delta S^E = \delta H^E \tag{21}$$

for small fluctuations. Hence, the excess entropy and enthalpy fluctuations are equal and therefore δS^E could be used in place of δH^E in the following equations.

If we substitute the expressions for δE and δV given by Equation (20) into the expression for $S^{(2)}$ found in Equation (5) and then collect identical terms the result can be written as

$$\begin{aligned} S^{(2)} = {} & F_{VV}(\delta V^E)^2 + F_{HH}(\delta H^E)^2 + F_{NN}(\delta N_1)^2 \\ & + 2F_{VH}\delta V^E\delta H^E + 2F_{NV}\delta N_1\delta V^E + 2F_{NH}\delta N_1\delta H^E, \end{aligned} \tag{22}$$

where the F's involve combinations of second entropy derivatives. Interestingly, the F's corresponding to the particle number fluctuations are given by

$$\begin{aligned} F_{NN} &\equiv S_{NN} + S_{VV} + S_{EE}(\beta E_1)^2 + 2S_{NV} + 2S_{NE}\beta E_1 \\ &\quad + 2S_{VE}\beta E_1 = 0, \\ F_{NV} &\equiv S_{VV} - S_{EE}\beta E_1 Z + S_{NV} - S_{NE}Z \\ &\quad + S_{VE}(\beta E_1 - Z) = 0, \\ F_{NH} &\equiv S_{EE}\beta E_1 + S_{NE} + 2S_{VE} = 0 \end{aligned} \tag{23}$$

and subsequently disappear. The other F's simplify considerably to give

$$\begin{aligned} F_{VV} &\equiv S_{VV} - 2S_{VE}Z + S_{EE}Z^2 = -S_0 b_{\varepsilon\varepsilon}, \\ F_{VH} &\equiv S_{VE} - S_{EE}Z = -S_0 b_{1\varepsilon}, \\ F_{HH} &\equiv S_{EE} = -S_0 b_{11} \end{aligned} \tag{24}$$

all of which involve quantities that are available experimentally. Therefore, the FST based Gaussian probability distribution is then

$$\begin{aligned} \ln P(N_1, V, E) = -\frac{S_0}{2N_1}\Big[& b_{\varepsilon\varepsilon}(\delta V^E)^2 + 2b_{1\varepsilon}\delta V^E\delta H^E \\ & + b_{11}(\delta H^E)^2\Big] \end{aligned} \tag{25}$$

which can be expressed using common thermodynamic properties as

$$\begin{aligned} \ln P(N_1, V, E) = -\frac{1}{2N_1}\frac{C_{p,m}}{C_{V,m}}\Bigg[& \frac{(\delta V^E)^2}{p\kappa_T/Z} \\ & - 2T\alpha_p\frac{\delta V^E}{p\kappa_T/Z}\frac{\delta H^E}{C_{p,m}/R} + \frac{(\delta H^E)^2}{C_{p,m}/R}\Bigg]. \end{aligned} \tag{26}$$

The relationship is valid for all the common ensembles of interest. Specific expressions for each ensemble are provided below.

The normalized Gaussian probability distribution result for ensembles where at least one extensive variable is fixed, i.e., all but the μpT ensemble, can be written in a general form given by

$$P(N_1, V, E) = (2\pi|\mathbf{Q}|^{1/2})^{-1}e^{-\frac{1}{2}\mathbf{q}^{\mathrm{T}}\mathbf{Q}^{-1}\mathbf{q}}, \tag{27}$$

where the covariance matrix, $\mathbf{Q}$, involves the pair fluctuations arising from a FST analysis of the thermodynamics according to

$$\mathbf{Q} \equiv \begin{pmatrix} \langle\delta N_1\delta N_1\rangle & -\beta\langle\delta N_1\delta\varepsilon\rangle \\ -\beta\langle\delta N_1\delta\varepsilon\rangle & \beta^2\langle\delta\varepsilon\delta\varepsilon\rangle \end{pmatrix}. \tag{28}$$

Alternatively, in terms of common thermodynamic properties one has

$$\mathbf{Q} = N_1\begin{pmatrix} p\kappa_T/Z & T\alpha_p \\ T\alpha_p & C_{p,m}/R \end{pmatrix}, \tag{29}$$

where $|\mathbf{Q}| = C_V p\kappa_T/(RZ) = N_1/S_0$. The thermodynamic properties in the above matrix provide the covariances corresponding to the fluctuations in the excess volume and enthalpy as indicated in Equation (26). The general nature of the above distribution is then provided by the corresponding column vector which takes the form

$$\mathbf{q} \equiv \begin{pmatrix} \delta V^E \\ \delta H^E \end{pmatrix} \tag{30}$$

in terms of the excess volume and enthalpy. This then reduces to the following values for the main ensembles:

$$\mathbf{q} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}_{NVE} = \begin{pmatrix} 0 \\ \beta\delta E \end{pmatrix}_{NVT} = \begin{pmatrix} \rho_1\delta V \\ \beta\delta H \end{pmatrix}_{NpT} = \begin{pmatrix} -\delta N_1 \\ \beta\delta\varepsilon \end{pmatrix}_{\mu VT} = \begin{pmatrix} \rho_1\delta V \\ 0 \end{pmatrix}_{NpH} \tag{31}$$

although the NpH ensemble is usually only of significant interest for computer simulation studies.[28]

Consequently, a Gaussian result valid for all ensembles has been provided in terms of an excess volume and enthalpy. The excess volume and enthalpy then reduce to the appropriate fluctuating quantities for each ensemble, while the covariance matrix itself does not change. The fluctuating quantities are essentially those expected for all but the GCE. In this latter case, the above relationships clearly indicate that the excess energy ε is the relevant quantity in the GCE ensemble required to describe the common thermodynamic properties in the simplest manner. The above equations also emphasize the cross correlation between the fluctuations in the NpT and GCE that only disappears when $\alpha_p = 0$ and hence $C_{p,m} = C_{V,m}$. Finally, we note that the determinant of the 3×3 matrix of entropy derivatives formed from the relationships provided in Equation (14) can be shown to be singular, and hence a general multivariate Gaussian distribution function expression for the (unbounded) fluctuations cannot be written for the corresponding μpT ensemble (as expected).

B. Non-Gaussian distribution

The corresponding non-Gaussian distribution is significantly more complicated. In particular, while the number of fluctuating quantities remains reasonable, the number of corresponding thermodynamic derivatives increases substantially. Nevertheless, from the previous results, it would appear that a general distribution valid for multiple ensembles should be available. Again, we use the substitutions denoted in Equation (20) for the $S^{(2)}$ and $S^{(3)}$ terms. The $S^{(2)}$ term remains the same as before and we can use the result presented in Equation (25) or (26). We will factor the $S^{(3)}$ term into two contributions, one (P_V) involving just the pair fluctuations (b's) that only appears for isobaric ensembles, and one (P_3) involving just the triplet fluctuations (c's), as indicated in Equations (15)–(18). The same type of manipulations used in Sec. III A is required again. The details are provided in Appendix D.

This approach leads to a general expression for the distribution

$$\ln P(N_1, V, E) = P_2 + P_V + P_3, \tag{32}$$

where

$$\begin{aligned} P_2 &= -\frac{S_0}{2N_1}\left[b_{\varepsilon\varepsilon}(\delta V^E)^2 + 2b_{1\varepsilon}\delta V^E\delta H^E + b_{11}(\delta H^E)^2\right], \\ P_V &= \frac{S_0}{2N_1^2}\left[b_{\varepsilon\varepsilon}(\delta V^E)^3 + 2b_{1\varepsilon}(\delta V^E)^2\delta H^E + b_{11}\delta V^E(\delta H^E)^2\right] \\ &= -P_2\delta V^E/N_1, \\ P_3 &= \frac{S_0^3}{6N_1^2}\left[c_{sss}(\delta V^E)^3 + 3c_{sst}(\delta V^E)^2\delta H^E + 3c_{stt}\delta V^E(\delta H^E)^2 \right. \\ &\quad \left. + c_{ttt}(\delta H^E)^3\right] \end{aligned} \tag{33}$$

in terms of the fluctuations defined in Equation (20) and where

$$\begin{aligned} s &\equiv \beta\varepsilon b_{1\varepsilon} - N_1 b_{\varepsilon\varepsilon}, \\ t &\equiv \beta\varepsilon b_{11} - N_1 b_{1\varepsilon}. \end{aligned} \tag{34}$$

We note that the P_V term only contributes for isobaric ensembles (even if the particle number or energy fluctuates). The intensive properties defined in Equation (8) indicate that the pair and triplet fluctuations are proportional to V and therefore N_1. Consequently, the P_V and P_3 terms are not intensive and appear to become negligible as the system size becomes infinite. Hence, the real distribution then tends towards that of a Gaussian distribution. This is as expected, although the infinite system limit is complicated by the additional requirement of an infinite "bath" to maintain a constant chemical potential, temperature, or pressure.[11] However, as we show below, this does not mean that the third and higher moments, obtained upon integration, are then zero.

P_2 and P_V can be expressed in terms of thermodynamic properties using the results given in Equation (11). In order to express the above result for P_3 in terms of thermodynamic properties, we need to express the triplet fluctuations in terms of thermodynamic properties. This process is described in Appendix D. The final result is

$$S_0^3 c_{ttt} = \left(\frac{R}{C_{V,m}}\right)^2 J_1, \tag{35}$$

$$S_0^3 c_{stt} = -\left(\frac{R}{C_{V,m}}\right)^2\left[\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1} J_1 + \frac{C_{V,m}}{R}J_2\right], \tag{36}$$

$$S_0^3 c_{sst} = \left(\frac{R}{C_{V,m}}\right)^2\left[\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1}^2 J_1 + 2\frac{C_{V,m}}{R}\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1} J_2 + \frac{C_{V,m}}{R}\frac{Z}{p\kappa_T}J_3\right], \tag{37}$$

$$S_0^3 c_{sss} = -\left(\frac{R}{C_{V,m}}\right)^2 \times \left[\begin{aligned} &\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1}^3 J_1 + 3\frac{C_{V,m}}{R}\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1}^2 J_2 \\ &+ 3\frac{C_{V,m}}{R}\frac{Z}{p\kappa_T}\left(Z\frac{T}{p}\frac{\partial p}{\partial T}\right)_{\rho_1} J_3 + \left(\frac{C_{V,m}}{R}\right)^2\frac{Z}{p\kappa_T}J_4 \end{aligned}\right], \tag{38}$$

where the J_n's are given by

$$\begin{aligned} J_1 &\equiv \frac{c_{ttt}}{b_{11}b_{tt}} = 2 + \frac{T}{C_{V,m}}\left(\frac{\partial C_{V,m}}{\partial T}\right)_{\rho_1}, \\ J_2 &\equiv \frac{c_{1tt}}{b_{11}b_{tt}} = 1 - Z\frac{R}{C_{V,m}}\frac{T^2}{p}\left(\frac{\partial^2 p}{\partial T^2}\right)_{\rho_1}, \\ J_3 &\equiv \frac{c_{11t}}{b_{11}^2} = 1 + \frac{T}{\kappa_T}\left(\frac{\partial \kappa_T}{\partial T}\right)_{\rho_1}, \\ J_4 &\equiv \frac{c_{111}}{b_{11}^2} = 1 + \rho_1\left(\frac{\partial^2\rho_1}{\partial p^2}\right)_T\Big/\left(\frac{\partial\rho_1}{\partial p}\right)_T^2 \end{aligned} \tag{39}$$

and have been written so that the distribution is Gaussian when all the J_n's are zero. Furthermore, the coefficients of the J_n's

are always positive, in the absence of an isobaric density maximum, and therefore the J_n's also determine the sign of the c's. The various combinations of derivatives in Equation (39) characterize the deviation from Gaussian behavior. The distribution provided by Equations (32), (33), and (35)–(39) is valid for all the ensembles described in Section III A. When α_p is zero then so is the thermal pressure coefficient. Under these conditions, the expressions simplify greatly as only the final term on the right hand side of Equations (35)–(38) survives. The J_n's are clearly related to the GCE triplet fluctuations and hence finite triplet fluctuations lead to non-Gaussian behavior. This is to be expected, but now it is clear which triplet fluctuations are important and how they relate to more common thermodynamic derivatives. Finally, the value of J_3 is given by a thermodynamic derivative not commonly provided by a typical equation of state. An alternative expression that employs more traditional derivatives is given by

$$J_3 = 1 + T\left[\left(\frac{\partial^2\rho_1}{\partial p\partial T}\right)\left(\frac{\partial\rho_1}{\partial p}\right)_T - \left(\frac{\partial^2\rho_1}{\partial p^2}\right)_T\left(\frac{\partial\rho_1}{\partial T}\right)_p\right]\Big/\left(\frac{\partial\rho_1}{\partial p}\right)_T^2 \tag{40}$$

and was obtained using the relationships provided in Equation (D5).

C. Experimental non-Gaussian behavior

To investigate the non-Gaussian behavior further, we have extracted the J_n's from accurate equations of state for the complex fluid water,[29] and the simple fluid argon,[30] that provide the required thermodynamic derivatives. The results are presented in Figures 1 and 2 and also in Tables I and II. It is immediately apparent that there is no point on the phase diagrams where all four J_n's are zero. While several of the J_n's do change sign, particularly in the liquid and supercritical regions, they do so at different pressures and temperatures. Non-Gaussian behavior can be quantified by the J_n values. When the J_n's are all zero, one observes a simple Gaussian distribution. This places firm restrictions on the allowed values of the thermodynamic derivatives in Equation (39). The data for water presented in Table II clearly show that the distribution displays significant non-Gaussian character. Hence, these restrictions are unlikely to be met for real fluids at a single state point and Gaussian behavior in pure fluids does not appear to be possible. The ideal gas results are also included in Table I to emphasize that Gaussian behavior is also not observed even in this limiting case. In fact, ideal gases actually follow a Poisson distribution.[9,14] It should be noted that the J_n's and c's in Table I represent triplet fluctuations in the GCE. As these are

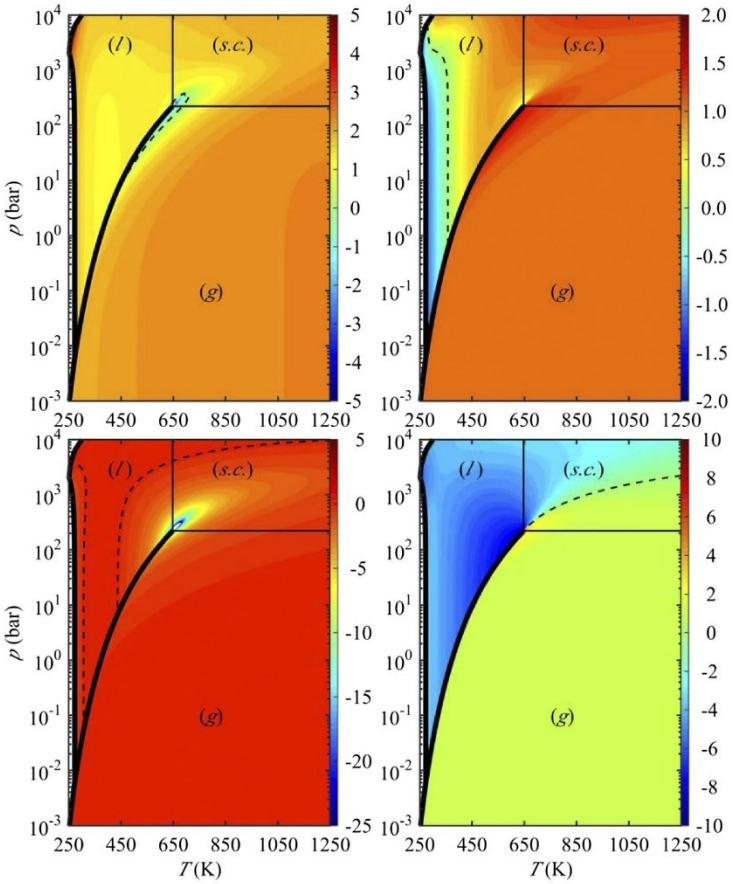

FIG. 1. Contour plots for the dimensionless non-Gaussian measures J_n as a function of p and T for fluid water. Top: J_1 (left) and J_2 (right). Bottom: J_3 (left) and J_4 (right). The thick solid black curves denote the phase boundaries. Zero contours are indicated with dashed black curves. The boundary of the supercritical region is marked by thin black lines. The data were generated using the IAPWS-95 equation of state,[29] as implemented in the NIST/ASME STEAM database version 2.2.[31]

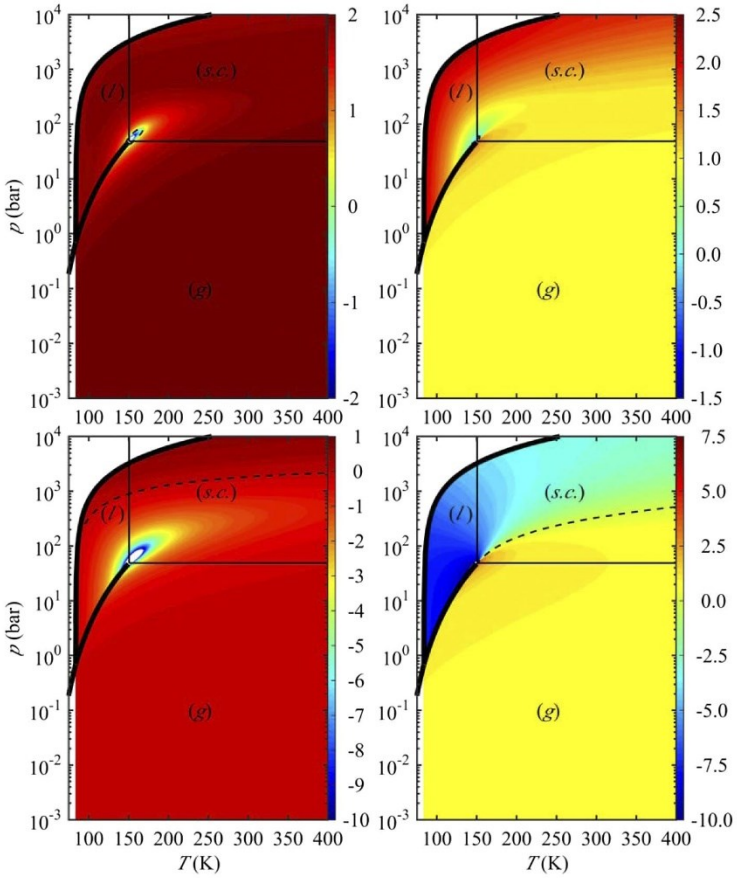

FIG. 2. Contour plots for the dimensionless non-Gaussian measures J_n as a function of p and T for fluid argon. Top: J_1 (left) and J_2 (right). Bottom: J_3 (left) and J_4 (right). The thick solid black curves denote the phase boundaries. Zero contours are indicated with dashed black curves. The boundary of the supercritical region is marked by thin black lines. The data were generated using the equation of state from Tegeler, Span, and Wagner,[30] as implemented in the NIST REFPROP database version 9.1.[32]

available and finite, this indicates that the third moments of the distribution, i.e., the measures of skewness, are non-zero for real fluids. Consequently, although the probability distribution may visually tend towards a Gaussian distribution for infinite systems, the resulting third moments obtained on integration are still finite—assuming that real fluid behavior represents the true thermodynamic limit.

D. Specific distributions for the common ensembles

The general result given in Equation (33) can be used to provide specific expressions for different ensembles. For the canonical ensemble we find

$$\ln P(E|N_1,V) = -\frac{S_0}{2N_1}\left[b_{11}(\beta\delta E)^2\right] + \frac{S_0^3}{6N_1^2}\left[c_{ttt}(\beta\delta E)^3\right]. \quad (41)$$

From Equation (21) or, equivalently, Equation (6), we also have $\delta S/R = \beta\delta E$ or $\delta A = 0$ for (infinitesimally) small fluctuations. Clearly, the Gaussian result corresponds to truncation after the first term on the right-hand side. The result can also be expressed in terms of common thermodynamic properties using the previous relationships.

For the isothermal-isobaric ensemble we have

$$\ln P(V,E|N_1) = -\frac{S_0}{2N_1}\left[b_{\varepsilon\varepsilon}(\rho_1\delta V)^2 + 2b_{1\varepsilon}\rho_1\delta V\beta\delta H + b_{11}(\beta\delta H)^2\right]\left(1-\frac{\rho_1\delta V}{N_1}\right) + \frac{S_0^3}{6N_1^2}\left[c_{sss}(\rho_1\delta V)^3 + 3c_{sst}(\rho_1\delta V)^2\beta\delta H + 3c_{stt}\rho_1\delta V(\beta\delta H)^2 + c_{ttt}(\beta\delta H)^3\right] \quad (42)$$

and $\delta S/R = \beta\delta H$ or $\delta G = 0$. One could expand the enthalpy to give the energy and volume fluctuations. However, this does not provide the simplest relationships for the corresponding moments.

In the GCE we have

$$\ln P(N_1,E|V) = -\frac{S_0}{2\langle N_1\rangle}\left[b_{\varepsilon\varepsilon}(\delta N_1)^2 - 2b_{1\varepsilon}\delta N_1\beta\delta\varepsilon + b_{11}(\beta\delta\varepsilon)^2\right] - \frac{S_0^3}{6\langle N_1\rangle^2}\left[c_{sss}(\delta N_1)^3 - 3c_{sst}(\delta N_1)^2\beta\delta\varepsilon + 3c_{stt}\delta N_1(\beta\delta\varepsilon)^2 - c_{ttt}(\beta\delta\varepsilon)^3\right] \quad (43)$$

TABLE I. Signs for the non-Gaussian measures in different fluid phases of water and argon. S.C. = supercritical; I.G. = classical ideal gas; $C_r = C_{V,m}/R$, where the heat capacity corresponds to that provided by the translational and intramolecular rotational and vibrational degrees of freedom, which are considered independent of temperature.

	Water			Argon			
Property	Gas	Liquid	S.C.	Gas	Liquid	S.C.	I.G.
S_0	+	+	+	+	+	+	$1/C_r$
b_{11}	+	+	+	+	+	+	1
$b_{1\varepsilon}$[a]	−	−	−	−	−	−	−1
$b_{\varepsilon\varepsilon}$	+	+	+	+	+	+	$1 + C_r$
J_1	+	+	+/−	+	+	+/−	2
J_2	+	+/−	+	+	+/−	+	1
J_3	−	+/−	+/−	−	−	−	0
J_4	+	−	+/−	+	−	+/−	1
c_{ttt}	+	+	+/−	+	+	+/−	$2C_r$
c_{stt}	−	+/−	−	−	+/−	+/−	$-C_r\,(2 + C_r)$
c_{sst}	+	+/−	+	+	+/−	+/−	$2C_r\,(1 + C_r)$
c_{sss}	−	+	+/−	−	+/−	+/−	$-C_r\,(1 + C_r)\,(2 + C_r)$

[a]The value of $b_{1\varepsilon}$ can be positive, negative, or zero (for liquid water) but is negative for the vast majority of the phase diagram. We have ignored changes in sign close to the vapor-liquid curve and the critical point.

and $\delta S^E/R = \beta\delta\varepsilon$ or $\delta A - \mu_1\delta N_1 = 0$. Hence, the excess energy is equivalent to the excess entropy for this ensemble. This condition can be rewritten to include the thermodynamic potential associated with the GCE to give $\delta G - \mu_1\delta N_1 - V\delta p = 0$. But, $\delta G = \mu_1\delta N_1$ for the GCE, and so we have $\delta p = 0$. Again, one can expand the excess energy term but, not only do the expressions become more complicated, they also lead to energy moments that are experimentally inaccessible.

Finally, the isobaric-isenthalpic ensemble leads to

$$\ln P(V|N_1,H) = -\frac{S_0}{2N_1}\left[b_{\varepsilon\varepsilon}(\rho_1\delta V)^2\right]\left(1-\frac{\rho_1\delta V}{N_1}\right) + \frac{S_0^3}{6N_1^2}\left[c_{sss}(\rho_1\delta V)^3\right] \tag{44}$$

and $\delta S = 0$. All the above equations can be expressed in terms of common thermodynamic properties using Equations (10) and (35)–(38).

E. Properties of the distributions

Previously we stated that, while the distribution tends towards a Gaussian for large systems, the higher moments (coskewness, excess cokurtosis) of the distribution are still finite. Here, we illustrate this behavior in more detail. Unfortunately, the entropy expansion through $S^{(3)}$ results in a distribution that cannot be integrated and hence the moments are unavailable. To provide a distribution that can be integrated, one requires higher terms in the expansion. As the general expression for the fluctuations leads to many complicated additional terms, we have focused on the simplest case—the energy distribution in the canonical ensemble—where we only require one extra derivative. This process is outlined in Appendix E. In particular, we compare the moments obtained from the current entropy expansion to the known (exact) moments obtained from the canonical ensemble partition function to ensure that the results are consistent.

TABLE II. Values for the non-Gaussian measures in liquid water at 298.15 K and 1 bar.

S_0	1.797	J_1	1.577	c_{ttt}	0.003
b_{11}	0.062	J_2	−0.800	c_{stt}	0.011
$b_{1\varepsilon}$	−0.077	J_3	−0.170	c_{sst}	−0.085
$b_{\varepsilon\varepsilon}$	9.060	J_4	−3.682	c_{sss}	10.484

Before using the expanded energy distribution, we note that the exact central moments (M_n) of the canonical ensemble energy distribution can be obtained by repeated differentiation of the partition function with respect to temperature.[2] This also provides the exact cumulants (K_n) of the distribution in terms of the central moments. The results are

$$\begin{aligned}
\langle\beta\delta E\rangle &= M_1 = 0,\\
\frac{C_V}{R} &= \left\langle(\beta\delta E)^2\right\rangle = M_2 = K_2 = \sigma^2,\\
2\frac{C_V}{R} + \frac{T}{R}\left(\frac{\partial C_V}{\partial T}\right)_{\rho_1} &= \left\langle(\beta\delta E)^3\right\rangle = M_3 = K_3,\\
6\frac{C_V}{R} + 6\frac{T}{R}\left(\frac{\partial C_V}{\partial T}\right)_{\rho_1} + \frac{T^2}{R}\left(\frac{\partial^2 C_V}{\partial T^2}\right)_{\rho_1} &= \left\langle(\beta\delta E)^4\right\rangle - 3\left\langle(\beta\delta E)^2\right\rangle^2 = M_4 - 3M_2^2 = K_4,
\end{aligned} \tag{45}$$

where σ^2 is the variance of the energy pdf and all the ensemble averages correspond to the canonical ensemble. The above expressions can be used to relate the fluctuations in the GCE and canonical ensembles as indicated in Equations (D4), (D6), and (E3).

All the pdfs presented in Secs. III A and III D actually depend on the system size. However, in most cases, one is only interested in the thermodynamic limit, i.e., $N_1 \to \infty$ with ρ_1 fixed. It is commonly noted that in the thermodynamic limit, the most probable and the average quantities coincide. However, while this is true in the Gaussian approximation, this is not true for the non-Gaussian counterparts. Hence, the equations presented in Sections III B and III D do not provide $\langle E\rangle = E^*$, etc. Clearly, for a finite system, we actually have $\delta E = E - E^*$ as the pdfs display a maximum at $\delta E = 0$. This does not invalidate the pdfs described previously, but it does have some subtle, and important, consequences that are discussed below.

Using the results presented in Equations (14), (17), (39), (E1), (E3), and (45) in the entropy expansion, we find that

$$\ln P(E|N_1,V) = -\frac{K_2}{2}\left(\frac{\beta\delta E}{K_2}\right)^2\left[1 - \frac{1}{3}\frac{K_3}{K_2}\left(\frac{\beta\delta E}{K_2}\right) - \frac{1}{12}\left(\frac{K_4}{K_2} - 3\frac{K_3^2}{K_2^2}\right)\left(\frac{\beta\delta E}{K_2}\right)^2\right]. \tag{46}$$

This is a general expression for the energy fluctuations in the canonical ensemble through fourth order. It is interesting to compare the central moments ($\tilde{M}_n$) and corresponding cumulants ($\tilde{K}_n$) of the distribution provided by Equation (46) with the known exact central moments and cumulants given in Equation (45). First, we note that the moments provided by Equation (46) depend on N_1, as does the visual appearance

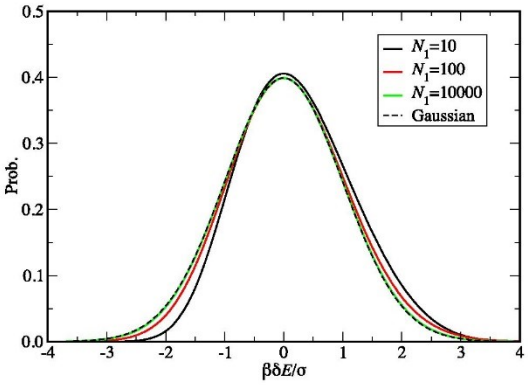

FIG. 3. Energy fluctuation probability distributions for the canonical ensemble as provided by Equation (46) in the high temperature limit. The normalized curves correspond to different systems sizes where $C_{V,m}/R = 1$, but the $C_V/R = N_1 = K_2 = \sigma^2$ values differ. The moments and cumulants obtained for infinite size systems correspond to exact values of $M_1/K_2 = 0$, $K_3/K_2 = 2$, and $K_4/K_2 = 6$.

of the distribution. This is illustrated in Figure 3 where we present the energy pdfs for different system sizes and compare the results with the corresponding Gaussian distribution. While the skewness and kurtosis are apparent for small system sizes, this appears to vanish for large systems and the distribution then visually resembles a Gaussian. This is due to the fact that the skewness and kurtosis become small in comparison with the variance. Numerical integration of the pdf for an infinite system provides

$$\frac{\tilde{M}_1}{\tilde{M}_2^{1/2}} = 0, \qquad \frac{\tilde{M}_3}{\tilde{M}_2^{3/2}} = 0, \qquad \frac{\tilde{M}_4}{\tilde{M}_2^{2}} = 3, \qquad \frac{\tilde{K}_4}{\tilde{M}_2^{2}} = 0 \tag{47}$$

for the standardized moments, as would be expected for a Gaussian distribution.

Upon integration, however, the asymmetry in the distribution is magnified and gives rise to non-zero values for the third and fourth cumulants. In fact, by numerical integration of Equation (46), we find that

$$\beta(\langle E\rangle - E^*) = \frac{1}{2}\frac{K_3}{K_2}, \qquad \frac{\tilde{K}_2}{K_2} = \frac{\tilde{K}_3}{K_3} = \frac{\tilde{K}_4}{K_4} = 1 \tag{48}$$

for $N_1 \to \infty$. The relationships are also good for $N_1 > \approx 1000$. Hence, the cumulants remain finite even though the distribution appears Gaussian. It is the cumulant ratios that give rise to the associated thermodynamic behavior, i.e., the changes in thermodynamic properties with changes in state point, as shown in Equation (45). For an infinite system, one obtains the same central moments and cumulants as provided by differentiation of the canonical ensemble partition function. The above results suggest that an analytical expression for the moments of the distribution should be available. However, we were unable to find one. The raw moments and corresponding cumulants ($\hat{K}_n$) are different from the central moments, as indicated by the first relationship in Equation (48), and are given by

$$\frac{\hat{K}_2}{K_2} = 1, \qquad \frac{\hat{K}_3}{K_2} = \frac{5}{2}\frac{K_3}{K_2}, \qquad \frac{\hat{K}_4}{K_2} = \frac{K_4}{K_2} + 2\left(\frac{K_3}{K_2}\right)^2 \tag{49}$$

for the distribution in Equation (46) in the infinite system limit.

As mentioned previously, if the $(\delta E)^4$ term in Equation (46) is set to zero, then the pdf cannot be (formally) integrated due to the behavior of the pdf as $\delta E \to \infty$. However, our numerical studies indicate that one actually finds a well-behaved Gaussian-like probability distribution followed by a divergence for energy fluctuations well removed from the main ($\delta E \approx 0$) distribution. The appearance of the divergence shifts to higher and higher energy fluctuations as N_1 increases (roughly as $N_1\ \sigma$). Consequently, for reasonably large values of N_1, the energy pdf is perfectly well behaved as long as one truncates the integration at 10 σ or so. In this case, the same results as described in Equation (48) are obtained.

The above relationships for the cumulants in terms of thermodynamic derivatives contain many terms. However, if we ignore the temperature dependence of the heat capacity, which is a valid approximation at high temperatures, then the fluctuations and entropy derivatives are simply given by the first term on the left hand side of Equation (45). The pdf can then be written as

$$\ln P(E|N_1, V) = \frac{C_V}{R}\left[-\frac{1}{2}\left(\frac{\beta\delta E}{C_V/R}\right)^2 + \frac{1}{3}\left(\frac{\beta\delta E}{C_V/R}\right)^3 - \frac{1}{4}\left(\frac{\beta\delta E}{C_V/R}\right)^4\right] \tag{50}$$

which takes the form of a $\ln(1+x)$ expansion. Assuming this is also true beyond fourth order, we find that for small ($|\beta\delta E| < C_V/R$) energy fluctuations, the un-normalized energy pdf is given by

$$P(E|N_1, V) = \left(1 + \frac{\beta\delta E}{C_V/R}\right)^{C_V/R} e^{-\beta\delta E} \tag{51}$$

through all orders, and also provides the entropy fluctuations via Equation (2). However, as the energy fluctuations can, in principle, be infinite, the more general distribution given by Equations (46) or (50) is required to determine the moments that characterize the actual truncated distribution. From Table II and Equation (39), it is clear that the high temperature limit is not valid for water under ambient conditions as $J_1 = 1.577 \ll 2$. However, this is primarily due to the intramolecular contributions to the heat capacity. For argon we find $J_1 \approx 2$ for large parts of the phase diagram, as shown in Figure 2. For a fluid at high temperature with unit variance ($K_2 = \tilde{K}_2 = 1$), the partition function provides cumulants of $K_3 = 2$ and $K_4 = 6$. The same results are obtained from Equation (46) for infinite size systems.

Ruppeiner has questioned the validity of a simple entropy expansion beyond second order.[4,11] This criticism is based on (at least) two aspects of the resulting pdf. First, with third order terms, the energy pdf is no longer symmetric and therefore $\langle E\rangle - E^*$ is no longer zero. This then violates the energy conservation and additivity laws between the system and the surrounding bath. Second, the resulting pdf is no longer covariant with respect to a change in thermodynamic coordinate

(property). From our numerical results using the pdf described in Equation (46), we find (as expected) that, even for infinite system sizes, the mean energy and most probable energy do not coincide. However, for large system sizes, the difference between the two is fixed by the third cumulant as indicated in Equation (48). Indeed, adding an (arbitrary) fifth and sixth order term to the entropy expansion in Equation (7) does not affect this result. Therefore, we find $\beta(E-E^*) = \beta(E-\langle E\rangle) + {}^1/_2\, K_3/K_2$ for large systems. Hence, the properties of the distribution, determined by $\beta\delta E/K_2$ in Equation (46), become insensitive to whether one uses $\delta E = E - E^*$ or $\delta E = E - \langle E\rangle$ as the difference just corresponds to a simple shift of the pdf along the energy coordinate. This is due to the fact that while the energy fluctuations are extensive, the K_3/K_2 ratio is intensive, and hence $(\langle E\rangle - E^*)/N_1$ then tends to zero, as indicated in Equation (48). Consequently, the pdf obeys the criteria required for energy conservation and additivity for all but very small ($N_1 < 1000$) systems, even though it possesses a finite skewness.

The issue of covariant behavior is more difficult to answer. A full discussion of the pdf presented here is hindered by the absence of a normalization factor. Normally, this is considered to be a simple constant. However, it is clear from the pdfs developed here that it must, at least, vary with system size and/or average energy. It is possible that any Jacobian factor arising from a change in coordinate may affect the results for very small system sizes but could lead to covariant behavior for large/infinite system sizes where the distribution resembles the properties of a simple Gaussian, as indicated by Equation (47). Alternatively, covariant behavior might be restricted to the pdf corresponding to an expansion of the entropy as a function of N_1, V, and E and not the conditional pdf generated by expanding in only one thermodynamic variable.

Consequently, the current approach seems to contradict the more general conclusion by Ruppeiner (and others) that simple entropy expansions are prone to difficulties. Unfortunately, this conflict has not been resolved here. Alternative approaches for simple systems have been suggested.[4,33] Nevertheless, the results presented here for the energy probability distribution indicate that the first four moments and cumulants obtained from the entropy expansion agree with those obtained from the partition function. Hence, assuming this also holds for other ensembles, it appears that the present non-Gaussian distribution can be useful, at least for infinite size systems.

IV. CONCLUSIONS

A simple general form for the probability distribution describing fluctuations in the particle number, volume, energy, and entropy has been presented. This single equation is valid for all the common ensembles of interest and has been extended to include the triplet fluctuations that describe non-Gaussian behavior. The probability distribution was developed using FST to manipulate the entropy derivatives into a convenient form and to then relate the results to experimentally available thermodynamic derivatives. It should be noted that FST is not required for this process and one could have used purely thermodynamic derivatives. However, the use of FST significantly simplifies the required manipulations. Visually, the resulting probability distribution tends towards the Gaussian result for large (infinite) system sizes. A set of four indicators for non-Gaussian behavior is proposed and investigated for fluid water and argon using equation of state data. It appears that the skewness of the distribution remains finite, even for infinite systems, and that Gaussian behavior is not possible for real fluids.

ACKNOWLEDGMENTS

The project described was supported by Grant No. R01GM118719 from the National Institute of General Medical Sciences to P.E.S. and Grant No. 1F32AG048690 from the National Institute on Aging to E.A.P. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Science, the National Institute on Aging, or the National Institutes of Health. The authors would like to thank John O'Connell and George Ruppeiner for interesting comments, and the reviewers for useful comments.

APPENDIX A: ENTROPY DERIVATIVES

A specific form of the entropy derivatives is used here. In particular, we define these derivatives to be dimensionless and intensive. The first derivatives of the entropy are obtained from Equation (6) and are given by

$$\begin{aligned} S_N &\equiv \frac{1}{R}\left(\frac{\partial S}{\partial N_1}\right)_{V,E} = -\beta\mu_1, \\ S_V &\equiv \frac{V_1}{R}\left(\frac{\partial S}{\partial V}\right)_{N_1,E} = Z, \\ S_E &\equiv \frac{1}{R\beta}\left(\frac{\partial S}{\partial E}\right)_{N_1,V} = 1. \end{aligned} \tag{A1}$$

A series of dimensionless intensive second derivatives are defined

$$\begin{aligned} S_{NN} &\equiv \frac{N_1}{R}\frac{\partial^2 S}{\partial N_1^2} = -N_1\left(\frac{\partial \beta\mu_1}{\partial N_1}\right)_{V,E}, \\ S_{VV} &\equiv \frac{N_1 V_1^2}{R}\frac{\partial^2 S}{\partial V^2} = N_1 V_1^2\left(\frac{\partial \beta p}{\partial V}\right)_{N_1,E}, \\ S_{EE} &\equiv \frac{N_1}{R\beta^2}\frac{\partial^2 S}{\partial E^2} = \frac{N_1}{\beta^2}\left(\frac{\partial \beta}{\partial E}\right)_{N_1,V} \end{aligned} \tag{A2}$$

together with a series of cross derivatives

$$\begin{aligned} S_{NV} &\equiv \frac{N_1 V_1}{R}\frac{\partial^2 S}{\partial N_1 \partial V} = -N_1 V_1\left(\frac{\partial \beta\mu_1}{\partial V}\right)_{N_1,E} \\ &= N_1 V_1\left(\frac{\partial \beta p}{\partial N_1}\right)_{V,E}, \\ S_{NE} &\equiv \frac{N_1}{R\beta}\frac{\partial^2 S}{\partial N_1 \partial E} = -\frac{N_1}{\beta}\left(\frac{\partial \beta\mu_1}{\partial E}\right)_{N_1,V} = \frac{N_1}{\beta}\left(\frac{\partial \beta}{\partial N_1}\right)_{V,E}, \\ S_{VE} &\equiv \frac{N_1 V_1}{R\beta}\frac{\partial^2 S}{\partial V \partial E} = \frac{N_1 V_1}{\beta}\left(\frac{\partial \beta p}{\partial E}\right)_{N_1,V} = \frac{N_1 V_1}{\beta}\left(\frac{\partial \beta}{\partial V}\right)_{N_1,E}, \end{aligned} \tag{A3}$$

where the equalities on the far right hand side are a result of Equation (A1). A similar set of dimensionless third derivatives

can also be defined where one includes a factor of $N_1{}^2$ and additional factors of V_1 for each volume derivative and β for each energy derivative. For example,

$$S_{NVE} \equiv \frac{N_1^2 V_1}{\beta R} \frac{\partial^3 S}{\partial N_1 \partial V \partial E} \tag{A4}$$

and similarly for the other nine unique third derivatives.

APPENDIX B: SECOND DERIVATIVES OF THE ENTROPY USING FST

The dimensionless second derivatives can be expressed using just the b's provided by FST. Here we provide the relationships used to generate the results presented in Equation (14). Our approach is to develop relationships between various thermodynamic derivatives and then use the known results from FST. First, we note that

$$\begin{aligned}\left(\frac{\partial \beta\mu_1}{\partial E}\right)_{N_1,V} &= H_1 \left(\frac{\partial \beta}{\partial E}\right)_{N_1,V} + \beta V_1 \left(\frac{\partial p}{\partial E}\right)_{N_1,V},\\ \left(\frac{\partial \beta\mu_1}{\partial V}\right)_{N_1,E} &= H_1 \left(\frac{\partial \beta}{\partial V}\right)_{N_1,E} + \beta V_1 \left(\frac{\partial p}{\partial V}\right)_{N_1,E}\end{aligned} \tag{B1}$$

are generated from the differential of $\beta\mu_1(\beta,p,N_1)$. A series of Euler chain rule relationships are also available

$$\begin{aligned}\left(\frac{\partial \beta\mu_1}{\partial N_1}\right)_{V,E} \left(\frac{\partial N_1}{\partial E}\right)_{\beta\mu_1,V} \left(\frac{\partial E}{\partial \beta\mu_1}\right)_{N_1,V} &= -1,\\ \left(\frac{\partial p}{\partial V}\right)_{N_1,E} \left(\frac{\partial V}{\partial E}\right)_{N_1,p} \left(\frac{\partial E}{\partial p}\right)_{N_1,V} &= -1,\\ \left(\frac{\partial \beta}{\partial V}\right)_{N_1,E} \left(\frac{\partial V}{\partial E}\right)_{N_1,\beta} \left(\frac{\partial E}{\partial \beta}\right)_{N_1,V} &= -1\end{aligned} \tag{B2}$$

together with the GCE differential,[16]

$$V_1 d\beta p = -E_1 d\beta + d\beta\mu_1. \tag{B3}$$

The central derivative in the first two relationships of Equation (B2) can be developed using the chain rule (in β) to give[27]

$$\left(\frac{\partial \beta\mu_1}{\partial N_1}\right)_{V,E} = -\left(\frac{\partial \beta\mu_1}{\partial E}\right)_{N_1,V} \frac{\langle \delta E \delta E\rangle}{\langle \delta N_1 \delta E\rangle} \tag{B4}$$

and[24]

$$\beta N_1 V_1^2 \left(\frac{\partial p}{\partial V}\right)_{N_1,E} = -S_0(b_{\varepsilon\varepsilon} + b_{1\varepsilon} Z). \tag{B5}$$

Finally, we require the known FST based expressions for two of the derivatives appearing above[24]

$$\begin{aligned}\frac{V_1}{RT}\left(\frac{\partial E_1}{\partial V_1}\right)_T &= -\frac{b_{1\varepsilon} + b_{11} Z}{b_{11}},\\ V_1 \left(\frac{\partial p}{\partial E_1}\right)_{\rho_1} &= -\frac{b_{1\varepsilon}}{b_{\varepsilon\varepsilon} b_{11} - b_{1\varepsilon}^2}.\end{aligned} \tag{B6}$$

Using these relationships, all the derivatives in Equation (14) can be expressed in terms of the fluctuating quantities for the equivalent GCE.

More specifically, S_{VE} was obtained from the last expression in Equation (B2) and the use of the first expression in Equation (B6); S_{NE} was obtained from the first expression in Equation (B1) and the use of the last expression in Equation (B6); S_{NN} was obtained from Equation (B4) and the expression for S_{NE}; S_{NV} was obtained from the last expression in Equation (B1) using the result of S_{VE} and also Equation (B5); S_{VV} was obtained from Equation (B3) using the results of S_{VE} and S_{NV}; and finally S_{EE} is provided directly by the last expression in Equation (D4).

APPENDIX C: THIRD DERIVATIVES OF THE ENTROPY USING FST

To obtain the third derivatives of the entropy as provided in Equations (15)–(18), we start from the second derivatives provided in Equation (14) and note that the differential of a number or energy density (X/V) can be written as[19]

$$d\left[\frac{\langle X\rangle}{V}\right] = \frac{\langle \delta X \delta N_1\rangle}{V} d\beta\mu_1 - \frac{\langle \delta X \delta E\rangle}{V} d\beta \tag{C1}$$

in the GCE, i.e., using $\beta\mu_1$ and β as independent variables. Extending this to include the pair fluctuations in the GCE, one finds that the required derivatives of the b's are provided by

$$\left(\frac{\partial B_{XY}}{\partial z}\right)_{z'} = \frac{\langle \delta X \delta Y \delta N_1\rangle}{V}\left(\frac{\partial \beta\mu_1}{\partial z}\right)_{z'} - \frac{\langle \delta X \delta Y \delta E\rangle}{V}\left(\frac{\partial \beta}{\partial z}\right)_{z'}, \tag{C2}$$

where z' indicates that all thermodynamic variables (in the NVE ensemble) except for z (= N_1, V or E) are held constant, and X and Y can be N_1 or E. This general result can be written in terms of our reduced fluctuations such that

$$V_1 \left(\frac{\partial B_{XY}}{\partial z}\right)_{z'} = c_{XY1}\left(\frac{\partial \beta\mu_1}{\partial z}\right)_{z'} - c_{XYE}\left(\frac{\partial \ln \beta}{\partial z}\right)_{z'}. \tag{C3}$$

The derivatives on the right hand side of Equation (C3) are provided in Equations (A2), (A3), and (14), and then lead to

$$\begin{aligned}N_1 V_1 \left(\frac{\partial B_{XY}}{\partial N_1}\right)_{V,E} &= -S_0 c_{XYn},\\ N_1 V_1^2 \left(\frac{\partial B_{XY}}{\partial V}\right)_{N_1,E} &= S_0 c_{XYv},\\ \frac{N_1 V_1}{\beta}\left(\frac{\partial B_{XY}}{\partial E}\right)_{N_1,V} &= S_0 c_{XYe}\end{aligned} \tag{C4}$$

after noting that

$$N_1\left(\frac{\partial E_1}{\partial N_1}\right)_{V,E} = -E_1, \quad \left(\frac{\partial E_1}{\partial V}\right)_{N_1,E} = 0, \quad N_1\left(\frac{\partial E_1}{\partial E}\right)_{N_1,V} = 1. \tag{C5}$$

Use of Equation (C4) provides, after some additional manipulation, the third derivatives of the entropy given in Equations (15)–(18).

APPENDIX D: NON-GAUSSIAN DISTRIBUTION FUNCTION TERMS USING FST

Here, we develop the third order term in the entropy expansion using FST to help manipulate the derivatives. If we focus on the triplet fluctuations, then only four terms survive on transforming the fluctuating variables. The contribution to $S^{(3)}$ from

the pair fluctuations is less clean as only one of the entropy derivative combinations cancels. However, if we exclude the μpT ensemble from consideration, i.e., only closed isobaric ensembles are considered, then the pair contributions term simplifies considerably and we find that

$$S^{(3)} = F_{VVV}(\delta V^E)^3 + 3F_{VVH}(\delta V^E)^2\delta H^E + 3F_{VHH}\delta V^E(\delta H^E)^2 + F_{HHH}(\delta H^E)^3 \tag{D1}$$

where,

$$\begin{aligned} F_{VVV} &\equiv S_{VVV} - 3S_{VVE}Z + 3S_{EEV}Z^2 - S_{EEE}Z^3 \\ &= S_0^3 c_{sss} + 3S_0 b_{\varepsilon\varepsilon}, \\ F_{VVH} &\equiv S_{VVE} - 2S_{EEV}Z + S_{EEE}Z^2 = S_0^3 c_{sst} + 2S_0 b_{1\varepsilon}, \\ F_{VHH} &\equiv S_{EEV} - S_{EEE}Z = S_0^3 c_{stt} + S_0 b_{11}, \\ F_{HHH} &\equiv S_{EEE} = S_0^3 c_{ttt} \end{aligned} \tag{D2}$$

and we have used the following relationships:

$$\begin{aligned} s &= v - eZ, \\ t &= e, \\ n - v - eE_1 &= 0. \end{aligned} \tag{D3}$$

All of the other (δN_1) terms are found to be zero. This leads to the relationships provided in Equations (32) and (33).

To express the results for the c's found in Equation (D2) in terms of thermodynamic derivatives that might be provided by an equation of state, we require known results from FST. The dimensionless thermal pressure coefficient is given by[24]

$$Z\frac{T}{p}\left(\frac{\partial p}{\partial T}\right)_{\rho_1} = Z\frac{T\alpha_p}{p\kappa_T} = -\frac{b_{1\varepsilon}}{b_{11}} \tag{D4}$$

together with the density derivatives[24]

$$\begin{aligned} \frac{p^2}{\rho_1 Z^2}\frac{\partial^2\rho_1}{\partial p^2} &= c_{111} - b_{11}^2, \\ \frac{Tp}{\rho_1 Z}\frac{\partial^2\rho_1}{\partial T\partial p} &= c_{11\varepsilon} - b_{11}(1 + b_{1\varepsilon}), \end{aligned} \tag{D5}$$

where the pressure derivatives are taken with T constant and the temperature derivatives with p constant. Finally, we also require the following isochoric derivatives:[24]

$$\begin{aligned} \frac{pT}{Z}\left(\frac{\partial\kappa_T}{\partial T}\right)_{\rho_1} &= c_{11\varepsilon} - c_{111}\frac{b_{1\varepsilon}}{b_{11}} - b_{11}, \\ Z\frac{T^2}{p}\left(\frac{\partial^2 p}{\partial T^2}\right)_{\rho_1} &= \frac{b_{\varepsilon\varepsilon}b_{11} - b_{1\varepsilon}^2}{b_{11}} - \frac{1}{b_{11}}\left[c_{1\varepsilon\varepsilon} - 2c_{11\varepsilon}\frac{b_{1\varepsilon}}{b_{11}} + c_{111}\left(\frac{b_{1\varepsilon}}{b_{11}}\right)^2\right], \\ \frac{T}{R}\left(\frac{\partial C_{V,m}}{\partial T}\right)_{\rho_1} &= -2\frac{b_{\varepsilon\varepsilon}b_{11} - b_{1\varepsilon}^2}{b_{11}} + \left[c_{\varepsilon\varepsilon\varepsilon} - 3c_{1\varepsilon\varepsilon}\left(\frac{b_{1\varepsilon}}{b_{11}}\right) + 3c_{11\varepsilon}\left(\frac{b_{1\varepsilon}}{b_{11}}\right)^2 - c_{111}\left(\frac{b_{1\varepsilon}}{b_{11}}\right)^3\right] \end{aligned} \tag{D6}$$

for pure fluids.

The isochoric derivatives presented in Equations (11) and (D6) can be written in a more condensed form using our definition of t to give

$$\begin{aligned} \frac{b_{tt}}{b_{11}^2} &= \frac{C_{V,m}}{R}, \\ \frac{c_{11t}}{b_{11}^2} &= 1 + \frac{T}{\kappa_T}\left(\frac{\partial\kappa_T}{\partial T}\right)_{\rho_1}, \\ \frac{c_{1tt}}{b_{11}^3} &= \frac{C_{V,m}}{R} - Z\frac{T^2}{p}\left(\frac{\partial^2 p}{\partial T^2}\right)_{\rho_1}, \\ \frac{c_{ttt}}{b_{11}^3} &= 2\frac{C_{V,m}}{R} + \frac{T}{R}\left(\frac{\partial C_{V,m}}{\partial T}\right)_{\rho_1}. \end{aligned} \tag{D7}$$

Then, using the relationships provided in Equations (D5), (13), and (D7), together with the fact that

$$s = t\frac{b_{1\varepsilon}}{b_{11}} - N_1\frac{C_{V,m}}{R} \tag{D8}$$

in Equation (33), provides the results given in Equations (35)–(39).

APPENDIX E: THE ENERGY PROBABILITY DISTRIBUTION FUNCTION THROUGH FOURTH ORDER

The intensive dimensionless fourth energy derivative of the entropy can be obtained from the third derivative provided in Equation (17) using the approach outlined in Appendix C. The result is

$$S_{EEEE} \equiv \frac{N_1^3}{R\beta^4}\left(\frac{\partial^4 S}{\partial E^4}\right)_{N_1,V} = S_0^4\left[d_{tttt} - 3\frac{c_{1tt}^2}{b_{11}} - 3\frac{c_{ttt}^2}{b_{tt}^2}\right], \tag{E1}$$

where

$$d_{tttt} = \frac{\left\langle(\delta t)^4\right\rangle - 3\left\langle(\delta t)^2\right\rangle^2}{\langle N_1\rangle}. \tag{E2}$$

To relate this new fluctuating quantity to more common thermodynamic derivatives, we take a temperature derivative of the last expression in Equation (D6) to give

$$6\frac{C_{V,m}}{R} + 6\frac{T}{R}\left(\frac{\partial C_{V,m}}{\partial T}\right)_{\rho_1} + \frac{T^2}{R}\left(\frac{\partial^2 C_{V,m}}{\partial T^2}\right)_{\rho_1} = d_{ttttt} - 3\frac{c_{1tt}^2}{b_{11}}. \tag{E3}$$

This, together with the relationships in Equation (D7), is sufficient to express the fourth derivative of the entropy in terms of more convenient thermodynamic derivatives. Alternatively, the energy derivatives of the entropy can also be evaluated by taking multiple derivatives of Equation (6), using the chain rule for $\partial C_V/\partial E = \partial C_V/\partial T \cdot \partial T/\partial E$, and noting that $\partial E/\partial T = C_V$ at fixed density. Hence, FST is not required in this case and the approach becomes totally thermodynamic in nature. Both approaches give the same results.

[1]T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956).
[2]D. A. McQuarrie, *Statistical Mechanics* (Harper & Row, New York, 1976).
[3]M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Oxford University Press, USA, 1987).
[4]G. Ruppeiner, Rev. Mod. Phys. **67**, 605 (1995).
[5]H. B. Callen and T. A. Welton, Phys. Rev. **83**, 34 (1951).
[6]G. Nicolis and I. Prigogine, Proc. Natl. Acad. Sci. U. S. A. **68**, 2102 (1971).

[7]P. Colonna, N. R. Nannan, A. Guardone, and T. P. van der Stelt, Fluid Phase Equilib. **286**, 43 (2009).
[8]M. Castier and V. F. Cabral, Fluid Phase Equilib. **334**, 128 (2012).
[9]L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part I*, 3rd ed., translated from the Russian by J. B. Sykes and M. J. Kearsley (Pergamon Press, Oxford, 1980).
[10]Y. Mishin, Ann. Phys. **363**, 48 (2015).
[11]G. Ruppeiner, Am. J. Phys. **78**, 1170 (2010).
[12]L. Diosi and B. Lukacs, Phys. Rev. A **31**, 3415 (1985).
[13]J. A. Tuszynski, M. J. Clouter, and H. Kiefte, Phys. Rev. B **33**, 3423 (1986).
[14]E. A. Ploetz, S. Karunaweera, and P. E. Smith, J. Chem. Phys. **142**, 044502 (2015).
[15]R. F. Greene and H. B. Callen, Phys. Rev. **83**, 1231 (1951).
[16]N. Davidson, *Statistical Mechanics* (McGraw-Hill, New York, 1962).
[17]J. G. Kirkwood and R. J. Goldberg, J. Chem. Phys. **18**, 54 (1950).
[18]E. A. Ploetz and P. E. Smith, J. Chem. Phys. **135**, 044506 (2011).
[19]E. A. Ploetz and P. E. Smith, Adv. Chem. Phys. **153**, 311 (2013).
[20]E. A. Ploetz and P. E. Smith, J. Phys. Chem. B **119**, 7761 (2015).
[21]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).
[22]A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press, New York, 2006).
[23]P. E. Smith, E. Matteoli, and J. P. O'Connell, *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics* (CRC Press, Boca Raton, 2013).
[24]E. A. Ploetz, G. N. Pallewela, and P. E. Smith, "Fluctuation Solution Theory of Pure Fluids," J. Chem. Phys. (to be published).
[25]A. Munster, *Classical Thermodynamics* (Wiley-Interscience, London, 1970).
[26]P. Schofield, Proc. Phys. Soc. **88**, 149 (1966).
[27]C. G. Gray and K. E. Gubbins, *Theory of Molecular Fluids: Fundamentals* (Oxford University Press, New York, 1984), Vol. 1
[28]H. C. Andersen, J. Chem. Phys. **72**, 2384 (1980).
[29]W. Wagner and A. Pruss, J. Phys. Chem. Ref. Data **31**, 387 (2002).
[30]C. Tegeler, R. Span, and W. Wagner, J. Phys. Chem. Ref. Data **28**, 779 (1999).
[31]A. H. Harvey, A. P. Peskin, and S. A. Klein, NIST Standard Reference Database 10: NIST/ASME Steam Properties, Version 2.22 (U.S. Department of Commerce, Gaithersburg, 2008).
[32]E. W. Lemmon, M. L. Huber, and M. O. McLinden, REFPROP: Reference Fluid Thermodynamic and Transport Properties, Version 9.1, DLL version number 9.1, U.S. Secretary of Commerce on behalf of the United States of America: Applied Chemicals and Materials Division, 2013.
[33]G. Ruppeiner, Phys. Rev. A **27**, 1116 (1983).