THEORY AND SIMULATION OF LIQUIDS AND LIQUID MIXTURES

by

GAYANI NADEERA PALLEWELA

B.S., University of Colombo, 2010

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

2016

Abstract

Kirkwood Buff (KB) theory is one of the most important theories of solutions. The theory can relate integrals over radial (pair) distribution functions (rdfs) in the grand canonical ensemble to common thermodynamic properties. An inversion of the KB theory has been proposed by Ben-Naim and this has led to the wide spread popularity of KB theory. The idea of the KB inversion procedure is to calculate KB integrals from available thermodynamic properties.

The KB theory can be used to validate the force field (ff) parameters used in molecular dynamics simulations. We have tested a series of small molecule ff parameters using KB theory that consists of both atom centered partial atomic charges and extra charge sites. The results indicate that using extra charge sites, derived from QM calculations, does not necessarily provide a more accurate representation of condensed phase properties. A further study aimed at an ongoing project of deriving new biomolecular ff parameters based on KB theory, has developed ff parameters for esters in order to represent the ester conjugation of the phospholipid molecule. The models were further tested against experimental properties.

Preferential solvation (PS) is an important concept of solution mixtures that can be described using KB theory. The difference between local composition and bulk composition in solution mixtures leads to the concept of PS. A generalized explanation based on local mole fractions was derived by Ben-Naim using KB theory. However, the original expressions have been modified over years. Here, we propose a new approach based on local volume fractions to explore PS in binary and ternary solution mixtures. Experimental and simulation data were used to examine different approaches to PS.

A relationship between the rdf and the triplet distribution function can be obtained using the Kirkwood Superposition Approximation (KSA). A combination of Fluctuation Solution

Theory and experimental rdfs are used to examine the KSA at a series of state points for pure water. The accuracy of several other approximate relationships between the pair and triplet correlation functions was also investigated and are in good agreement for regions of the phase diagram where the compressibility is small.

THEORY AND SIMULATION OF LIQUIDS AND LIQUID MIXTURES

by

GAYANI NADEERA PALLEWELA

B.S., University of Colombo, 2010

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

2016

Approved by:

Major Professor
Professor Paul E. Smith

Copyright

Abstract

Kirkwood Buff (KB) theory is one of the most important theories of solutions. The theory can relate integrals over radial (pair) distribution functions (rdfs) in the grand canonical ensemble to common thermodynamic properties. An inversion of the KB theory has been proposed by Ben-Naim and this has led to the wide spread popularity of KB theory. The idea of the KB inversion procedure is to calculate KB integrals from available thermodynamic properties.

The KB theory can be used to validate the force field (ff) parameters used in molecular dynamics simulations. We have tested a series of small molecule ff parameters using KB theory that consists of both atom centered partial atomic charges and extra charge sites. The results indicate that using extra charge sites, derived from QM calculations, does not necessarily provide a more accurate representation of condensed phase properties. A further study aimed at an ongoing project of deriving new biomolecular ff parameters based on KB theory, has developed ff parameters for esters in order to represent the ester conjugation of the phospholipid molecule. The models were further tested against experimental properties.

Preferential solvation (PS) is an important concept of solution mixtures that can be described using KB theory. The difference between local composition and bulk composition in solution mixtures leads to the concept of PS. A generalized explanation based on local mole fractions was derived by Ben-Naim using KB theory. However, the original expressions have been modified over years. Here, we propose a new approach based on local volume fractions to explore PS in binary and ternary solution mixtures. Experimental and simulation data were used to examine different approaches to PS.

A relationship between the rdf and the triplet distribution function can be obtained using the Kirkwood Superposition Approximation (KSA). A combination of Fluctuation Solution

Theory and experimental rdfs are used to examine the KSA at a series of state points for pure water. The accuracy of several other approximate relationships between the pair and triplet correlation functions was also investigated and are in good agreement for regions of the phase diagram where the compressibility is small.

Table of Contents

List of Figures

List of Tables

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor, Professor Paul E. Smith for his guidance, support, motivation and patience throughout my graduate study. I learned so much from your immense knowledge and unique perspective towards the research and achieving goals. I feel myself blessed and truly believe that it is been an amazing period of my life that I have spent under your supervision. I hope to pursue my studies with all the experience and knowledge that have obtained from you to become a great scientist one day.

I would also like to thank my Ph.D. advisory committee, Professor Christine Aikens, Professor Eric A. Maatta, Professor Om Prakash, and Professor Peter H. Pfromm for their valuable time, support and input.

I like to convey a very special thanks to my undergraduate advisor Professor Samantha Weerasinghe for his support, motivation and advices throughout this time. Thank you for believing in me. A special gratitude is extended to all my teachers for sharing their knowledge and for the enormous support to fulfill my dreams.

Big thanks to all the past and present Smith group members; Dr. Yuanfang Jiao, Dr. Elizabeth Ploetz, Dr. Shu Dai, Sadish Karunaweera, Mohomed Nawavi Mohomed Naleem, Nilusha Kariyawasam Manachchige. Thank you for the support and wonderful time that we shared together. My sincere gratitude also go to all my friends and relatives.

At last, but not least I would like to acknowledge my father, mother, brother and Vishara for their unconditional love, support and encouragement. I finally feel pleased as I have done something to make my father proud. Thank you!!! This would not be possible without you.

Dedication

To my dearest father and mother

For their endless support, love and encouragement

To my brother

For being a best friend and a very supportive brother

To *Vishara*

For believing in me and completing my life with your love

Chapter 1 - Introduction

1.1 Molecular Dynamics Simulations

Molecular simulations have been used for over five decades to study various types of biological systems in order to provide extreme detailed atomic level understanding that cannot be easily obtained from the other methods.[1,2-5] There are two main types of computer simulation techniques available as of today: molecular dynamics (MD) and Monte Carlo (MC) simulations.[6-9] These two main computer simulation techniques contain their own distinct features. Which one is more suitable for a particular study depends on the system and the properties that are going to be studied. MC simulations attempt to predict the time independent properties whereas MD simulations also predict the time dependent behavior of systems. Besides these two techniques, people have also been trying to develop hybrid techniques using the features from the above two techniques.[2, 7] This dissertation is solely based on molecular dynamics simulations.

The most important use of computer simulations is to relate macroscopic properties to microscopic properties. It is obvious that experiments play the central role in science. However, by depending only on experiments, it is not possible to get clear explanations for all the observations of complex systems. Therefore, computer simulations have become a powerful tool to obtain a better understanding of unexplained phenomenon. At the same time, people use computer simulations to test and validate theories. Therefore, simulations act as a bridge between experiments and theories. Furthermore, it is possible to conduct computer simulations, under difficult, hazardous or expensive conditions, that cannot be easily accessible under laboratory conditions. For instance, under extremes of pressure and temperature, or systems involving biologically hazardous chemicals, etc.[2]

MD simulations have been used to study a wide variety of important phenomenon in physics, chemistry and biology. Over the years, due to the revolutionary advances in computers and algorithms, MD simulations have become one of the most important and reliable techniques to study complex biological systems.[10-13] The molecules and their behavior is best described by quantum mechanical models. However, it is not possible to model complex biological systems that contain a large number of molecules using quantum mechanics due to inaccessible computer demand. In most MD simulations, therefore, classical equations of motion are solved for a finite time period.[7] Atoms and molecules are represented using simple balls and springs. The main objective is to find the forces acting on each molecule in the system. Thus, we can determine the change in the properties of the system with the time. The derivative of the potential energy (U) is used to calculate the forces acting on each molecule as shown in Equation (1.1) and Equation (1.2).

$$m_i\ddot{\vec{r}}_i = \vec{f}_i \tag{1.1}$$

$$\vec{f}_i = -\frac{\partial U}{\partial \vec{r}_i} \tag{1.2}$$

where $\vec{f}_i$ is the force exerted on ith particle that has mass of m_i, $\vec{r}_i$ is the position of the particle, and the second derivative of $\vec{r}_i$ with respect to time provides the acceleration ($\ddot{\vec{r}}_i$). High accuracy of the potential energy calculation is very important in MD simulations as it determines the next state of the system. Numerical integration methods are used to solve the equations of motions.[7, 14] Finite difference methods are one of the widely applied methods to solve the above differential equations. The general idea is to predict the new positions and velocities of the molecules using current and previous details of the system and a finite step in time. The Verlet algorithm is one of the most popular algorithms that is used in MD to solve the equations of motion.[14-15] Equation (1.3) and

Equation (1.4) can be obtained using a Taylor expansion about $\vec{r}(t)$ to predict the future position $\vec{r}(t+\delta t)$ from the previous position $\vec{r}(t-\delta t)$ of the atoms in the system,

$$\vec{r}(t+\delta t) = \vec{r}(t) + \delta t\vec{\dot{r}}(t) + 1/2\,\delta t^2\vec{\ddot{r}}(t) + \cdots \tag{1.3}$$

$$\vec{r}(t-\delta t) = \vec{r}(t) - \delta t\vec{\dot{r}}(t) + 1/2\,\delta t^2\vec{\ddot{r}}(t) - \cdots \tag{1.4}$$

where δt, $\vec{r}(t)$, $\vec{\dot{r}}(t)$, $\vec{\ddot{r}}(t)$ represent integration time step, current position, velocity and acceleration, respectively. The velocity can be removed from the addition of Equation (1.3) and Equation (1.4) to give Equation (1.5),

$$\vec{r}(t+\delta t) = 2\vec{r}(t) - \vec{r}(t-\delta t) + \delta t^2\vec{\ddot{r}}(t) \tag{1.5}$$

Even though the determination of a new trajectory does not need the velocity, the determination of the kinetic energy needs the velocity. Velocities can be obtained from Equation (1.6) if necessary.

$$\vec{\dot{r}}(t) = \frac{\vec{r}(t+\delta t) - \vec{r}(t-\delta t)}{2\delta t} \tag{1.6}$$

The characteristics of the simulation system determine the magnitude of the time step. Figure 1.1 shows the simplified version of the main steps in a molecular dynamics simulation.

Not only equilibrium systems, but also non equilibrium systems can be studied using MD simulations. There is no net transport of mass, momentum and heat in equilibrium systems. But, if the system is far away from the equilibrium, the link between microscopic dynamical properties and non equilibrium macroscopic states is not easy to establish.[16] The perturbations from the equilibrium states could be seen. These type of systems can be studied using non equilibrium molecular dynamics (NEMD) simulations. Therefore, it is possible to obtain microscopic

dynamics and macroscopic non equilibrium properties. In non equilibrium molecular dynamics simulations external forces are applied to the system.[14, 16] This dissertation is solely based on equilibrium MD simulations.

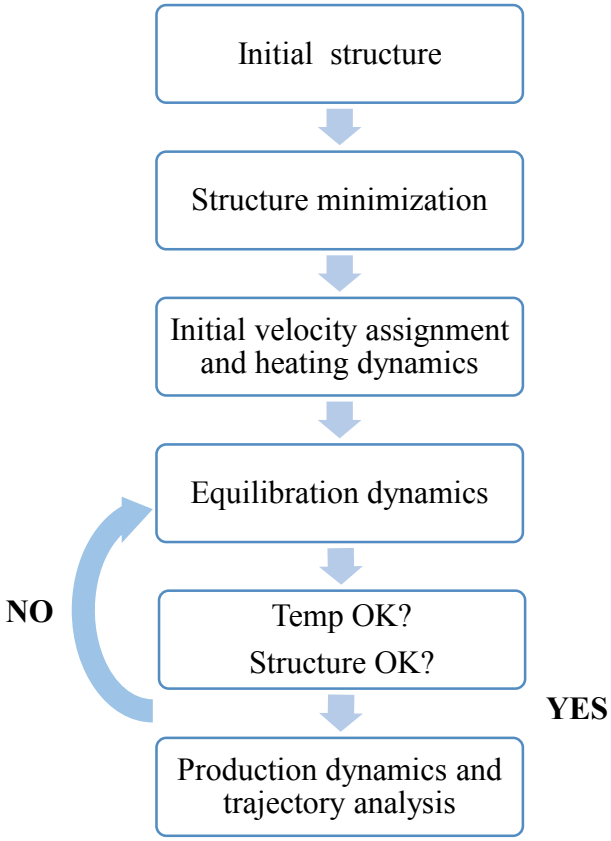

Figure 1.1 Simplified schematic of the main steps in a molecular dynamics simulation[17]

1.2 Force Fields

A force field is defined by a set of equations and parameters that are used to calculate the potential energy of the system.[6, 18-22] The accuracy of the simulation results mainly depend on the accuracy of the force field.[23] Most force fields use atomistic models instead of considering electrons and nuclei to represent the molecules. Thus, we may consider atoms as the building blocks of the large biological macromolecules. Even though it has been over four decades since

the first molecular dynamics simulation was conducted,[24] people consistently try to improve force field parameters.[25-28]

In classical MD simulations we may separate the potential energy function in to two groups: bonded interactions and non bonded interactions. Bond, angle, proper and improper dihedral rotations are categorized under the bonded interactions. On the other hand, van der Waals interactions and Coulomb interactions are categorized under non bonded interactions. The following set of equations typically represent the potential energy of the system.

$$V_{bond} = \sum \frac{1}{2} k_b (r - r_0)^2 \tag{1.7}$$

$$V_{angle} = \sum \frac{1}{2} k_a (\theta - \theta_0)^2 \tag{1.8}$$

$$V_{proper\ dihedral} = \sum k_\phi [1 + cos(n\phi - \phi_s)] \tag{1.9}$$

$$V_{improper\ dihedral} = \sum \frac{1}{2} k_\xi \left(\xi_{ijkl} - \xi_0\right)^2 \tag{1.10}$$

$$V_{electrostatics} = \sum \frac{q_i q_j}{r_{ij}} \tag{1.11}$$

$$V_{van\ der\ Waals} = \sum_{LJ} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6} \right) \tag{1.12}$$

A Cartesian coordinates set R is used to define the position of all the atoms. Therefore, one can obtain internal coordinates for bond lengths (r), bond angles (θ), proper dihedral angles (ϕ), and interatomic distances (r_{ij}). A hypothetical molecule that illustrates the potential energy

functions is shown in Figure 1.2. Even though the potential energy function remains similar, the parametrization process is different in most of the currently available force fields. Most importantly, the above simple terms that are used to obtain interaction energies of complex biomolecular systems represent a compromise between simplicity and accuracy.

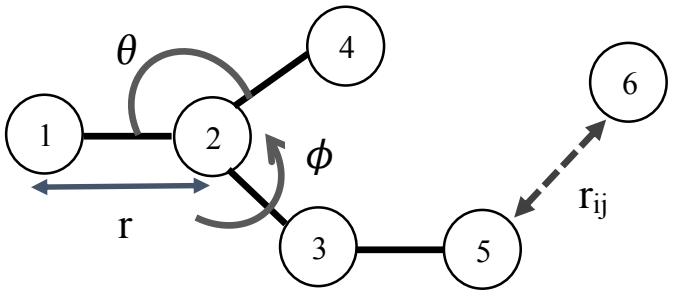

Figure 1.2 Illustration of several potential energy terms using a simple model[23]

As denoted in Equation (1.7), Equation (1.8) and Equation (1.10), harmonic energy functions have been used to model bond stretching, angle bending and improper dihedral interactions. This helps to keep the bonds and angles near their equilibrium values during the simulation. Equilibrium bond length, angle and improper dihedral angles are represented by r_0, θ_0, ξ_0, respectively. Furthermore, associated force constants for bond stretching, angle bending and improper dihedral interactions are represented by k_b, k_a, k_ξ, respectively. Proper dihedral interactions can be calculated using Equation (1.9). This energy term represents the energy barriers associated with rotation around a bond which results in a change in the relative positions of atom 1 and atom 5 in space. The force constant, periodicity and phase are represented by k_ϕ, n, ϕ_s. Non bonded interactions are calculated between different molecules and also within the same molecule. When considering the same molecule, the two atoms should be separated by at least three bonds. Most importantly, when these two atoms are separated by exactly three bonds, interactions are often adjusted using a scaling factor. Different force fields use different scaling factors.[18, 20] The

Coulomb interactions (electrostatic interactions) can be determined using partial atomic charges (q_i , q_j). Generally, the van der Waals interactions can be determined using a Lennard-Jones (12-6) potential as shown in Equation (1.12). The parameters ε_{ij}, σ_{ij} represent the well depth which is also known as minimum interaction energy and the radius where the LJ potential energy is zero, respectively. The repulsion interaction between two atoms due to the overlap of electron clouds is represented using $1/r^{12}$ term. Moreover, the attractive London dispersion interactions or instantaneous induced dipole-induced dipole interactions are represented using $-1/r^6$.

Most of these force field parameters are usually obtained using either experimental data or quantum mechanical calculations.[23] Equilibrium bond lengths and angles can be determined using experimental structural data, such as crystal structures and electron diffraction methods.[23] Multiplicities and phases can be obtained from quantum mechanical calculations.[23] Spectroscopic data such as IR, Raman and QM data are often used to obtain force constants.[23] Van der Waal parameters can be obtained from X ray diffraction and neutron diffraction.[23] In fixed charge force fields the electrostatic interactions are mainly expressed using Coulomb law.[23] The charge distribution around a molecule is then described by simple partial atomic charges.[23] Unfortunately, partial atomic charges are not experimental observables.[29] Moreover, there is no universally approved way of obtaining partial atomic charges in force fields.[18] The most common approach is to use QM electrostatic potentials to obtain partial atomic charges.[18, 30]

There are several popular force fields available to perform MD simulations of biological systems such as: CHARMM (Chemistry at HARvard Macromolecular Mechanics),[31] AMBER (Assisted Model Building with Energy Refinement),[20] GROMOS (GROningen MOlecular Simulation program package),[32] OPLS (Optimized Potentials for Liquid Simulations).[33]

With the CHARMM force field the partial atomic charges were obtained using minimum interaction energies and different interaction orientations of a water molecule with model compounds.[18, 23, 25] In this case they have used one water molecule and model compound at a time.[23, 25] Then they change the position of the water molecule in order to cover all the interaction sites.[23] In their geometric optimization process they have used HF/6-31G* basis set.[25] These partial atomic charges are optimized to reproduce the QM interaction offset distances, scaled energies and dipole moments. Experimental heats of vaporization and molecular volumes are generally used as the target data for the van der Waals parameters.[34] Force constants were optimized to reproduce experimental or QM vibrational spectra.[23] For dihedral parameter optimization, they have used conformational energetics of the model compounds that are derived using HF/6-31G* level or higher basis set.[23]

Usually, atom centered partial atomic charges are derived using the 6-31G* basis set and a restrained electrostatic potential (RESP) fitting in the AMBER force field.[20] These charges were tested to reproduce interaction energies, free energy of solvation, liquid enthalpies, densities and conformational energies of small molecules.[20] For small molecules bond angle parameters were obtained by fitting to structural and vibrational frequency data.[20] Subsequently, they were readjusted to reproduce experimental normal mode frequencies.[20] Lattice energies and crystal structures were used to obtain van der Waals parameters and tested against liquid properties.[20] Dihedral parameters were obtained from quantum mechanical data.[20]

The GROMOS force field parameters were developed using the following approach. Partial atomic charges were obtained from quantum mechanical electron densities and further optimized using experimental dielectric data. The crystal structures of small molecules were used to obtain equilibrium bond lengths and angles.[3, 21] Gas phase IR spectroscopic data were used to

obtain force constants.[21] Non bonded van der Waals parameters were obtained by fitting to the experimental thermodynamic data such as: heats of vaporization, densities of pure liquids and free energy of solvation.[3] Dihedral parameters were obtained using quantum mechanical dihedral potentials.[3, 21]

The OPLS force field parameters have been derived primarily to reproduce experimental liquid properties.[20, 33] Experimental densities and heats of vaporization of liquids have been used as the target data for parameter optimization.[33] Partial atomic charges were derived empirically to best represent the condensed phase properties of small molecules.[20, 35] Bond stretching and angle bending terms were adopted from AMBER force field. Structural parameters were tested against vibrational spectroscopy and diffraction data.[33] However, later, improved partial atomic charges were obtained by adjusting the partial atomic charges using QM data.[36] Here, they have used hydration free energies as the target data.[37] Rotational energy profiles obtained from ab initio molecular orbital calculations were used to determine torsional parameters.[35]

It is important to point out that most of these force fields were developed and tested using a specific water model. Therefore, when performing bimolecular simulations, we should use that particular water model with the selected force field. TIP3P[38], SPC[39] and SPC/E[40] are the most common water models that have been used widely in simulations. AMBER, OPLS and CHARMM force fields mainly use the TIP3P water model, whereas the GROMOS force field uses the SPC water model.[18]

As explained above, most of the biomolecular force fields have very similar equations for calculating potential energy of the system. Yet, the way that they derive parameters and the values of the parameters are drastically different. Most of these force field parameters were obtained for small molecules.[23] The small molecules are treated as the building blocks of biological

macromolecules such as proteins, lipids, nucleic acids etc. First, small molecular analogues are selected to represent the functional groups of the biomolecules. Then, force field parameters are developed to model these small molecules. Force fields for small molecular analogues are usually considered as transferable, assuming that these parameters can be assigned to larger molecules with similar chemical structures. Therefore, transferability is one of the significant assumptions in force field development.[23] Furthermore, additivity, which assumes the potential energy of the systems can be calculated using the sum of the above described potential terms, is also a very important assumption that used in force fields.[27] Hence, these fixed charge force fields are also known as additive force fields.[27]

1.2.1 Polarizable Force Fields

Most of the force fields use fixed atomic charges to represent the charge distribution around a molecule.[23] These are known as fixed charge, effective charge, or non polarizable force fields. As we know, real molecules should be able to change the charges according to the environment. Therefore, people have tried to embed this feature in to force fields for decades.[41-49] Molecular polarization can be divided in to three components: electronic polarization (resulting from the redistribution of electrons around an atom or molecule), geometric polarization (resulting from the changes of the molecular geometry), and orientational polarization (resulting from the rigid rotation of the molecule due to an electric field).[45,48] These three components of polarization are interrelated. Geometric polarization can be included by using flexible molecules.[45] Molecules are usually rotating during the simulation and therefore orientational polarization is always countered. Force fields that include electronic polarization explicitly are known as polarizable force fields.[45] There are several methods of including the explicit polarization to the system. Commonly, these methods can be categorize in to three main groups: induced point dipole models, fluctuation charge

models and Drude oscillator models.[45] The commonly used non polarizable force fields like CHARMM, OPLS, AMBER, GROMOS use one of the above methods in their polarizable versions of the force fields.[42,45,48-49]

In the induced point dipole model, polarization is represented using point dipoles that are added to some or all atomic sites of the molecule in the system. Generally atomic sites are atomic interaction centers. The interaction between the dipoles and the environment provide the total polarization of the system.[42,45,48-49]

The idea behind the fluctuating charge model is to treat the atomic partial charges as dynamic quantities. This model is able to produce molecular polarizability to all orders in the charge moments. Intermolecular and intramolecular charge transfer is allowed by this method and the electrostatic energy of the system is minimized at each step to find the instantaneous values of partial charges.[45]

The Drude oscillator model is also known as the shell model. In this model two charged particles are used to implement the polarizable site. One charge particle is defined as the heavy core particle and the other particle is a very light or massless shell particle. These two particles are connected using a harmonic spring that has a force constant k. The core charge and shell charge does not change during the simulation. The shell charge is defined as $-q_s$ ($q_s \geq 0$). The core charge is defined as $+q_s$. To represent the polarization in this model, the position of the core and shell particles change with respect to each other.[45, 49]

It is important to point out the main issues with these methods. The polarization "catastrophe" is one of the problems associated with the induced point dipole model.[45, 48] If two point dipoles are aligned closely in the same direction, it can result in unphysical, strong interactions, which is known as the polarization catastrophe. One way of avoiding the polarizable

catastrophe is by modifying the polarizability tensor so that the polarizability of the point dipoles are damped to a finite value at short seperations.[45] The main drawback of the fluctuation charge model is the nonphysical large charge transfer at large distances resulting in an overestimation of the effect resulting in large dipole moments for single molecules.[41] In practice, the fluctuation charge model allows a redistribution of charge only within each molecule, or just certain parts of the molecules.[45] In addition, this method does not mimic out of plane polarization in planar or linear chemical moieties, because the electronegativity equalization can only proceed along the bonds.[41]

Comparatively, polarizable force fields are computationally more demanding than non polarizable force fields.[44-45] Theoretically, these models are supposed to increase the accuracy of the simulation results. However, ongoing efforts are devoted to validating the polarizable force fields in bimolecular simulations.[44, 46, 49-50] Nevertheless, non polarizable force fields still play the main role in biomolecular simulations. This dissertation is solely based on non polarizable force fields.

1.2.2 Problems Associated with Currently Available Force Fields

Although there are several well established force fields available for MD simulations of biological systems, there is still significant room for improvement.[28] In the past few decades people have put a significant effort in order to obtain more accurate force field parameters. This includes changing the approach of obtaining force field parameters.[28] Most of the biologically important processes are affected by non bonded interactions. Hence, an accurate description of non bonded interactions is one of the main force field challenges.[28]

Binding free energies and protein ligand interaction energies can be used as a test to validate non bonded interactions. When MD is applied to computational drug design, it is possible to predict correct ligand receptor poses. However, quantitative prediction of binding free energies and binding affinities of different ligands is not always possible.[51-52] As the non bonded interactions are mainly responsible for these ligand receptor binding affinities one can argue that the non bonded parameters of currently available force fields need further optimization.[53] Protein folding is also one of the most important phenomena that have been studied extensively for the past three decades.[20, 54-57] Most proteins are only slightly stable at room temperature.[58-60] Most of the thermodynamic properties associated with protein folding require an accurate description of non bonded interactions.[28] The melting temperature, entropy and enthalpy contributions should be correctly predicted when we study protein folding with MD simulations. However, with current force fields it is somewhat difficult to reproduce all the thermodynamic properties to compare with experiment values.[28]

1.2.3 Towards the Development of a New Force Field

Our desire to develop more accurate force field for biomolecular simulations came in the early 90's from the study of cosolvent and biomolecules. Smith and coworkers have performed MD simulations of mixtures of small solutes, ions and biomolecules to obtain a deeper understanding of the Hofmiester series using several force fields.[61-63] Unfortunately, there was no straight forward way of validating the observed simulation results with experimental data, because there was no distinct binding sites available to obtain experimental structural data. There was a clear need of a way to validate computer simulation data with experimental thermodynamic data.[28] In order to overcome this problem, Smith and coworkers have decided to use Kirkwood Buff

theory, one of the most important theory of solutions, to relate computer simulation data to common thermodynamic properties. At this point I would like to discuss the details of Kirkwood Buff theory.

1.3 Theory of Solutions

In comparison to gases and solids, liquids and liquid solution mixtures display more complicated behavior due to the intermolecular interactions.[64] The behavior of gases and solids are less problematic as they have very weak and strong intermolecular interactions, respectively. Modeling liquids and solution mixtures is more challenging. However, many important systems are in the solution phase. Thus, studying properties of the condensed phase has become an important research area for many years.

There are mainly two theories for solutions: McMillan-Mayer theory (MM)[65] and Kirkwood Buff theory (KB).[66] The McMillan-Mayer theory was first derived in 1945 by W. G. McMillan and J. E. Mayer.[65] In the theory of imperfect gases we use an expansion of the pressure in a power series in the density. In MM theory, we use an expansion of the osmotic pressure in a power series in the solute density.[66] MM theory is exactly a generalization of the theory of imperfect gases. For a two component system of solute i and solvent j we have

$$\beta\pi = \rho_i + B_2^*(T, \lambda_j)\rho_i^2 + B_3^*(T, \lambda_j)\rho_i^3 + \cdots \tag{1.13}$$

where B_2^*, B_3^* … are the virial coefficients of the osmotic pressure that depend on temperature (T), $\rho_i = \frac{N_i}{V}$ is the number density of molecule i and λ_j is solvent activity. B_2^* and B_3^* require a knowledge of the pair correlation function and triplet correlation function, respectively. Thus, higher correlation functions are required for further terms in Equation (1.13). Unfortunately, there is no clear way of determining higher correlation functions.[66] Thus, MM theory can be applied

only for very dilute solutions. KB theory, is a more generalized theory that can be applied to solution mixtures at any concentration. A large part of this dissertation involves the use of KB theory.

1.3.1 Kirkwood Buff Theory/Fluctuation Theory

Kirkwood-Buff (KB) theory, which is also known as fluctuation theory, was derived by John G. Kirkwood and Frank P. Buff and published in 1951.[67] Unarguably, KB theory is the most important theory of solutions provides the best representation of a solution mixture. This theory contains several advantages. KB theory is an exact theory, without any approximations, that can be applied to any stable solution mixture that contains any number of components. Additionally, it can be applied to molecules of any size and any concentration.[28, 66]

In KB theory, we provide relationships between common thermodynamic quantities and the particle distribution functions in an equivalent grand canonical ensemble (μVT).[66] Hence, it is possible to obtain the isothermal compressibility, partial molar volumes, and derivatives of the chemical potential of a liquid solution mixture from the details of underlying molecular distributions.[66]

The pair correlation function, which is also known as radial distribution function (rdf), is the most important function in the theory of liquids. The radial distribution function, $g_{ij}(r)$, is a measure of the probability of finding a j particle at a distance r around a central i particle.[66] Importantly, it is a measure of how the density of a particular species varies with the distance from a selected central particle. The radial distribution function can be measured using center of mass-center of mass or atom-atom distances.

Figure 1.3 illustrates a typical rdf that we obtain for a liquid. It is evident that at very short distances the rdf is zero. This area is known as the exclusion volume. It means that there is no probability to find another particle at this small distance of r due to strong electron-electron repulsion. After that we may see the first solvation sphere. This peak represents a high density of finding particles from the central particle. The second solvation sphere is usually a broader peak than the first solvation sphere. We may see several solvation spheres that vary in character with the system. After a few molecular diameters the rdf converges to 1, which indicates the local distribution approaches the bulk distribution.

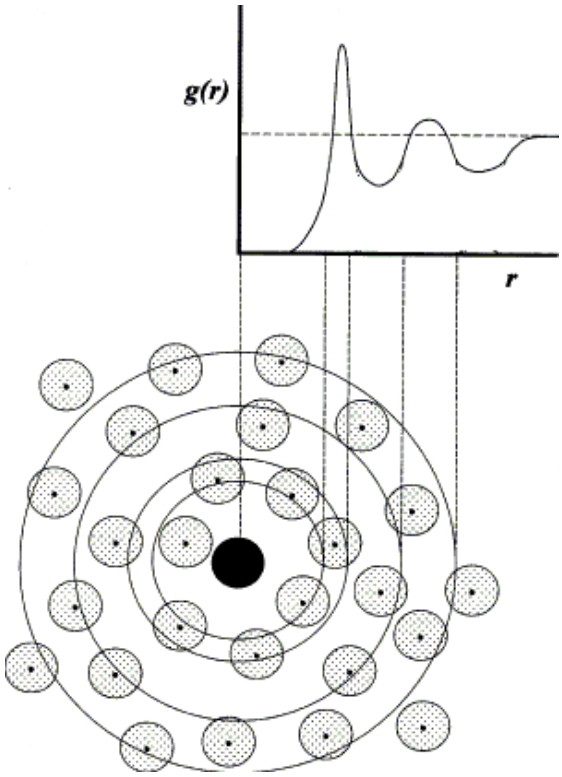

Figure 1.3 Schematic diagram of radial distribution function[68]

The radial distribution function can be obtained from neutron and X-ray scattering experiments.[69-71] The rdf is a measure that provides structural information concerning a liquid.[66]

The average number of j particles in a spherical volume of radius r and width dr, around a central j particle can be obtained by integration of the rdf to a distance R. This is known as the coordination number,

$$CN_{ij} = \rho_j \int_0^R g_{ij}(r) 4\pi r^2 dr \tag{1.14}$$

A KB integral can be defined using the radial distribution function in the grand canonical ensemble as shown in Equation (1.15),

$$G_{ij} = 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 dr \tag{1.15}$$

where G_{ij} is the Kirkwood Buff integral (KBI) between i and j species, $g_{ij}^{\mu VT}$ is the corresponding center of mass-center of mass radial distribution function in a grand canonical ensemble. From the value and the sign of the KB integrals we can obtain an idea about how the species i and j interact with each other in a solution. A value of G_{ij} >0 arises due to a favorable net interaction between i and j. In other words, an excess of j molecules around i molecules is observed. On the other hand, when G_{ij} <0 an unfavorable interaction between species i and j is observed and represents a depletion of j molecules around a central i molecule. Generally speaking, from the KB integrals we can obtain information concerning the local properties of a solution mixture. As explained above, typically g_{ij} converges to one within a few molecular diameters, except for systems near critical points or for solids.[28] Hence, the value of G_{ij} is mainly determined by small local regions around a central i molecule. An excess coordination number N_{ij} can be obtained from the KB integrals as shown in Equation (1.16).

$$N_{ij} = \rho_j G_{ij} = \rho_j 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 dr \tag{1.16}$$

As we can see above the KB integrals are derived for an open system. Most of the MD simulations are performed in closed systems (NpT, NVT, NVE ensembles). For closed systems, G_{ij}=-1 and 0 for i=j and i≠j respectively. Therefore, it is not possible to use Equation (1.15) directly for closed systems.[66] Although, theoretically it is possible to perform MD simulations in an open system, there are some complications with the calculations due to particle insertion during the simulations.[72-75] Thus, a different approach has been considered to obtain KB integrals in a closed system (NpT) simulation.[76] Here, the idea is to truncate g_{ij}^{NpT} at a distance R, where the rdf converges to one. In this case, an important assumption has to be made that the truncated g_{ij}^{NpT} contains all the necessary features of the full $g_{ij}^{\mu VT}$.Therefore, we can use Equation (1.17) employing the approximation[76-78]

$$G_{ij} = 4\pi \int_0^R [g_{ij}^{NpT}(r) - 1] r^2 dr \tag{1.17}$$

where, R is the distance that the rdf converges to one. This approximation seems to be a reasonable one and a very important assumption for the study of closed systems.[66]

The particle fluctuation approach is another method that we can use to obtain the KB integrals.[66]

$$G_{ij} = V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) \tag{1.18}$$

where N_i is the number of i particles inside V, <...> denotes grand canonical averages, and δ_{ij} is the Kronecker delta. Recently, Schnell and coworkers have used an approach based on particle number fluctuations to obtain KB integrals for finite volumes.[79-81] Here, an expression for the finite volume KB integrals are obtained for a finite volume (cube or sphere) and then these integrals are linearly extrapolated to obtain a value corresponding to an infinite system.[79-81] This approach can be beneficial for systems that contain convergence problems with the traditional expression.[79] In our experience, both approaches result in similar values for the KB integrals.

Using KB theory it is possible to link macroscopic thermodynamic properties to the distribution of the molecules in solution mixtures. For a stable binary solution that contains species 1 and 2, the relationships between the three KB integrals (G_{11}, G_{22}, G_{12}) and the thermodynamic properties ($\bar{V}_1$ partial molar volume of species 1, $\bar{V}_2$ partial molar volume of species 2, κ_T isothermal compressiblity of the solution mixture, μ_{12} derivative of the chemical potential of species 1 with respect to the number of molecules of species 2 while holding the fixed number of molecules 1) can be obtained using following equations.[66]

$$\bar{V}_1 = \frac{1 + \rho_2(G_{22} - G_{12})}{\eta} \tag{1.19}$$

$$\bar{V}_2 = \frac{1 + \rho_1(G_{11} - G_{12})}{\eta} \tag{1.20}$$

$$\kappa_T = \frac{\xi\beta}{\eta} \tag{1.21}$$

$$\mu_{12} = \left(\frac{\partial \mu_1}{\partial N_2}\right)_{P,T,N_2'} \tag{1.22}$$

$$a_{11} = \left(\frac{\partial ln a_1}{\partial ln \rho_1}\right)_{p,T} = 1 + \left(\frac{\partial ln y_1}{\partial ln \rho_1}\right)_{p,T} = \frac{1}{1 + \rho_1 (G_{11} - G_{12})} \tag{1.23}$$

$$f_{11} = \left(\frac{\partial ln f_1}{\partial ln x_1}\right)_{p,T} = -\frac{\rho_2 x_1 (G_{11} + G_{22} - 2G_{12})}{1 + \rho_2 x_1 (G_{11} + G_{22} - 2G_{12})} \tag{1.24}$$

where,

$$\eta = \rho_1 + \rho_2 + \rho_1 \rho_2 (G_{11} + G_{22} - 2G_{12}) \tag{1.25}$$

$$\xi = 1 + \rho_1 G_{11} + \rho_2 G_{22} + \rho_1 \rho_2 (G_{11} G_{22} + -G_{12}^2) \tag{1.26}$$

where $\beta = 1/(RT)$, R is gas constant, a_1 is the molar activity of the species 1, f_1 is the mole fraction scale activity coefficient of the species 1, and x_1 is the mole fraction of the species 1.

1.3.2 Inversion of the Kirkwood Buff theory

Although KB theory is the most important theory of solutions, it was not widely used in the early days. At the time this theory was first derived, obtaining the radial distribution function for solution mixtures was demanding. The lack of computational power and inaccessibility of the experimental diffraction techniques were the main reasons behind the difficulty of obtaining the rdf for desired solution mixtures.

Ben-Naim proposed an inversion of the KB theory in 1977 and this represents a turning point in the wide spread popularity of KB theory.[66, 82] The idea of the KB inversion procedure is to reverse the calculations of the original KB theory. In other words, calculating KB integrals from available thermodynamic data. The two theories can be expressed as follows. For KB theory we have

$$\{G_{ij}\} \rightarrow \{\bar{V}_i, \kappa_T, \partial \mu_i / \partial \rho_i\}$$

while for Inversion of KB theory we have

$$\{\bar{V}_i, \kappa_T, \partial\mu_i/\partial\rho_i\} \rightarrow \{G_{ij}\}$$

After the inversion of the Kirkwood Buff theory appeared this theory was more popular among scientific community because it is comparatively easier to measure bulk thermodynamic properties rather than obtaining the rdf of solution mixtures. The experimental KB integrals can then be obtained using experimental quantities as shown in following equations.

$$G_{12} = kT\kappa_T - \frac{\rho\bar{V}_1\bar{V}_2}{D} \tag{1.27}$$

$$G_{11} = kT\kappa_T - \frac{1}{\rho_1} + \frac{\rho_2\bar{V}_2^2\rho}{\rho_1 D} \tag{1.28}$$

where

$$D = \frac{x_1}{kT}\left(\frac{\partial\mu_1}{\partial x_1}\right)_{p,T} \tag{1.29}$$

However, one major issue with the calculation of KB integrals using thermodynamic data is that these calculated KB integrals are very sensitive to the accuracy of the thermodynamic data.[28] To obtain a better understanding of the uncertainty of the KB integrals, it is very important to have an idea of a relative importance of the input thermodynamic data. The compressibility, partial molar volumes and derivatives of the chemical potential are the thermodynamic properties that we use to obtain KB integrals. Derivatives of the chemical potential can be related to the activity coefficients and excess Gibbs free energy as shown in following equations.

$$\beta\left(\frac{\partial\mu_1}{\partial lnx_1}\right)_{p,T} = 1 + \left(\frac{\partial lnf_1}{\partial lnx_1}\right)_{p,T} = 1 + f_{11} \tag{1.30}$$

$$g^{EX} = \frac{G^{EX}}{N_1 + N_2} = x_1\mu_1 + x_2\mu_2 - x_2(\mu_1^P + kTlnx_1) - x_2(\mu_2^P + kTlnx_2) \quad (1.31)$$

$$D = \frac{x_1}{kT}\left(\frac{\partial \mu_1}{\partial x_1}\right)_{p,T} = 1 + \frac{x_1 x_2}{kT}\left(\frac{\partial^2 g^{EX}}{\partial x_1^2}\right)_{p,T} \quad (1.32)$$

The accuracy of the excess Gibbs free energies or activity coefficients are very important to obtain the correct KB integrals for solution mixtures.[83] Generally, the contribution from the $kT\kappa_T$ term is very small when we calculate G_{12} for solution mixtures. The partial molar volume data provides a moderate impact on the G_{12} values. The key quantity that is needed precisely is the excess Gibbs free energy or activity coefficient data since the derivation of KB integrals required multiple differentiations of the functions that include these properties.[28] Furthermore, it is necessary to have a reliable model equation that can accurately fit the experimental data. The Wilson, NRTL, van Laar and Redlich-Kister are some of the most popular fitting equations widely used to fit activity coefficient data.[84]

1.4 Towards Accurate Force Fields Based on Kirkwood Buff Theory

Here, we would like to discuss our ongoing effort to develop more accurate force fields based on Kirkwood Buff theory. As we discussed above, current force fields need to be further optimized to obtain more reliable simulation data. Although KB theory is the most important theory that describes the relative distribution of components in a solution phase, people have not used this theory when they derive or test force field parameters for the condensed phase. Generally, they use quantum mechanically derived charges and scaled them to obtain macroscopic thermodynamic properties.[28] Furthermore, most of these force field parameters are based on pure liquids and not solution mixtures. Several studies have been carried out to investigate the ability

to reproduce the experimental KB integrals with commonly available force fields. These studies have shown that experimental KB integrals for solution mixtures were not well reproducible.[76, 85-86] In other words, the correct balance between solvent-solvent, solvent-solute and solute-solute molecules are not represented. Furthermore, an incapability to model the correct cosolvent biomolecule distribution was also found in these studies.[87] Although the obtained KB integrals vary drastically among different force fields, most of the other traditional properties such as densities, diffusion coefficients, dielectric properties gave somewhat similar values.[76, 88]

Previously, Smith and coworkers have performed a case study in order to check the accuracy of the current biomolecular force fields by studying binary solution mixtures.[28] The solutes selected were small molecular analogues for the amino acid side chains. They have studied following binary mixtures: methanol-water, benzene-methanol, N-methyl acetamide-water, zwitterionic glycine-water to represent two polar molecules, aromatic amino acid side chains, peptide group, salt-bridge forming ions, respectively.[28] Furthermore, the benzene-methanol mixture was chosen to model phenyl alanine molecule in two different environment (solvent exposed and buried environment). The KBIs were calculated using the AMBER99sb[89], CHARMM27[90-91], OPLSAA[33, 92], GROMOS54a7[93] force fields. These force fields have been continuously updating their force field parameters in an effort to improve their simulation results. Some force fields were able to well reproduce the experimental KBI for particular systems. However, the most important outcome of this study was that none of the force fields were capable of reproducing the correct experimental KBI (excess coordination numbers) for all four systems. Generally, the deviation from the experimental data was too positive for the solute-solute excess coordination numbers in aqueous mixtures.[28] This indicates the lack of a correct representation of interaction between the components in solution mixtures.[28]

When we consider current biomolecular force fields, most of the uncertainly lies in the representation of the electrostatic interactions.[28] Ordinarily, non polarizable force fields use Coulomb interactions to calculate electrostatic interactions. The argument is that partial atomic charges that are derived from gas phase quantum mechanical calculations are not able to reproduce the correct condense phase interactions. Thus, it is very important to use the properties of solution mixtures when obtaining partial atomic charges to use in condense phase molecular dynamics simulations. Moreover, instead of just studying the properties of pure liquids, we need to focus on the properties of solution mixtures over a wide composition range. For example, the free enthalpy of solvation and free energy of solvation of model compounds has been used to validate force field parameters.[21, 36, 94-95] However, these parameters only provide the details about solute-solvent interactions. When considering solution mixtures, the solute activity is the main thermodynamic property which describes the correct balance between solvent-solvent and solute-solute interactions. The composition dependent changes of the solute activity can be used as a guide to force field validations.[96] The KB integrals are very sensitive to the charge distribution of the molecule.[97] Based on this idea, Smith and coworkers have been trying to obtain a new force field for biomolecular simulations based on Kirkwood Buff integrals.[26, 87, 96, 98-105] This force field is well known as the Kirkwood Buff derived Force Field (KBFF). It is a non polarizable force field. In this approach, the experimental KB integrals were determined and compared with the simulated KB integrals. During the parametrization, the partial atomic charges were adjusted to reproduce the experimental KB integrals as shown in Figure 1.4.

Here, indirectly we try to mimic the correct condensed phase polarization using an effective charge distribution. We may have to test several charge distributions to obtain the optimized charge distribution that gives the closest agreement with the experimental KB integrals. The ultimate goal

is to perform more accurate MD simulations of proteins. As a first step, the Kirkwood Buff derived force fields were developed to represent small molecular analogues of proteins. Table 1.1 shows the completed models to date.

On the other hand, the KB integrals provide additional information about the solution mixtures. For example, the composition dependent changes in the interactions between molecules in the system. Some mixtures show higher self-associations at certain compositions. If this to be correctly reproduced using models, it will be helpful to validate simulation data compared to experiments.[28]

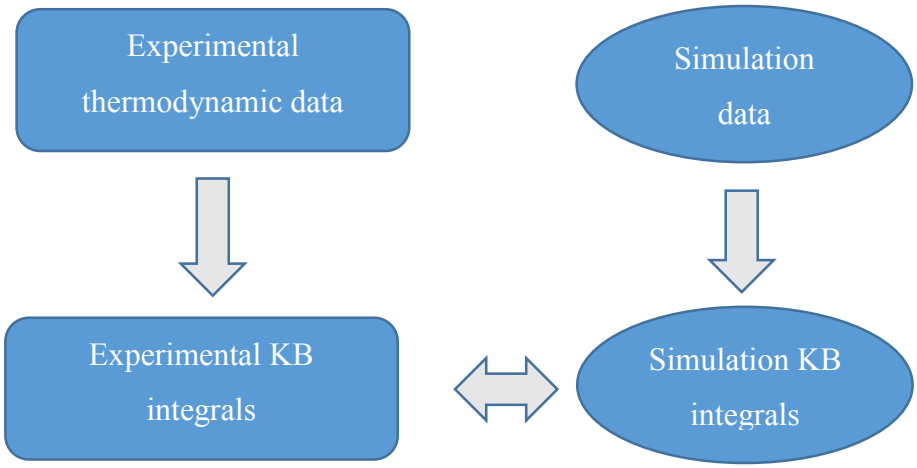

Figure 1.4 Scenario of parameterizing the force field

Despite all the advantages of using KB theory in the force field parametrization, there are some practical difficulties with this approach.[28] Comparatively larger simulation boxes may be necessary for aggregating systems, although with the current development of computational power this is not such a big problem. Sometimes, it is not possible to obtain converged KB integrals with short simulations. Therefore, longer simulation times are needed to obtain converged KB integrals.

Table 1.1 Currently available KBFF parameters

Solute	Solvent	Reference
Acetone	water	98
Urea	water	99
NaCl	water	100
Guanidinium Chloride	water	101
Methanol	water	102
Amides	water	87
Thiols and sulfides	methanol	103
Aromatics, Heterocycles	methanol, water	104
Alkali halides	water	105
Alcohols	water	To be published
Amino acids	water	To be published
Alkaline earth halides	water	To be published
Esters	water, methanol, ethanol	To be published

Several charge distributions should be tested to find the optimized effective charges. There is no well defined way of adjusting the charges. It is totally dependent on experience and basic chemical knowledge. One of the biggest drawbacks is the lack of reliable experimental data for a wide range of systems to obtain KB integrals.[28]

KB integrals have become one of the most important pieces of data to check the validity of condensed phase force field parameters. For example, Mackeral and coworkers have studied their new polarizable force field parameters using KB integrals.[106] They have been developing a polarizable force field based on Drude ocillaters.[106] They have done simulations of NMA water mixtures with their polarizable force field and non polarizable force fields. According to their simulation results the polarizable force field reproduces the experimental KB integrals more

accurately compared to an additive force field.[106] However, they couldn't reproduce activity derivatives and Gibbs free energy data. In contrast, our KBFF force field could get reasonable agreement with all the data, although it is a non polarizable force field.[106]

One might argue that polarizable force fields would reproduce the experimental KB integrals more accurately than non polarizable force fields. Theoretically, this argument is logical because when we include explicit electronic polarization it should represent the charge distribution more accurately. Therefore all the interactions should be mimicked correctly. Generally, polarizable force fields are more computationally expensive than non polarizable force fields due to increase number of charge sites, complex potential energy functions, small time steps, etc.[41] Therefore, the computational cost is one of the main drawbacks associated with polarizable force fields when performing longer biomolecular simulations. As explained above, the KBFF non polarizable NMA model could reproduce the KB integrals and other important properties with the same accuracy as a polarizable model. Therefore, we believe our KBFF models represent the balance between solvent and solute interactions more accurately.

1.5 Preferential Solvation in Solution

Preferential interactions are one of the most promising aspects of liquid solution mixtures. Most experiments are carried out in condensed phase. Therefore, solvation is an important concept. In a liquid solution mixture, the local composition around any molecule is different from the bulk composition, and this idea leads to the concept of preferential interactions and preferential solvation.[66] It is possible to obtain a clear understanding of preferential solvation using Kirkwood Buff theory and it is one of the most important applications of the KB integrals. Ben Naim has described a general approach to define preferential solvation based on the concept of solvation

thermodynamics and KB integrals.[107-109] However, several others have modified the original derivation and this has resulted in a great argument regarding the subject.[28]

Different approaches for treating preferential solvation in solution mixtures have been reported in the literature.[110-111] Some of the earliest approaches were not based on the concept of solvation thermodynamics. For example, Grunwald and coworkers have studied dioxone-water mixtures in the presence of electrolytes.[110] Here, they have calculated the standard partial molar free energy by changing the solvent.[110] According to the traditional concept of solvation, this phenomena could only be applied to very dilute systems; such as three (or more component system). In 1983, Marcus proposed a model to obtain preferential interactions using quasi lattice, quasi chemical theory (QLQC).[112] The preferential solvation of ions in mixed solvents were studied. Quasi lattice theory uses the number of nearest neighbors (lattice parameter) of the pure components to determine the lattice parameters of the components in a solution mixture. The number of nearest neighbors of a molecule (solute or solvent) in a mixture can be calculated from the weighted mean of the lattice parameters of pure components. Moreover, the theory assumes the interaction energies are independent of the other neighbors leading to ideal volumes and an ideal entropy of mixing takes place. The relationship between the number of unlike neighbors to the number of like neighbors and the interaction energies were given by quasi chemical theory. Molar volumes and excess Gibbs energies are required as input in this approach. Several binary mixtures were studied using this method.[113] Qualitative agreement between QLQC approach and KBI approach was discussed using binary solution mixtures.[114] UV-visible spectroscopic methods can also be used to obtain the preferential solvation parameters of mixed solvents.[115-119] In this approach, the transition energy value corresponding to the maximum absorption in mixed solvents is the measure of local composition.

Preferential solvation helps us to understand many chemical and physical properties in solution mixtures such as chemical reactivity, spectroscopy, diffusion coefficients, etc. We may have a three component solution mixture with a solute S in a mixture of solvents A and B. It would be interesting to determine the effect from each solvent component of the system on the properties of the solute. Our property of interest (σ_{AB}) for solute S in the mixed solvent can be expressed as,

$$\sigma_{AB} = x_A\sigma_A + x_B\sigma_B \tag{1.33}$$

where σ_A, σ_B values are our property of interest in pure solvents A and B, respectively, and x_A, x_B are mole fraction of component A and B, respectively.[66] Equation (1.33) would be valid for a perfect gas system where all the intermolecular interactions are negligible.[66] Nevertheless, there are non-negligible interactions between the solute S and solvent components. Therefore, it is possible to define a spherical volume around the solute molecule S, where the solvent composition is different from the bulk composition. This is known as the local mole fraction. With this quantity it is possible to rewrite Equation (1.33) as,

$$\sigma_{AB} = x_A^L\sigma_A + x_B^L\sigma_B \tag{1.34}$$

where x_A^L and x_B^L are the local mole fraction of component A and component B, respectively.

Preferential solvation of the solvent A molecules around solute S can be defined by δ_{AS},

$$\delta_{AS} = x_A^L - x_A \tag{1.35}$$

where x_A^L is the local mole fraction of A component around solute S.

A general explanation of preferential solvation based on local mole fractions was derived by Ben-Naim using KB theory. In particularly, we may consider a binary solution mixture that

contains components i and j. It is possible to define a spherical volume around the central particle i beyond which the distribution of j molecules around i molecule is equal to the bulk distribution. This volume can be defined as the correlation volume V_{cor}. The average number of j particles around a central i particle within the correlation volume can be defined using Equation (1.36),

$$\langle N_j \rangle = N_{ij} + \rho_j V_{cor} \tag{1.36}$$

The local mole fraction of j particles within the correlation volume is then given by,

$$x_j^L = \frac{\langle N_j \rangle}{\sum_k \langle N_k \rangle} = \frac{N_{ij} + \rho_j V_{cor}}{\sum_k (N_{ik} + \rho_k V_{cor})} \tag{1.37}$$

Thus, the preferential solvation of i particle by j particles can be expressed using KB integrals as shown in Equation (1.38),

$$\delta x_{ji} = x_j^L - x_j = \frac{x_j \sum_k x_k (G_{ij} - G_{ik})}{V_{cor} + \sum_k x_k G_{ik}} \tag{1.38}$$

There are some attempts to obtain the correlation volume explicitly to determine the preferential parameters. However, Ben Naim used an approach to obtain a qualitative understanding of the preferential solvation by expanding in a power series about $1/V_{cor}$. The first order coefficient of the power series expansion is δ_{ji}^0,

$$\delta_{ji}^0 = x_j^L - x_j = \frac{x_j \sum_k x_k \left(G_{ij} - G_{ik}\right)}{V_{cor}} + \cdots \tag{1.39}$$

The sign of the numerator of the above equation provides the sign of the preferential solvation as we approach from an infinitely large distance. According to his view the sign gives the information about the relative distribution of species in the solution. Thus, the magnitude is less important.

However, various approaches have been suggested to obtain V_{cor} explicitly and these are further discussed in Chapter 4.

In 1997 Matteoli suggested correcting the KB integrals with respect to a reference mixture in order to determine the preferential interaction.[120] His argument was that for an ideal solution mixture, where all the interactions are equal, there should be no preferential interaction. However, one can get a value for the preferential interactions using Equation (1.39) due to the difference in sizes of the molecules. A great debate has been going on concerning this correction.[37] Ben Naim has critically explained the meaninglessness of these corrections.[37, 121] Furthermore, according to his opinion the KB integrals provide information about local densities around a central particle without providing any explicit information about molecular interactions.[37] He further pointed out that preferential interactions and ideal solution behavior are two different properties of the solution and no one should be surprised to see preferential solvation in an ideal solution. According to the definition of the ideal solution mixtures by Ben Naim the necessary condition is

$$\Delta_{ij} = G_{ii} + G_{jj} - 2G_{ij} = \left(G_{ii} - G_{ij}\right) + \left(G_{jj} - G_{ij}\right) = 0 \tag{1.40}$$

However, Matteoli's correction has been followed by several other authors.[122] A detailed discussion of this area can be found in Chapter 4.

1.6 Summary and Organization of the Dissertation

Computer simulations play a very important role in modern scientific research. We have used the most important theory of solution, Kirkwood Buff theory, in order to develop new force field parameters and to validate existing force field parameters. Additionally, a promising application of KB theory to investigate preferential solvation will be discussed.

In Chapter 2 the Kirkwood Buff theory is used to validate a set of charge models. Molecular mechanics force fields use relatively simple partial atomic charges that are assigned at nuclear sites in order to represent the charge distribution of molecules. This represents a compromise between accuracy and computational efficiency. It is important, therefore, to establish whether force fields are improved by the addition of these extra charge sites, especially due to the increased computational effort that is required. Mobley and coworkers have studied set of chloroethane molecules with and without extra charge models. In their study they have used the hydration free energy to validate the charge models. Here, we examine if these models represent a significant improvement in the known liquid state properties of mixtures of these molecules. In particular, approaches based on Kirkwood-Buff theory of solution mixtures have become important tests for developing and validating force field parameters.

In Chapter 3 a Kirkwood Buff derived force field for esters is developed. The main objective of this work is to obtain a correct charge distribution to represent the ester linkage of the phospholipid molecule. The long term objective is to perform more accurate simulations of protein membrane systems. Methyl acetate-water, methyl acetate-methanol, methyl acetate-ethanol, ethyl acetate-methanol, methyl propionate-methanol systems are studied to obtain optimized partial atomic charges for the ester linkage. We have used alcohol solvents because most of these esters are immiscible in aqueous medium. The models are further validated using comparison with other thermodynamic and transport properties.

In Chapter 4 we derive a new approach to obtain preferential interactions in binary and ternary mixtures using Kirkwood Buff integrals. Preferential interactions are one of the important phenomena that describe the properties of liquid solution mixtures. There are several distinct approaches derived to obtain these parameters and a great argument is ongoing about this subject

area. Here, we have compared our new approach with the existing approaches. Several binary solution mixtures are used to test and validate the different approaches.

In Chapter 5 we have used an approach based on Fluctuation Solution Theory (FST) to investigate the validity of the Kirkwood Superposition Approximation (KSA), which is an important approximation that appears in many liquid state theories. Recent advances in FST have provided rigorous expressions for integrals over the triplet and pair distributions from bulk thermodynamic data. A combination of the FST and experimental rdfs are used to examine the KSA at a series of state points for pure water. Moreover, several other approximations between triplet and pair distribution are also tested. The analysis indicates, that it is possible to obtain good agreement with the fluid thermodynamic result for regions of the phase diagram where the compressibility is small. Furthermore, the distant dependent accuracy of these approximations were further explored using MD simulation data.

In Chapter 6 we have provided summary and future work.

1.7 References

1. Lumb, J. *Mol. Cell. Biochem.* **1987,** *73* (2), 91-98.

2. Allen, M. P. *Introduction to Molecular Dynamics Simulation*. John von Neumann Institute for Computing: Computational Soft Matter, 2004; Vol. 23.

3. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006,** *45* (25), 4064-4092.

4. van Gunsteren, W. F.; Dolenc, J. *Biomolecular simulation: historical picture and future perspectives*. 2008; Vol. 36, p 11-15.

5. van Gunsteren, W. F.; Dolenc, J.; Mark, A. E. *Curr. Opin. Biotechnol.* **2008,** *18* (2), 149-153.

6. Karplus, M.; Petsko, G. A. *Nature* **1990,** *347* (6294), 631-639.

7. Rapaport, D. C. *The Art of Molecular Dynamics Simulations*. Cambridge University Press: 1995.

8. LaBerge, L. J.; Tully, J. C. *Chem. Phys.* **2000,** *260* (1–2), 183-191.

9. Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. *Q. Rev. Biophys.* **2012,** *45* (4), 427-491.

10. Karplus, M.; McCammon, J. A. *Nat. Struct. Mol. Biol.* **2002,** *9* (9), 646-652.

11. Durrant, J. D.; McCammon, J. A. *BMC Biol.* **2011,** *9* (1), 1-9.

12. Kandt, C.; Ash, W. L.; Peter Tieleman, D. *Methods* **2007,** *41* (4), 475-488.

13. Lindorff-Larsen, K.; Trbovic, N.; Maragakis, P.; Piana, S.; Shaw, D. E. *J. Am. Chem. Soc.* **2012,** *134* (8), 3787-3791.

14. Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*. Oxford University Press: 1987.

15. Verlet, L. *Phys. Rev.* **1967,** *159* (1), 98-103.

16. Giovanni Ciccotti, R. K.; Sergi, A. In *Handbook of Materials Modeling*, Yip, S., Ed. Springer: 2005; pp 745–761.

17. Setting up and running a Molecular Dynamics Simulations. http://www.ch.embnet.org/MD_tutorial/pages/MD.Part3.html (accessed February 28, 2016).

18. Mackerell, A. D. *J. Comput. Chem.* **2004,** *25* (13), 1584-1604.

19. Karplus, M. *J. Comput. Chem.* **1983,** *4*, 187217.

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* (19), 5179-5197.

21. Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004,** *25* (13), 1656-1676.

22. van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. R. Soc. Interface.* **2008,** *5* (Suppl 3), 173-190.

23. Becker, O. M. M.; MacKerell, A. D. J.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics* Marcel Dekker, inc.: New York, 2001.

24. Monticelli, L.; Tieleman, D. P. Force Fields for Classical Molecular Dynamics. In *Biomolecular Simulations*, E. S., Ed. Humana Press: 2013; Vol. 924, pp 197-204.

25. Zhu, X.; Lopes, P. E. M.; MacKerell, A. D. *Wiley Interdiscip Rev Comput Mol Sci* **2012,** *2* (1), 167-185.

26. Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. In *Developing force fields from the microscopic structure of solutions: the Kirkwood-Buff approach*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 55-76.

27. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. *Chem. Rev.* **2016**.

28. Smith, P. E.; Matteoli, E.; O'Connell, J. P. *Fluctuation Theory of solutions*. CRC Press: Boca Raton, 2013.

29. Li, J.; Zhu, T.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem. A* **1998,** *102* (10), 1820-1831.

30. Andres Cisneros, G.; Karttunen, M.; Ren, P.; Sagui, C. *Chem. Rev.* **2013,** *114*, 779-814.

31. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

32. Scott, W. R. P.; Hünenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Krüger, P.; van Gunsteren, W. F. *J. Phys. Chem. A* **1999,** *103* (19), 3596-3607.

33. Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988,** *110* (6), 1657-66.

34. Feller, S. E.; Yin, D.; Pastor, R. W.; MacKerell, A. D. *Biophys. J.* **1997,** *73* (5), 2269-2279.

35. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.

36. Xu, Z.; Luo, H. H.; Tieleman, D. P. *J. Comput. Chem.* **2007,** *28* (3), 689-697.

37. Ben-Naim, A. *J. Phys. Chem. B* **2007,** *111* (11), 2896-2902.

38. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

39. Ahlström, P.; Wallqvist, A.; Engström, S.; Jönsson, B. *Mol. Phys.* **1989,** *68* (3), 563-581.

40. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

41. Cieplak, P.; Dupradeau, F. Y.; Duan, Y.; Wang, J. M. *J. Phys.: Condens. Matter* **2009,** *21* (33).

42. Halgren, T. A.; Damm, W. *Curr. Opin. Biotechnol.* **2001,** *11* (2), 236-242.

43. Harder, E.; Anisimov, V. M.; Whitfield, T. W.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2008,** *112* (11), 3509-3521.

44. Baker, C. M. *Wiley Interdiscip Rev Comput Mol Sci* **2015,** *5* (2), 241-254.

45. Antila, H. S.; Salonen, E. *Methods Mol. Biol. (N. Y., NY, U. S.)* **2013,** *924* (Biomolecular Simulations), 215-241.

46. Chowdhary, J.; Harder, E.; Lopes, P. E. M.; Huang, L.; MacKerell, A. D.; Roux, B. *J. Phys. Chem. B* **2013,** *117* (31), 9142-9160.

47. Stern, H. A.; Kaminski, G. A.; Banks, J. L.; Zhou, R.; Berne, B. J.; Friesner, R. A. *J. Phys. Chem. B* **1999,** *103* (22), 4730-4737.

48. Yu, H.; van Gunsteren, W. F. *Comput. Phys. Commun.* **2005,** *172* (2), 69-85.

49. Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Theor. Chem. Acc.* **2009,** *124* (1-2), 11-28.

50. Zgarbová, M.; Rosnik, A. M.; Luque, F. J.; Curutchet, C.; Jurečka, P. *J. Comput. Chem.* **2015,** *36* (25), 1874-1884.

51. Lazaridis, T. *Curr. Org. Chem.* **2002,** *6* (14), 1319-1332.

52. Bonnet, P.; Bryce, R. A. *Protein Sci.* **2004,** *13* (4), 946-957.

53. Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins: Structure, Function, and Bioinformatics* **2002,** *47* (2), 194-208.

54. Bowman, G. R.; Voelz, V. A.; Pande, V. S. *Curr. Opin. Biotechnol.* **2011,** *21* (1), 4-11.

55. Chen, Y.; Ding, F.; Nie, H.; Serohijos, A. W.; Sharma, S.; Wilcox, K. C.; Yin, S.; Dokholyan, N. V. *Arch. Biochem. Biophys.* **2008,** *469* (1), 4-19.

56. Englander, S. W.; Mayne, L. *Proc. Natl. Acad. Sci.* **2014,** *111* (45), 15873-15880.

57. Wirth, A. J.; Liu, Y.; Prigozhin, M. B.; Schulten, K.; Gruebele, M. *J. Am. Chem. Soc.* **2015,** *137* (22), 7152-7159.

58. Dill, K. A. *Biochemistry* **1990,** *29* (31), 7133-7155.

59. Dill, K. A.; Ghosh, K.; Schmit, J. D. *Proc. Natl. Acad. Sci.* **2011,** *108* (44), 17876-17882.

60. Acampora, G.; Hermans, J. *J. Am. Chem. Soc.* **1967,** *89* (7), 1543-7.

61. Smith, P. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1991,** *113* (16), 6029-6037.

62. Smith, P. E.; Pettitt, B. M. *Biopolymers* **1992,** *32* (12), 1623-1629.

63. Smith, P. E.; Marlow, G. E.; Pettitt, B. M. *J. Am. Chem. Soc.* **1993,** *115* (16), 7493-7498.

64. Ploetz, E. A.; Smith, P. E. Local Fluctuations in Solution: Theory and Applications. In *Adv. Chem. Phys.*, John Wiley & Sons, Inc.: 2013; pp 311-372.

65. McMillan, W. G.; Mayer, J. E. *J. Chem. Phys.* **1945,** *13* (7), 276-305.

66. Ben-Naim, A. *Molecular theory of solutions*. Oxford university press, 2006.

67. Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951,** *19*, 774-7.

68. http://en.wikipedia.org/wiki/File:Lennard-Jones_Radial_Distribution_Function.svg-RDF

69. Soper, A. K. *Chem. Phys.* **2000,** *258* (2), 121-137.

70. Soper, A. K. *isrn phys. chem.* **2013,** *2013*, 67.

71. Clark, G. N. I.; Hura, G. L.; Teixeira, J.; Soper, A. K.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (32), 14003.

72. Çagin, T.; Pettitt, B. M. *Mol. Phys.* **1991,** *72* (1), 169-175.

73. Çagin, T.; Pettitt, B. M. *Mol. Simul.* **1991,** *6* (1-3), 5-26.

74. Beutler, T. C.; Mark, A. E.; van Schaik, R. C.; Gerber, P. R.; van Gunsteren, W. F. *Chem. Phys. Lett.* **1994,** *222* (6), 529-539.

75. Zacharias, M.; Straatsma, T. P.; McCammon, J. A. *J. Chem. Phys.* **1994,** *100* (12), 9025.

76. Chitra, R.; Smith, P. E. *J. Chem. Phys.* **2001,** *114* (1), 426-435.

77. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2002,** *106* (6), 1491-1500.

78. Weerasinghe, S.; Pettitt, B. M. *Mol. Phys.* **1994,** *82* (5), 897-912.

79. Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. *J. Phys. Chem. Lett.* **2013,** *4* (2), 235-238.

80. Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. *J. Phys. Chem. B* **2011,** *115* (37), 10911-10918.

81. Schnell, S. K.; Vlugt, T. J. H.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. *Chem. Phys. Lett.* **2011,** *504* (4–6), 199-201.

82. Ben-Naim, A. *J. Chem. Phys.* **1977,** *67* (11), 4884-4890.

83. Matteoli, E.; Lepori, L. *J. Chem. Phys.* **1984,** *80* (6), 2856-63.

84. *Physical and chemical equilibriums for chemical engineers*. John Wiley & sons: New Jersey, 2012.

85. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2001,** *105* (46), 11513-11522.

86. Chitra, R.; Smith, P. E. *J. Chem. Phys.* **2001,** *115* (12), 5521-5530.

87. Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006,** *27* (13), 1477-1485.

88. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2000,** *104* (24), 5854-5864.

89. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.

90. Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004,** *25* (11), 1400-1415.

91. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010,** *6* (2), 459-466.

92. Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (19), 6665-6670.

93. Schmid, N.; Eichenberger, A.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.; van Gunsteren, W. *Eur. Biophys. J.* **2011,** *40* (7), 843-856.

94. Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002,** *23* (5), 548-553.

95. Murzyn, K.; Bratek, M.; Pasenkiewicz-Gierula, M. *J. Phys. Chem. B* **2013,** *117* (51), 16388-16396.

96. Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010,** *290* (1–2), 43-47.

97. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *118* (13), 5901-5910.

98. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *118* (23), 10663-10670.

99. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003,** *107* (16), 3891-3898.

100. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *119* (21), 11342-11349.

101. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004,** *121* (5), 2180-2186.

102. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005,** *109* (31), 15080-15086.

103. Bentenitis, N.; Cox, N. R.; Smith, P. E. *J. Phys. Chem. B* **2009,** *113* (36), 12306-12315.

104. Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011,** *13* (40), 18154-18167.

105. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weeerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011,** *7* (5), 1369-1380.

106. Lin, B.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *J. Chem. Phys.* **2013,** *139* (8), 084509.

107. Ben-Naim, A. *Cell Biophys.* **1988,** *12* (1), 255-269.

108. Ben-Naim, A. *J. Phys. Chem.* **1989,** *93* (9), 3809-3813.

109. Ben-Naim, A. *Pure Appl. Chem.* **1990,** *62* (1), 25.

110. Grunwald, E.; Baughman, G.; Kohnstam, G. *J. Am. Chem. Soc.* **1960,** *82* (22), 5801-5811.

111. Covington, A. K.; Newman, K. E. In *Thermodynamic Behavior of Electrolytes in Mixed Solvents*, American Chemical Society: 1976; Vol. 155, pp 153-196.

112. Marcus, Y. *Aust. J. Chem.* **1983,** *36* (9), 1719-1731.

113. Marcus, Y. *J. Chem. Soc., Faraday Trans. 1* **1989,** *85* (2), 381-388.

114. Marcus, Y. *J. Chem. Soc., Faraday Trans.* **1991,** *87* (12), 1843-1849.

115. Chatterjee, P.; Laha, A. K.; Bagchi, S. *J. Chem. Soc., Faraday Trans.* **1992,** *88* (12), 1675-1678.

116. Banerjee, D.; Laha, A. K.; Bagchi, S. *J. Chem. Soc., Faraday Trans.* **1995,** *91* (4), 631-6.

117. Laha, A. K.; Banerjee, D.; Bagchi, S. *Indian J. Chem., Sect. A: Inorg., Bio-inorg., Phys., Theor. Anal. Chem.* **1995,** *34A* (5), 335-41.

118. Laha, A. K.; Das, P. K.; Banerjee, D.; Bagchi, S. *J. Chem. Soc., Faraday Trans.* **1996,** *92* (9), 1499-1502.

119. Laha, A. K.; Das, P. K.; Bagchi, S. *J. Phys. Chem. A* **2002,** *106* (13), 3230-3234.

120. Matteoli, E. *J. Phys. Chem. B* **1997,** *101* (47), 9800-9810.

121. Ben-Naim, A. *J. Phys. Chem. B* **2007,** *111* (11), 3072-3072.

122. Shulgin, I. L.; Ruckenstein, E. *J. Chem. Phys.* **2005,** *123* (5), 054909.

Chapter 2 - Are Molecular Mechanics Force Fields Improved by Using More Accurate Electrostatics Potentials?

2.1 Introduction

Nowadays, molecular dynamics simulations are widely applied to obtain a better understanding of complex biological systems.[1-15] When using molecular dynamics simulations, the precision and accuracy of the simulation results are very important.[1, 16-18] The degree of consistency of the results attained during the simulations defines the precision, which is determined by the degree of sampling.[17-19] The accuracy of the simulation results primarily depend on the force field parameters.[14-15] Therefore, in order to obtain more accurate simulation results, people are constantly attempting to improve force field parameters.[2, 5-7, 16, 19-23]

Molecular mechanics force fields use relatively simple partial atomic charges that are assigned at nuclear sites in order to represent the charge distribution of molecules.[2, 5-7, 22, 24-25] This represents a compromise between accuracy and computational efficiency.[2, 16, 22, 25] It is important, therefore, to establish whether force fields can be improved by the addition of extra charge sites used to provide an improved description of the real electron density, especially due to the increased computational effort that is then required.[7, 22, 25-26]

Hydration free energy calculations are one of the widely applied tests for force field accuracy.[18, 26-34] Most hydration free energy studies have been performed on amino acids side chains as it is closely associated with the protein folding.[18, 26, 32, 34-35] Nevertheless, recently, diverse sets of small molecules have been studied using hydration free energies.[26, 28] In particular, alchemical free energy calculations have been used to calculate the hydration free energy from molecular dynamics simulations.[18, 26, 36] This approach involves with the sampling of the molecules using molecular dynamics simulations, in water and in the gas phase, together with a series of

intermediate unphysical states (alchemical states) spanning the gas and aqueous states.[18, 26, 36] Hence, the potential energy differences between each of these alchemical states can be used to calculate hydration free energies.[26] Mobley and coworkers have performed a test of force field accuracy by calculating hydration free energies of a set of small organic compounds.[26, 28] Their test set consisted of ethane, biphenyl, and dibenzyl dioxin, as well as a series of chlorinated derivatives of each molecule.[26] Furthermore, they have investigated several charge models and determined their hydration free energies.[26] The calculated hydration free energy values were then compared with the experimental values.[26] Their standard charge model was the AM1-BCC charge model. According to their results they found that high quality partial charges from MP2/cc-PVTZ SCRF RESP fits to the quantum mechanical electron density (MP2/SCRF charge model) provided slightly better agreement with experiment compare to the standard AM1-BCC. The MP2/SCRF charge model for chlorinated ethane derivatives was also extended by including additional virtual sites designed to more accurately describe the quantum derived molecular electrostatic potentials. In particular, virtual sites were added to better represent the potential around the chlorine atoms leading to the MP2/ExpSQ-Q charge model.[26]

Generally, the calculated hydration free energies were more positive than the experimental values for all the charge models.[26] This was moderated somewhat for the MP2/ExpSQ-Q charge model (with extra charge sites). The addition of virtual sites shifts the hydration free energies of many of these compounds to be more favorable. These extra sites provide larger C-Cl bond dipoles even while the overall polarity of the molecule remains low. Even for the extra site models the free energies were still too positive on the average.[26]

However, when we calculate hydration free energies, it should be noted that this quantity does not give information about solute-solute interactions. Only solute-solvent and modified

solvent-solvent interactions are included in the hydration free energy values. Furthermore, most common force fields are developed considering only the properties of pure liquids.[6] Therefore, when these force fields are used for mixtures the accuracy is initially unknown.

Here, we examine if the addition of extra (virtual) charge sites represent a significant improvement in the known liquid state properties of mixtures containing these molecules. In particular, approaches based on the Kirkwood-Buff theory of solution mixtures have become important tests for developing and validating force field parameters.[37-44] Hence, we have determined the Kirkwood-Buff integrals of chloroethane and methanol mixtures using the AM1-BCC (without extra sites), MP2/SCRF (without extra sites) and MP2/ExpSQ-Q (with extra sites) models.

2.2 Methods

2.2.1 Kirkwood Buff Theory

The Kirkwood-Buff (KB) theory of solution, also known as Fluctuation Solution Theory, was first proposed by Kirkwood and Buff in 1951.[45] This theory is an exact theory that can be applied to any stable solution mixture.[16, 46] This theory has several important advantages such as: it can be applied to any number of components, at any concentration, and for any type of molecule.[16, 19, 45] Basically, using this theory we can relate the microscopic solution structure to macroscopic thermodynamic properties.[46] The KB inversion procedure was developed by Ben-Naim.[19, 46-47] Here, available thermodynamic properties such as partial molar volumes, chemical potential derivatives and isothermal compressabilities can be related to the microscopic distribution of molecules in solution.[16, 47]

The Kirkwood-Buff integrals (G_{ij}) are defined as,

$$G_{ij} = 4\pi \int_0^{\infty} [g_{ij}^{\mu VT}(r) - 1] r^2 \, dr \approx 4\pi \int_0^{R} [g_{ij}^{NpT}(r) - 1] r^2 dr \tag{2.1}$$

where g_{ij} is the corresponding radial distribution function (rdf).[16, 37-39, 46] The distance between center of mass of component i and j is represented by r. R is the cut off distance where $g_{ij}^{NpT}(r)$ converges to one. The integrals measure the deviation of the intermolecular distribution from a random or bulk distribution.[16, 46] Chemical potentials, partial molar volumes, and compressibilities of solution mixtures can be used to obtain the KB integrals according to following equations,[42]

$$G_{12} = RTk_T - \frac{\bar{V}_1 \bar{V}_2}{(1 + f_{22})V_m} \tag{2.2}$$

$$G_{11} = G_{12} + \frac{1}{x_1}\left(\frac{\bar{V}_2}{(1 + f_{22})} - V_m\right) \tag{2.3}$$

where R is the gas constant, x_1 is the mole fraction of component 1, $\bar{V}_1$ and $\bar{V}_2$ are the partial molar volumes of component 1 and 2, respectively. k_T is the isothermal compressibility. $V_m = V/(N_1 + N_2)$ is the molar volume, and

$$\beta\left(\frac{\partial \mu_2}{\partial ln x_2}\right)_{p,T} = 1 + \left(\frac{\partial ln f_2}{\partial ln x_2}\right)_{p,T} = 1 + f_{22} \tag{2.4}$$

with ($\beta = 1/RT$) and f_2 is the activity coefficient of the component 2 on the mole fraction scale with the pure solvents as the standard states, μ_2 is the chemical potential of component 2 .[42] Calculations of the excess molar volumes is used to obtain partial molar volumes from the experimental density data as shown in Equation (2.5),[42]

$$X_m^E = X_m - x_1 X_{m,1}^0 - x_2 X_{m,2}^0 \tag{2.5}$$

where X is the volume (V) of the solution and $V_{m,1}^0$ is the molar volume of pure component 1. The excess volume and excess molar Gibbs free energy values were fitted to Redlich-Kister equation or Wilson equation.[48-50] The Redlich-Kister fitting equation is shown in Equation (2.6),[49]

$$X_m^E = x_1 x_2 \sum_{i=0}^{n} a_i \, (x_1 - x_2)^i \tag{2.6}$$

where a_i are fitting constants, x_1 and x_2 are the mole fractions, and X is either the volume or Gibbs free energy. The Wilson fitting equation is shown in Equation (2.7),[48]

$$g = \frac{G^E}{RT} = -x_1 \ln(x_1 + \Lambda_{12} x_2) - x_2 \ln(x_2 + \Lambda_{21} x_1) \tag{2.7}$$

where G^E is the excess Gibbs free energy, $\Lambda_{12}, \Lambda_{21}$ are the Wilson fitting constant. The partial molar quantities at any composition can be calculated using the standard relationship as shown in Equation (2.8),

$$Y_1 = X_m^E - x_2 \left(\frac{\partial X_m^E}{\partial x_2} \right)_{p,T} \tag{2.8}$$

where X = V or βG and Y is corresponding to partial molar volume or excess chemical potential of component 1, respectively.

Normally, the KB integrals are not sensitive to the values of the isothermal compressibility and for that reason the following equation was used to obtain the isothermal compressibility,[51]

$$k_T = \phi_1 k_{T,1}^0 + \phi_2 k_{T,2}^0 \tag{2.9}$$

where $\phi_1 = \rho_1 \bar{V}_1$ is the volume fraction of the component 1 in the solution. Isothermal compressibilities for the pure components ($k_{T,1}^0$) were obtained from the literature.[52-54]

2.2.2 Molecular Dynamics Simulations

All mixtures were simulated using classical molecular dynamics techniques and the GROMACS package (version 4.6).[55] The simulations were performed in the isothermal isobaric (NpT) ensemble at 323 K and 1 bar. The weak coupling techniques was used to modulate the temperature and pressure with relaxation times of 0.1 and 5 ps.[26] Van der Waals interactions were steadily switched off between 0.9 and 1.0 nm.[26] The particle mesh Ewald technique was used to calculate the electrostatics interactions.[56] The real space cut off value was 1.2 nm.[26] Bonds were constrained using the LINCS algorithm.[26] An integration time step of 2 fs was used. Random initial configurations were generated in a 10 nm cubic simulation box. The steepest decent method was used for energy minimization. This was followed by several equilibration runs and then performed a production run of 10 ns.

Mobley and coworkers have studied several chloroethane molecules with different charge models.[26] Their approach for calculating hydration free energies used molecules in the gas phase and in the aqueous phase, and also at several intermediate states (alchemical states) spanning between gas phase and aqueous phase.[26, 28] Their general protocol is well described in their previous studies.[26, 28] We have selected three charge models from their study: AM1-BCC, MP2/SCRF, MP2/ExpSQ-Q.[26] Their standard charge model was AM1-BCC charge model.[26] In addition to this standard charge model, they have also used a MP2/SCRF charge model. This model is based on MP2/cc-pVTZ calculations with a self consistent reaction field (SCRF) continuum electrostatic model to represent the solvent.[26] Both of these charge models contain atom centered partial atomic charges. The MP2/ExpSQ-Q charge model was an extension of the MP2/SCRF charge model with additional virtual charge sites attached to each chlorine atom. The position of the virtual charge site is along the carbon-chlorine bond axis at a distance 30% of the carbon

chlorine bond length away from the chlorine atom.[26] In this study, the AM1-BCC, MP2/SCRF and MP2/ExpSQ-Q charge models will be referred to AM1, MP2, MP2-Q respectively.

In this study we have analyzed chloroethane and methanol mixtures. The main reason that we have selected chloroethane was the presence of available experimental thermodynamic data required to obtain the experimental KB integrals of chloroethane and methanol mixtures. We have studied 1,1 dichloroethane, 1,2 dichloroethane, 1,1,1 trichloroethane and 1,1,2,2 tetrachloroethane molecules with methanol using the models developed by Mobley and coworkers based on the AMBER force field. These chloroethane and methanol mixtures were studied at 323 K. The AMBER methanol model was used in our simulations.[57]

In addition to the KB integrals, several other solution properties were examined. The self diffusion coefficients were computed using the mean square fluctuation approach.[58] Furthermore, we have calculated the relative permittivities of these mixtures from the dipole moment fluctuations.[59] The enthalpy of mixing is computed from Equation (2.10),

$$\Delta H_m = H_{sol} - x_1 H_1 - x_2 H_2 \tag{2.10}$$

where ΔH_m is the enthalpy of mixing, H_{sol} is the molar enthalpy of the solution, $H_1$and H_2 are the molar enthalpy of the pure components 1 and 2, respectively.

2.3 Results

Partial atomic charges of the three charge models and simulated dipole moments obtained from the pure liquids are displayed in Table 2.1. The dipole moment values are not significantly different. In particular, after they include additional charge sites (MP2-Q charge model) this results in a comparatively higher negative partial atomic charge for Cl.

Table 2.1 Partial atomic charges of the three charge models and simulated dipole moment values of chloroethane molecules

Molecule	C	H	Cl	E	μ (D)
1,1 dichloroethane					
AM1	-0.1124/0.1563*	0.0622/0.0939*	-0.1622		2.40
MP2	-0.2739/-0.0701*	0.1273/0.2289*	-0.1335		2.83
MP2-Q	-0.4999/0.5836*	0.1536/0.0504*	-0.5814	0.2839	2.61
1,2 dichloroethane					
AM1	0.0201	0.0761	-0.1722		2.37
MP2	-0.1247	0.1590	-0.1934		2.85
MP2-Q	0.0401	0.1053	-0.4159	0.1651	2.83
1,1,1 trichloroethane					
AM1	-0.1120/0.2640*	0.0730	-0.1236		2.26
MP2	-0.4914/-0.2683*	0.2247	0.0285		2.38
MP2-Q	-0.6434/0.9553*	0.1926	-0.6638	0.3671	2.17
1,1,2,2 tetrachloroethane					
AM1	0.1226	0.1135	-0.1180		1.71
MP2	-0.3949	0.3700	0.0125		1.00
MP2-Q	0.2145	0.1476	-0.4226	0.2416	1.52

For asymmetric molecules, charges on the Cl side are denoted using a *. In these simulations the number of exclusions is equal to 3.

Computational cost is another important factor that we have to consider with force fields. An increase in the number of charge sites is expected to result in an increase in computational cost. Hence, the number of nanoseconds per day was calculated for simulation of the pure chloroethane compounds. The timings suggest that the MP2-Q charge model was computationally more

Table 2.2 Simulated and experimental properties of the pure liquid chloroethanes. All simulation and experimental data correspond to 323 K unless otherwise noted. ρ (g/cm^{-3}), D (10^{-5} cm^2/s), E_{pot} (kJ/mol), $E_{pot}^{(inter)}$ (kJ/mol),ε indicate the density, diffusion coefficient, potential energy per molecule , intermolecular interaction energy per molecule and relative permittivity, respectively.

System	Property	Expt.	Molecular dynamics		
			AM1	MP2	MP2-Q
1,1 dichloroethane	ρ	1.18[60]	1.11	1.14	1.14
	D		4.01	3.34	3.47
	E_{pot}		-18.88	-7.20	-967.02
	$E_{pot}^{(inter)}$		-26.09	-27.65	-28.65
	ε		6	10	10
1,2 dichloroethane	ρ	1.20[50]	1.15	1.19	1.19
	D (313 K)	2.11[61]	3.04	2.39	2.61
	D		3.39	2.37	2.53
	E_{pot}		-6.60	-2.58	-322.96
	$E_{pot}^{(inter)}$		-28.78	-34.77	-33.61
	ε		9	15	12
1,1,1 trichloroethane	ρ	1.29[50]	1.32	1.32	1.30
	D		2.07	2.02	2.34
	D (303 K)	1.58[62]	1.52	2.01	2.34
	E_{pot}		-34.91	15.82	-2200.17
	$E_{pot}^{(inter)}$		-31.26	-31.89	-31.11
	ε		5	5	5
1,1,2,2 tetrachloroethane	ρ	1.54[50]	1.53	1.55	1.53
	D		1.16	1.08	1.19
	E_{pot}		-11.95	46.55	-925.35
	$E_{pot}^{(inter)}$		-39.99	-40.79	-40.77
	ε		4	2	4

expensive (about 10% for mixtures and 20% for pure systems) compared to the AM1 and MP2 charge models. However, this does not represent a prohibitively higher computational cost.

The properties of the pure chloroethane liquids are listed in Table 2.2 and compare to experimental values where possible. These results do not show a clear improvement in the simulated density values after the inclusion of extra charge sites on chlorine atoms. In fact, all of these charge models predict somewhat similar results for the density.

We could not find experimental self diffusion coefficients at 323 K. The experimental self diffusion coefficients for pure 1,2 dichloroethane and 1,1,1 trichloroethane molecules at 313 K and 303 K, respectively are available.[61-62] Therefore, we have performed several additional MD simulations for these two systems. The self diffusion coefficients predicted using the MP2 charge model is closer to the experimental value than the other two charge models. The MP2 and MP2-Q charge models result in very similar self diffusion coefficients.

The calculated potential energy per molecule values show a very large negative value for the MP2-Q charge model compared to the other two charge models. It is due to the higher number of interactions within the chloroethane molecules when using the extra virtual charge sites. Although we use exclusions of three bonds when performing the molecular dynamics simulations, the extra virtual charge sites are not included in the exclusion rules. Hence, we have calculated potential energy values corresponding to intermolecular interactions, $E_{pot}^{(inter)}$ by removing all the intramolecular contribution to the total potential energy. Thus, we observe all three charge models result in very similar intermolecular interaction energies.

The three charge models predict very similar values for the above compared properties. Therefore, it is difficult to distinguish which charge model provides the best representation of the liquid structure based on these properties.

The relative free energy for rotation around the central dihedral angles were computed in order to explore whether the extra charge sites effect the conformational preferences of the chloroethane molecules. The relative free energy (W) can be determined from the probability distribution (P), using $W = -RT \ln P$.

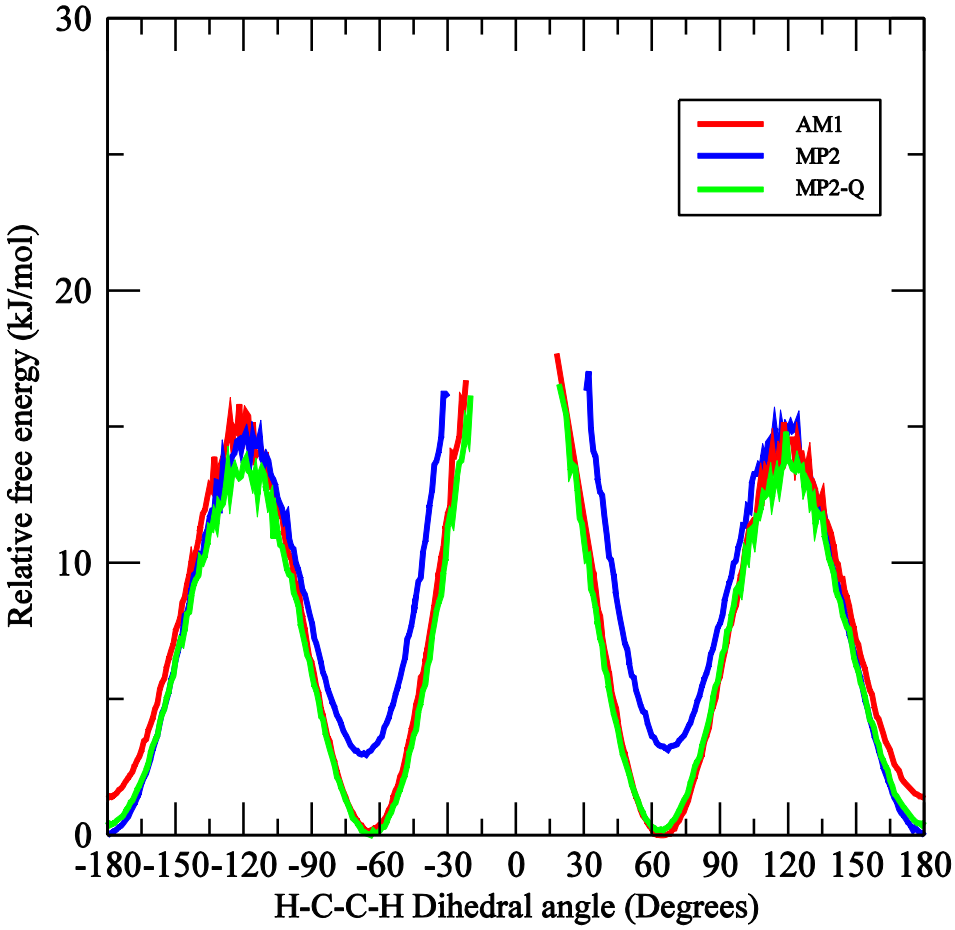

Figure 2.1 Relative free energy of rotation around H-C-C-H dihedral angle of 1,1,2,2 tetrachloroethane molecule for x_1=0.4 simulation. Simulation data correspond to 323 K

We have selected a solution mixture contains 1,1,2,2 chloroethane as this molecule contains four extra charge sites. The relative free energy for rotation around the H-C-C-H dihedral angle of 1,1,2,2 tetrachloroethane molecule is shown in Figure 2.1. The MP2 charge model results in a comparatively high free energy for the gauche (g+ and g-) conformations. This can be due to the

higher positive charge on the H atom in MP2 charge model. Hence, g+ and g- conformations are less stable. However, the AM1 and MP2-Q charge distributions result in very similar free energies concluding that the extra charge sites do not display a significant effect on the conformational changes.

Furthermore, a conformational analysis of 1,2 dichloroethane and 1,1,2,2 tetrachloroethane molecules was performed and the results are given in Table 2.3. The results correspond to mixtures of 0.4 mole fraction chloroethane and pure system. The same behavior of the conformational distribution can be observed with the other compositions as well. We have studied Cl-C-C-Cl and H-C-C-H dihedral angles of 1,2 dichloroethane and 1,1,2,2 tetrachloroethane molecules, respectively. In the 1,2 dichloroethane system we observe a similar conformational distribution when comparing all three charge models. In the 1,1,2,2 tetrachloroethane system we see a relatively low probability for g+ and g- conformations in the MP2 charge model. Therefore, the most probable conformation is the trans conformation. This can be due to the higher positive charge on H atoms in the 1,1,2,2 tetrachloroethane molecule as assigned in the MP2 charge model In this case, the trans conformation is more stable than the g- and g+ conformations.

However, the trans conformation ought to be the most stable conformation conformation as it gives the minimum steric repulsion. According to the results, we observe the opposite behavior. This can be due to solvation effects. It might be possible to observe the most probable trans conformations in gas phase, where the intermolecular interaction are very low. But in the condensed phase intermolecular interaction play a major role in conformational equilibria.

Table 2.3 A conformational analysis of Cl-C-C-Cl and H-C-C-H dihedral angles at 323 K

Chloroethane Mole fraction	Charge model	1,2 dichloroethane Cl-C-C-Cl		1,1,2,2 tetrachloroethane H-C-C-H	
		g-/g+	Trans	g-/g+	trans
0.4	AM1	0.791	0.209	0.649	0.350
	MP2	0.847	0.152	0.359	0.641
	MP2-Q	0.791	0.209	0.663	0.336
1.0	AM1	0.789	0.210	0.706	0.294
	MP2	0.710	0.289	0.211	0.789
	MP2-Q	0.781	0.220	0.539	0.461

The methanol-methanol Kirkwood-Buff integrals are displayed in Figure 2.2. When we increase the chloroethane mole fraction we observe a higher aggregation of methanol molecules as indicated by a higher positive value for the KB integrals. For the 1,1,2,2 tetrachloroethane system we did not find the experimental thermodynamic data required to obtain the experimental KB integrals. However, we have included the results obtained for 1,1,2,2 tetrachloroethane-methanol system, as we can observe similar trends compare to the other three systems. The AM1 charge model reproduced the experimental KB integrals more closely for the 1,1 dichloroethane and 1,2 dichloroethane systems. In contrast, the MP2-Q charge model is best for the 1,1,1 trichloroethane system. However, all three charge models overestimate the methanol aggregation for the 1,1,1 trichloroethane system with an increase in chloroethane mole fraction. Clearly, there is no significant improvement in the simulated KB integrals with the MP2-Q charge model.

The methanol-chloroethane KB integrals are displayed in Figure 2.3. Again there is no notable difference in the KB integrals between the MP2 and MP2-Q charge models except for the 1,1,1 trichloroethane-methanol system.

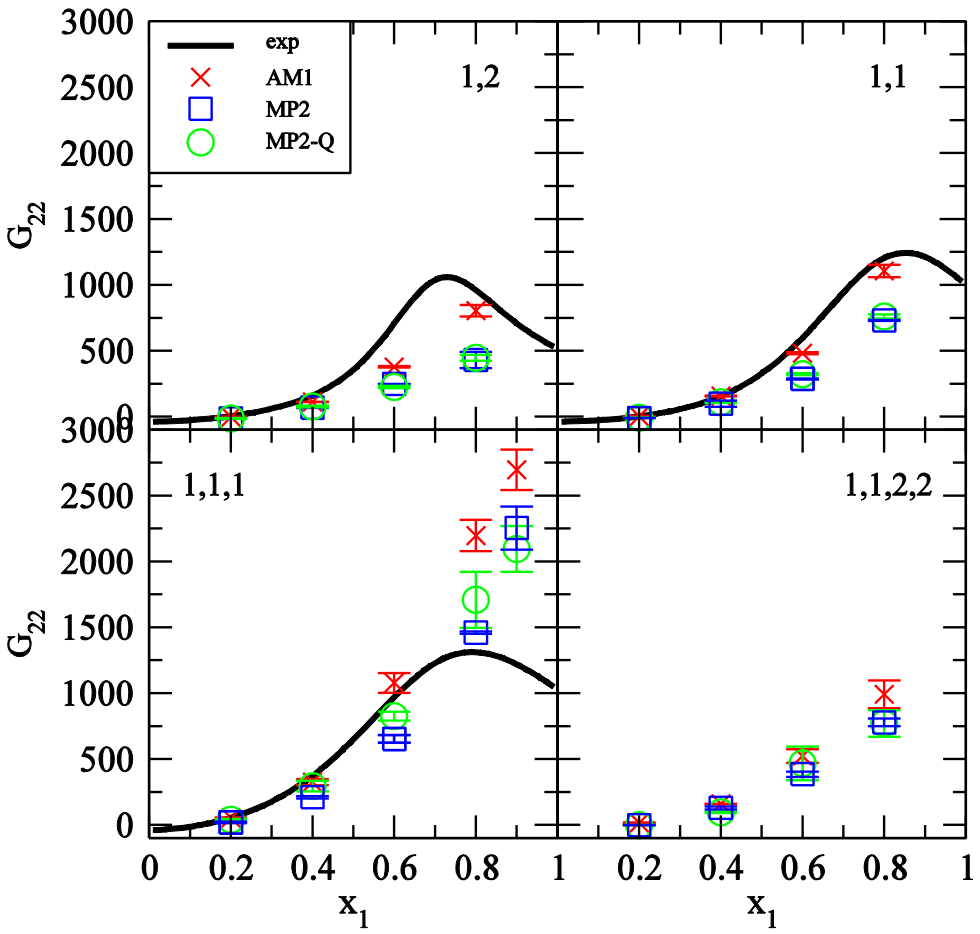

Figure 2.2 Methanol-methanol KB integrals (G_{22}, cm^3/mol) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). The solid lines correspond to the experimental data. Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

The chloroethane - chloroethane KB integrals are shown in Figure 2.4. Here, also we do not observe a significant improvement in the simulation results after the inclusion of additional charge sites on chlorine atoms.

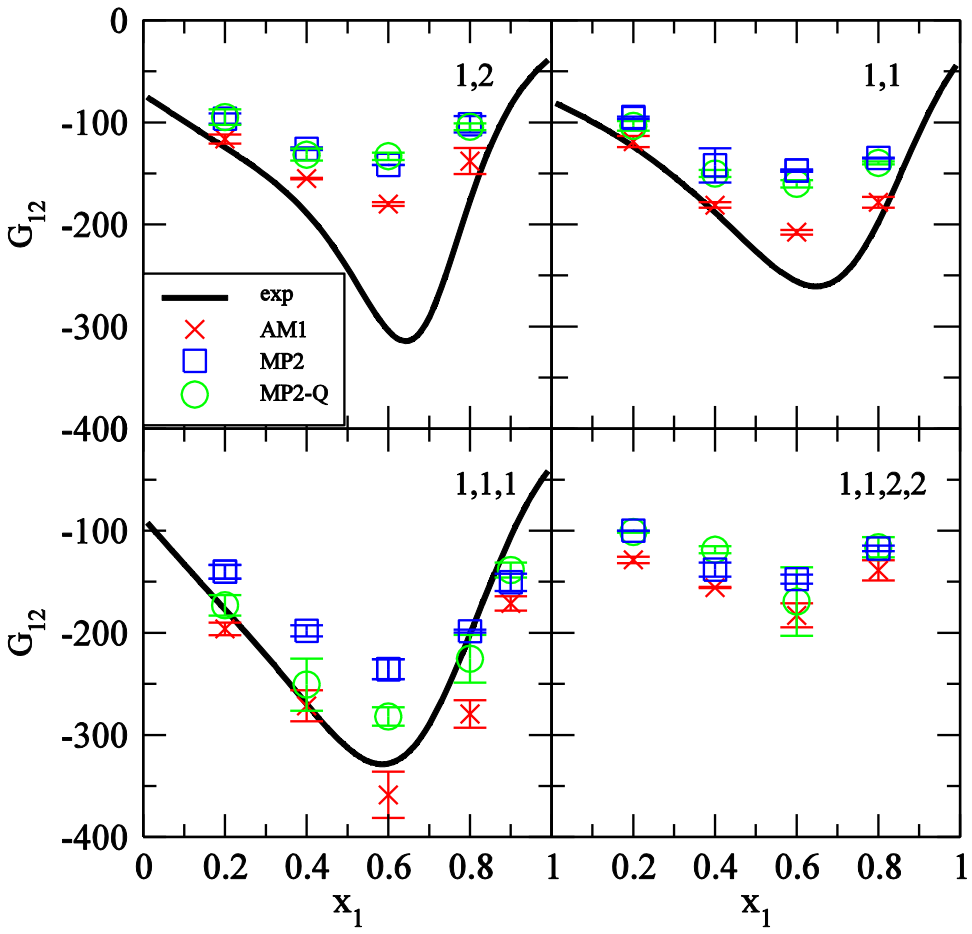

Figure 2.3 Chloroethane-methanol KB integrals (G_{12}, cm^3/mol) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). The solid lines correspond to the experimental data. Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

The simulated KB integrals provide information concerning the variation in the relative distribution of molecules over all solvation shells. It is interesting to determine the contribution from the first solvation shell to the full KB integrals. Most importantly, we wanted to investigate whether, after including extra charge site on the Cl atoms, we can observe a significant difference in the local arrangement of the molecules. Figure 2.5 shows the first shell KB integrals compared to the full KB integral. We can see that there is a correlation between the first shell KB integrals

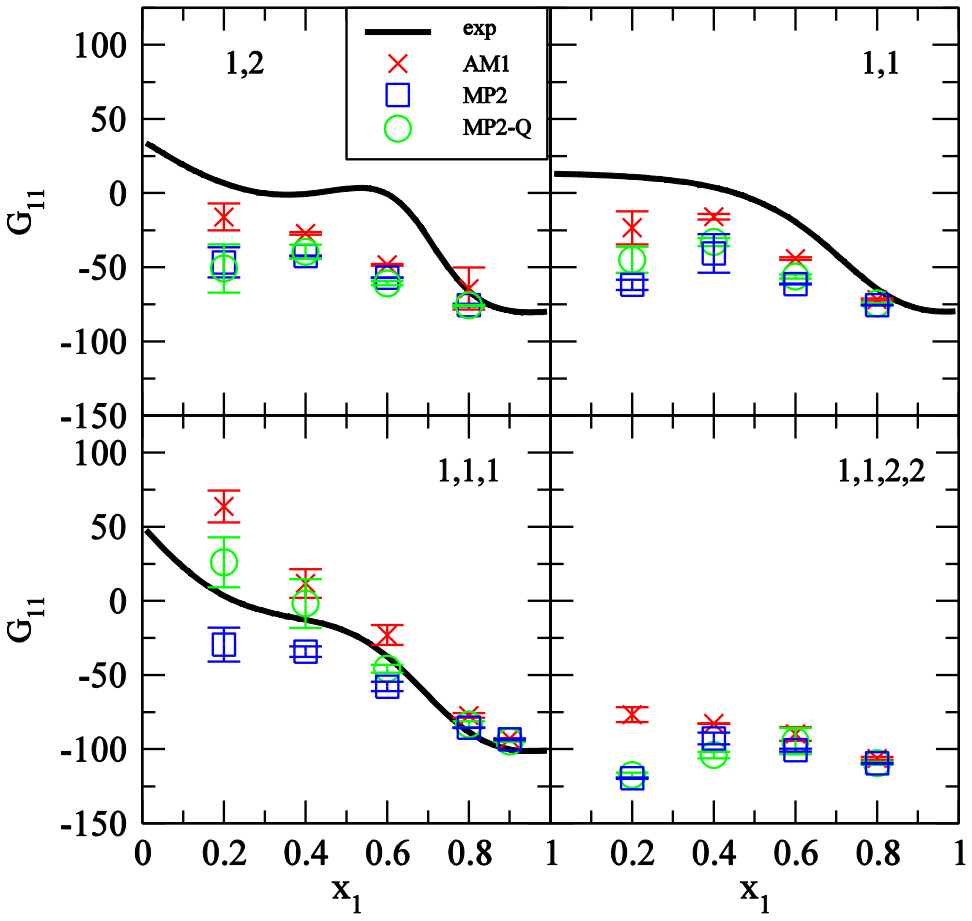

Figure 2.4 Chloroethane-chloroethane KB integrals (G_{11}, cm^3/mol) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). The solid lines correspond to the experimental data. Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

and the full KB integrals. However, we do not observe a significant change in the data when using additional charge sites. Furthermore, the relationship between long range and short range KB integrals is not 1:1.

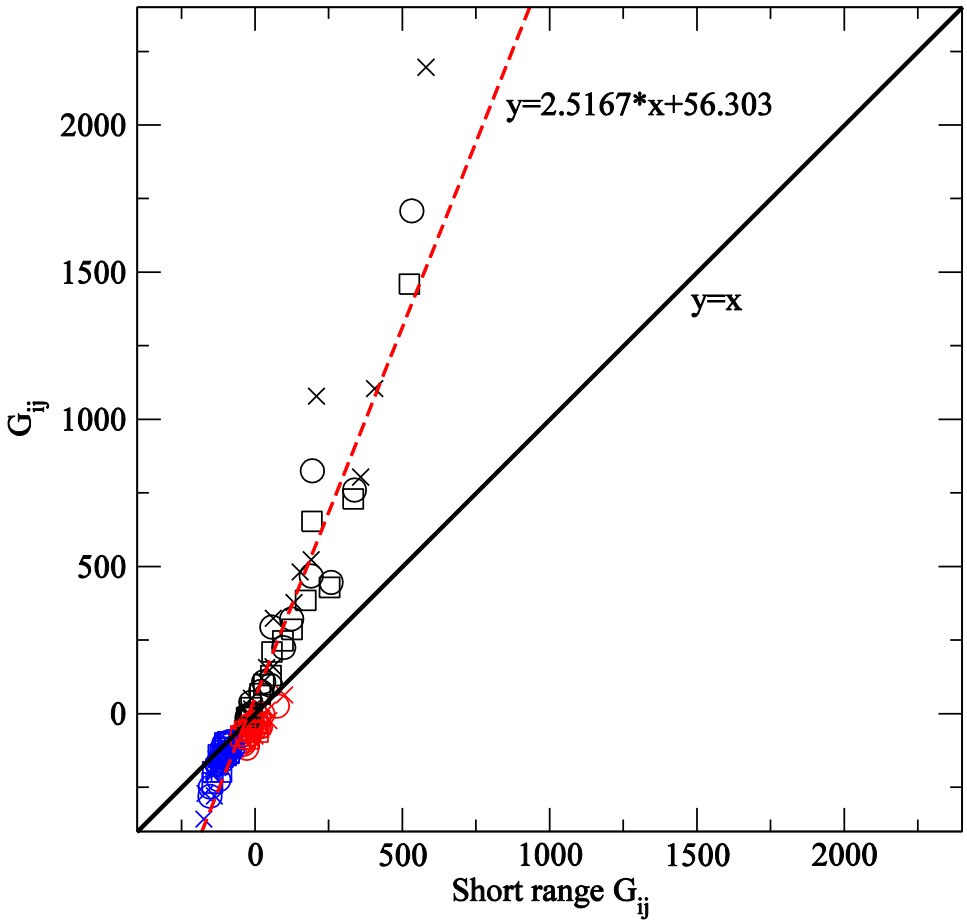

Figure 2.5 The correlation between simulated KB integrals and the contribution from the first shell. Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively. Black, red and blue symbols correspond to methanol-mathanol, chloroethane-chloroethane, chloroethane-methanol KB integrals respectively for all the compositions. Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

The calculated relative permittivities of these mixtures are displayed in Figure 2.6. We can see that the AM1 charge model displayed different results for the 1,2 dichloroethane-methanol and 1,1 dichloroethane-methanol systems than the MP2 charge models. However, we do not observe a significant difference for the MP2 and MP2-Q charge model. Unfortunately, we do not have experimental relative permittivity values to compare.

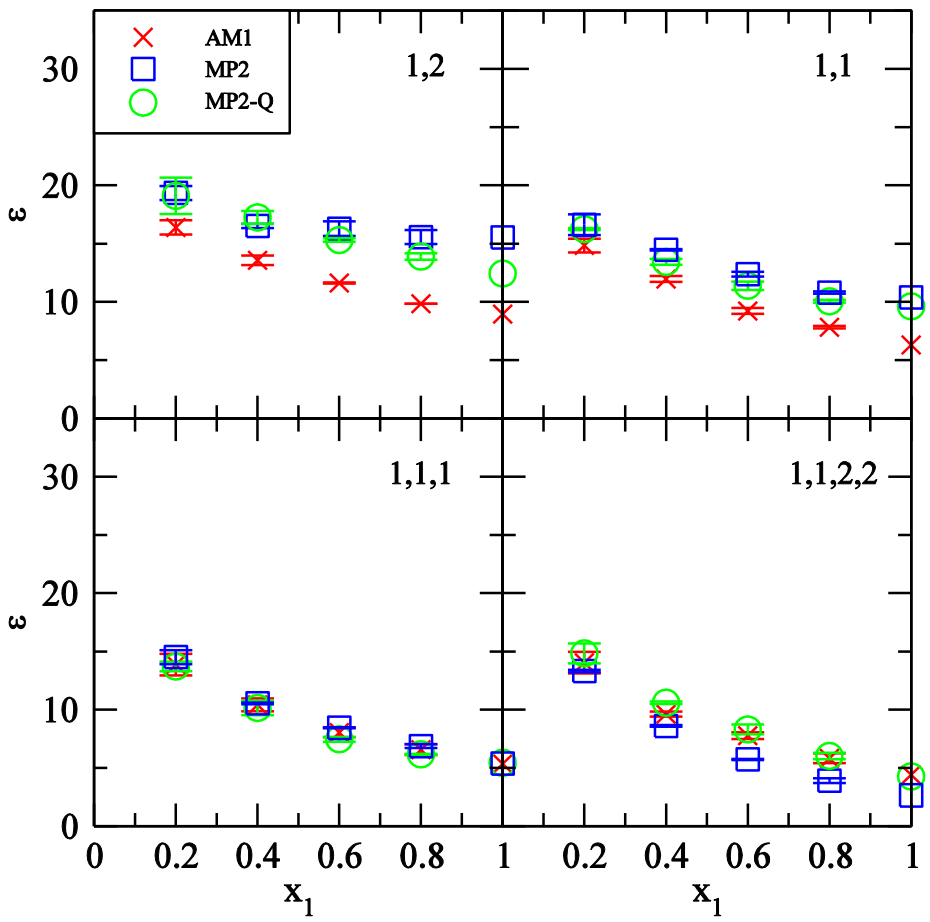

Figure 2.6 Relative permittivity (ε) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

The self diffusion coefficients for the chloroethane and methanol molecules are displayed in Figure 2.7 and Figure 2.8, respectively. Unfortunately, the corresponding experimental values are unavailable. We do not observe significant difference in chloroethane self diffusion coefficients for the three charge models in the 1,1,1 trichloroethane and 1,1,2,2 tetrachloroethane systems.

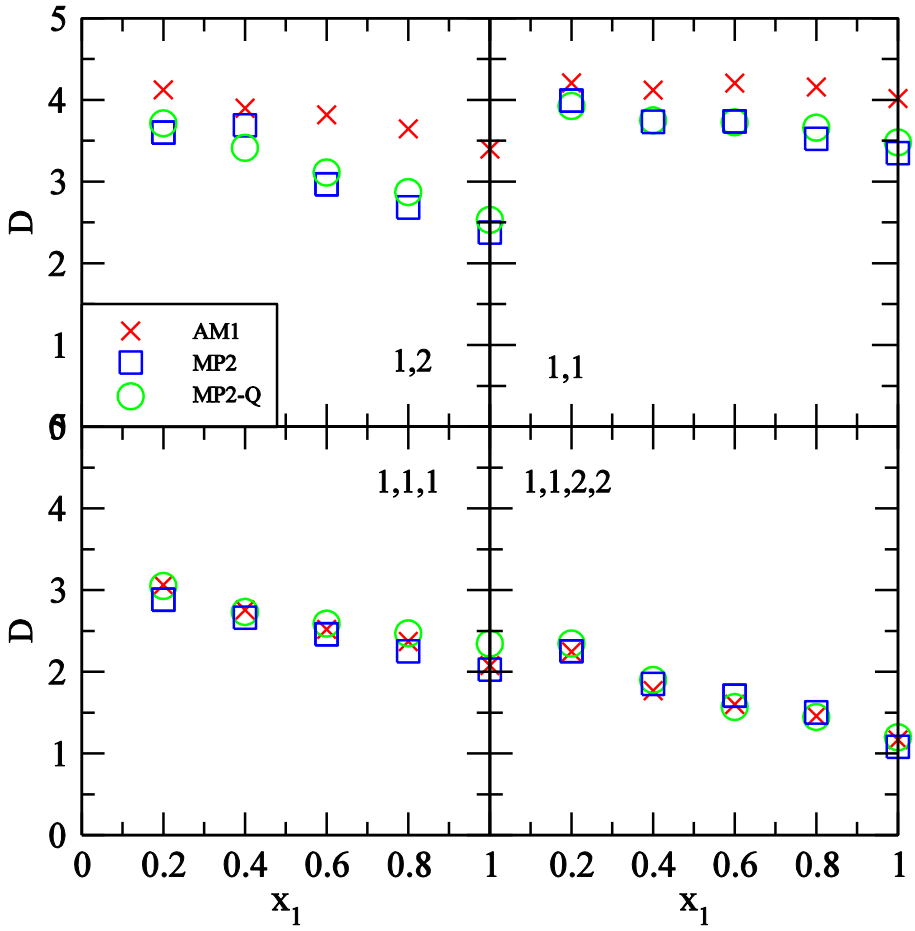

Figure 2.7 Chloroethane self diffusion coefficient (D, 10^{-5} cm^2/s) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

However, the AM1 charge model result in a higher chloroethane self diffusion coefficient for the 1,2 dichloroethane and 1,1 dichloroethane systems. Interestingly, for all of these systems we do not see any significant difference in chloroethane self diffusion coefficients between the MP2 and MP2-Q charge models. Furthermore, all three models predict very similar methanol self diffusion coefficients.

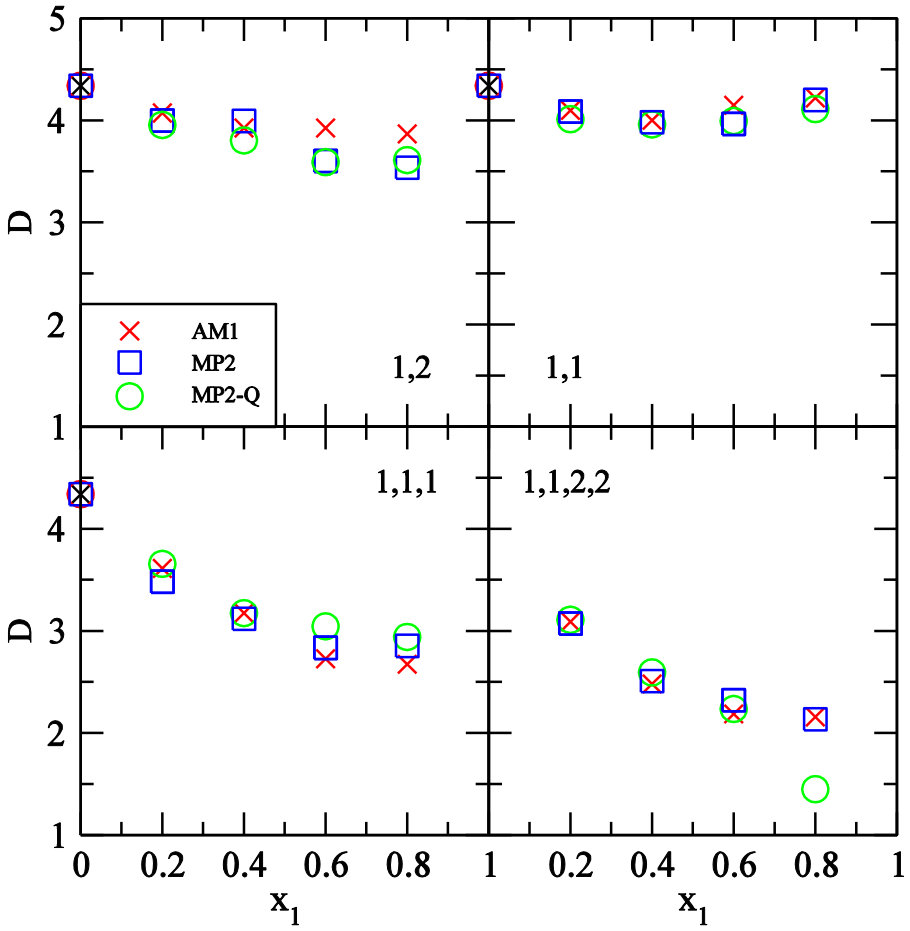

Figure 2.8 Methanol self diffusion coefficient (D, 10^{-5} cm^2/s) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

The enthalpy of mixing values for these mixtures are displayed in Figure 2.9. Generally, the enthalpy of mixing values are sensitive to the charge distribution.[63] However, with these chloroethane-methanol mixtures we cannot observe a significantly different enthalpy of mixing value with the different charge models. In 1,2 dichloroethane system we can see that the AM1 charge model produces a comparatively higher enthalpy of mixing value compared to the other two models. Unfortunately, the experimental enthalpy of mixing values are also not available.

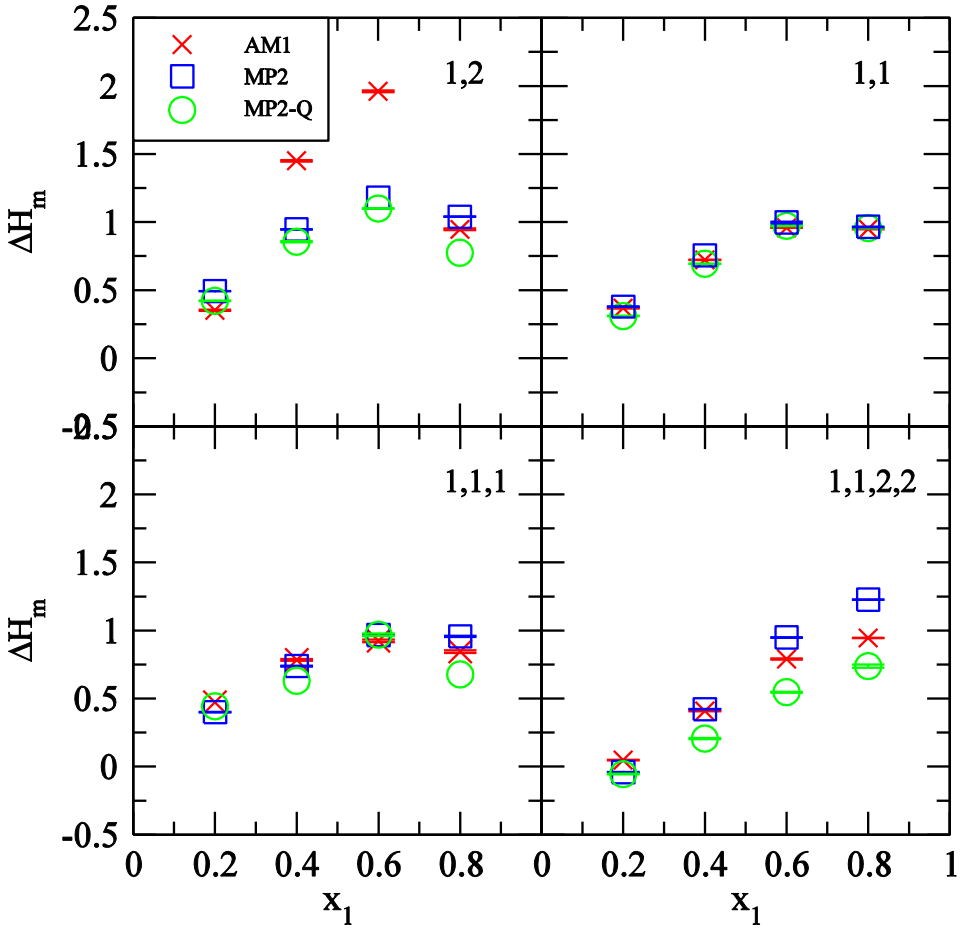

Figure 2.9 Simulated enthalpy of mixing values (ΔH_m, kJ/mol) for 1,2 dichloroethane (top left), 1,1 dichloroethane (top right), 1,1,1 trichloroethane (bottom left), 1,1,2,2 tetrachloroethane (bottom right) systems as a function of chloroethane mole fraction (x_1). Crosses, squares and circles correspond to simulation data of AM1, MP2, MP2-Q charge models respectively.

2.4 Conclusions

The results indicate a significantly higher degree of extra association between chloroethane and methanol molecules when using the MP2 charge models (with and without extra sites).The comparison of experimental and simulation KB integrals from the MP2 charge models shows a slight improvement in the simulation results when including the virtual sites. However, the AM1

charge model reproduces the experimental KB integrals more accurately than the MP2 charge models. It is difficult to draw clear conclusions for the other calculated properties of the three charge models. This can be due to the low polarity of the chloroethane molecules. Therefore, it may not be possible to observe significant change in the observed properties with the inclusion of extra charge sites. However, using extra charge sites, derived from QM calculations, clearly does not necessarily provide more accurate condensed phase properties for liquid mixtures.

2.5 References

1. Karplus, M.; McCammon, J. A. *Nat. Struct. Mol. Biol.* **2002,** *9* (9), 646-652.

2. van der Kamp, M. W.; Shaw, K. E.; Woods, C. J.; Mulholland, A. J. *J. R. Soc. Interface.* **2008,** *5* (Suppl 3), 173-190.

3. Lazaridis, T. *Curr. Org. Chem.* **2002,** *6* (14), 1319-1332.

4. Lazaridis, T.; Masunov, A.; Gandolfo, F. *Proteins: Structure, Function, and Bioinformatics* **2002,** *47* (2), 194-208.

5. van Gunsteren, W. F.; Dolenc, J. *Biomolecular simulation: historical picture and future perspectives*. 2008; Vol. 36, p 11-15.

6. van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glättli, A.; Hünenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F. A.; Yu, H. B. *Angew. Chem. Int. Ed.* **2006,** *45* (25), 4064-4092.

7. Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. *J. Chem. Theory Comput.* **2010,** *6* (2), 459-466.

8. Bonnet, P.; Bryce, R. A. *Protein Sci.* **2004,** *13* (4), 946-957.

9. Durrant, J. D.; Keranen, H.; Wilson, B. A.; McCammon, J. A. *PLoS Negl Trop Dis* **2010,** *4*.

10. Durrant, J. D.; Urbaniak, M. D.; Ferguson, M. A.; McCammon, J. A. *J. Med. Chem.* **2010,** *53*.

11. Durrant, J. D.; McCammon, J. A. *BMC Biol.* **2011,** *9* (1), 1-9.

12. Durrant, J. D.; Cao, R.; Gorfe, A. A.; Zhu, W.; Li, J.; Sankovsky, A.; Oldfield, E.; McCammon, J. A. *Chem Biol Drug Des* **2011,** *78*.

13. Kandt, C.; Ash, W. L.; Peter Tieleman, D. *Methods* **2007,** *41* (4), 475-488.

14. Meller, J.; Petrenko, R. Molecular Dynamics. John Wiley & Sons, Ltd: 2001.

15. Becker, O. M. M., MacKerell, A. D. J.; Roux, B.; Watanabe, M. *Computational Biochemistry and Biophysics* Marcel Dekker, inc.: New York, 2001.

16. Smith, P. E.; Matteoli, E.; O'Connell, J. P. *Fluctuation Theory of solutions*. CRC Press: Boca Raton, 2013.

17. van Gunsteren, W. F.; Mark, A. E. *J. Chem. Phys.* **1998,** *108* (15), 6109-6116.

18. Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003,** *119* (11), 5740-5761.

19. Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. In *Developing force fields from the microscopic structure of solutions: the Kirkwood-Buff approach*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 55-76.

20. Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. *Chem. Rev.* **2016**.

21. 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*.

22. Cisneros, G. A.; Karttunen, M.; Ren, P.; Sagui, C. *Chem. Rev.* **2014,** *114* (1), 779-814.

23. Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010,** *290* (1–2), 43-47.

24. Ren, P.; Chun, J.; Thomas, D. G.; Schnieders, M. J.; Marucho, M.; Zhang, J.; Baker, N. A. *Q. Rev. Biophys.* **2012,** *45* (4), 427-491.

25. Baucom, J.; Transue, T.; Fuentes-Cabrera, M.; Krahn, J. M.; Darden, T. A.; Sagui, C. *J. Chem. Phys.* **2004,** *121* (14), 6998-7008.

26. Mobley, D.; Liu, S.; Cerutti, D.; Swope, W.; Rice, J. *J. Comput. Aided Mol. Des.* **2012,** *26* (5), 551-562.

27. Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2011,** *115* (5), 1329-1332.

28. Mobley, D. L.; Dumont, É.; Chodera, J. D.; Dill, K. A. *J. Phys. Chem. B* **2007,** *111* (9), 2242-2254.

29. Nicholls, A.; Mobley, D. L.; Guthrie, J. P.; Chodera, J. D.; Bayly, C. I.; Cooper, M. D.; Pande, V. S. *J. Med. Chem.* **2008,** *51* (4), 769-779.

30. Mobley, D. L.; Bayly, C. I.; Cooper, M. D.; Shirts, M. R.; Dill, K. A. *J. Chem. Theory Comput.* **2009,** *5* (2), 350-358.

31. Klimovich, P.; Mobley, D. *J Comput Aided Mol Des* **2010,** *24* (4), 307-316.

32. Villa, A.; Mark, A. E. *J. Comput. Chem.* **2002,** *23* (5), 548-553.

33. Fennell, C. J.; Wymer, K. L.; Mobley, D. L. *J. Phys. Chem. B* **2014,** *118* (24), 6438-6446.

34. Xu, Z.; Luo, H. H.; Tieleman, D. P. *J. Comput. Chem.* **2007,** *28* (3), 689-697.

35. Chang, J.; Lenhoff, A. M.; Sandler, S. I. *J. Phys. Chem. B* **2007,** *111* (8), 2098-2106.

36. Mobley, D. L.; Klimovich, P. V. *J. Chem. Phys.* **2012,** *137* (23), 230901.

37. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *118* (23), 10663-10670.

38. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003,** *107* (16), 3891-3898.

39. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *119* (21), 11342-11349.

40. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004,** *121* (5), 2180-2186.

41. Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006,** *27* (13), 1477-1485.

42. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005,** *109* (31), 15080-15086.

43. Bentenitis, N.; Cox, N. R.; Smith, P. E. *J. Phys. Chem. B* **2009,** *113* (36), 12306-12315.

44. Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011,** *13* (40), 18154-18167.

45. Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys* **1951,** *19* (6), 774-777.

46. Ben-Naim, A. *Molecular theory of solutions*. Oxford university press, 2006.

47. Ben-Naim, A. *J. Chem. Phys.* **1977,** *67* (11), 4884-4890.

48. Srinivas, C.; Venkateshwara Rao, M.; Prasad, D. H. L. *Fluid Phase Equilib.* **1991,** *61* (3), 285-297.

49. Chaudhari, S. K.; Katti, S. S. *Fluid Phase Equilib.* **1989,** *50* (3), 329-338.

50. Prakash, D. J.; Lakshmi, D. S.; Rao, M. V.; Prasad, D. H. L. *Phys. Chem. Liq.* **1996,** *33* (4), 249-254.

51. Matteoli, E.; Lepori, L. *J. Chem. Phys.* **1984,** *80* (6), 2856-63.

52. Findenegg, G. H.; Kohler, F. *Trans. Faraday Society* **1967,** *63* (0), 870-878.

53. Chandrasekhar, A. S., Kn ; Krishnaiah, A. *ACH Models Chem.* **1990,** *127* (3), 345-351.

54. Easteal, A. J.; Woolf, L. A. *J. Chem. Thermodyn.* **1985,** *17* (1), 49-62.

55. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013,** *29* (7), 845-54.

56. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

57. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins: Structure, Function, and Bioinformatics* **2006,** *65* (3), 712-725.

58. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2000,** *104* (24), 5854-5864.

59. Allen, M.P; Tildesley D. J. *Computer Simulations of Liquids*. Oxford University Press: 1987.

60. Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. *J. Chem. Theory Comput.* **2012,** *8* (1), 61-74.

61. Malhotra, R.; Price, W. E.; Woolf, L. A.; Easteal, A. J. *Int. J. Thermophys.* **1990,** *11* (5), 835-61.

62. Grochulski, T.; Pszczółkowski, L.; Kempka, M. *Phys. Rev. Lett.* **1992,** *68* (24), 3635-3637.

63. Chitra, R.; Smith, P. E. *J. Chem. Phys.* **2001,** *115* (12), 5521-5530.

Chapter 3 - Kirkwood Buff Derived Force Fields for Esters

3.1 Introduction

Recently, we have been using Kirkwood Buff (KB) theory to develop new force fields to perform more accurate biomolecular simulations.[1-10] We have developed force field parameters for a set of small organic molecules that can be used as small molecule analogues for large macromolecules. As a consequence, we have determined all the force field parameters required for peptides and proteins. In order to perform molecular dynamics simulations of lipid membrane systems, we have decided to develop force field parameters for lipids. Lipid bilayers are very important components in cells.[11-16] They mainly act as barriers to maintain the balance of molecules inside and outside the cells.[12, 17-18] Protein lipid membrane interactions are also very important.[12, 14-15, 19-24] Using our new force field we expect to obtain more reliable simulation data. As a first step to obtain force field parameters for lipid molecules we have partitioned these molecules in to four groups; the phosphate group, the glycerol group, hydrocarbon chain and an ester group. Then we have attempted to obtain force field parameters for these small molecule analogues. In this work, we present the force field parameters for the ester linkage.

There are several force fields of esters that have been developed in past few years. Kamath and coworkers have developed the TraPPE-UA force field for esters.[25] They have applied the TraPPE-UA force field for the determination of vapor liquid equilibria of carboxylate esters.[25] In this work they have calculated vapor-liquid coexistence curves, vapor pressures, boiling points and critical points of methyl acetate, ethyl acetate, methyl propionate and vinyl acetate. Furthermore, pressure-composition diagrams have been calculated for methyl acetate + ethyl acetate at 313.15 K and methyl acetate + methanol at 323.15 K. The main objective of this TraPPE force field is transferability. This means that the parameters for a particular functional group do

not change from molecule to molecule. In their work the Lennard-Jones parameters were taken from the TraPPE-UA force field for alkanes,[26] alcohols,[27] ethers[28] and carboxylic acids[25], and they have used partial charges obtained from the OPLS-UA[29] force field. They have mainly developed this force field to determine vapor-liquid equilibria of carboxylate esters.[25] There are a lot of industrial applications of esters such as lubricants, plasticizers, agricultural chemicals, plastic production, etc.[25] Thus, vapor-liquid equilibrium data are essential for the design of separation processes required in the purification of esters. Using this force field, phase equilibria of different pure linear esters are accurately predicted. In contrast, overestimation of the bubble pressure has been observed for mixtures.[25]

OPLS (optimized potential for liquid simulations) is one of the popular force fields that has been used in computer simulations for a long time.[30-35] In 1991, they provided force field parameters for esters.[29] They have used Monte Carlo simulations to obtain parameters for liquid methyl acetate at 25 °C.[29] Here, they have chosen potential parameters in order to reproduce experimental thermodynamic and physical properties.[29] The OPLS parameters were developed using calculations on gas phase molecules.[29] After that, these charges were adjusted in order to reproduce experimental and thermochemical and structural information for liquids.[29] A Mulliken population analysis of 6-31G(d) wave functions were used initially to obtain the atomic charges of the molecules.[29] In this study, they have performed a Monte Carlo simulations for the pure liquids and checked for densities and heat of vaporization during the fitting process.[29] In addition, they have calculated the isothermal compressibility, the coefficient of thermal expansion, and the heat capacity.[29] The main objective was to obtain relevant parameters to represent common solvents and terminal groups for polypeptides.[29] When they derive the parameters only liquid phase

properties at room temperature have been considered.[29, 36] Therefore, the transferability of this model is potentially problematic.

SPASIBA is another force field that has developed parameters for esters in order to investigate biomolecules related to lipids.[37] They have derived potential energy parameters by minimizing the average error between the observed and calculated structures, conformational energy differences, vibrational frequencies and predicted quantities for a series of esters.[37] Charges were derived from ab initio full geometry optimizations with the B3LYP/6-31G** level of theory for esters. In their study, they have investigated the moment of inertia and dipole moment to compare with the experimental values.[37] The calculated values were in good agreement with the experimental values.[37]

A new transferable united atom force field for esters was developed as an extension of the transferable AUA4 force field.[38] The main objective of the AUA4 force field was to obtain phase equilibria of pure compounds and mixtures.[38] They have developed parameters for many major organic molecules including linear alkanes,[39] branched alkanes,[40] olefins,[41] benzene,[42] etc. Partial atomic charges were calculated using an ab initio calculations.[38] Here, they have place the selected molecule in a dielectric media that has the dielectric constant of the neat liquid for the given molecule, and the partial charges were adjusted to reproduce the dipole moment.[38] The main objective of this force field development was to accurately predict the phase equilibrium of pure esters as well as mixtures.[38] They have investigated several pure compound properties such as saturated liquid densities, vapor pressures, vaporization enthalpies, critical properties, liquid-vapor surface tensions.[38] Moreover, they have obtained good agreement for the estimation of binary-mixture pressure composition diagrams.[38] Here, they have not introduced empirical binary interaction parameters.[38] Thermophysical properties of compounds that contain esters are widely

applied in industry such as for food chemistry, pharmaceutics, oil and gas industry. Therefore, in order to design the industrial processes it is very useful to have a good understanding of relevant phase equilibria.[38]

MacKerell and coworkers have also derived parameters for esters, with the overall goal of extending the CHARMM force field for lipids.[13] CHARMM is one of the most promising force fields for the simulation of biological systems.[43-48] They have used methyl acetate, ethyl acetate and methyl propionate as model compounds.[13] Ab-initio calculations were used to obtain interaction parameters.[13] They have tested the interaction parameters using pure liquid simulations. Here, they have calculated the heat of vaporization and molecular volume to compare with experimental values. Energy differences between conformers were used to modify torsional parameters. Moreover, intramolecular parameters were fit to the experimental geometry and frequencies.[13] They have recently updated their parameters in order to reproduce experimental properties for several lipid types.[49]

There are several different approaches to obtain force field parameters.[43] In our previous work, we have pointed out many draw backs of currently available force fields.[50] As a result we have started to develop a new force field for biomolecular simulations.[50-51] Our approach is quite different from the traditional approaches and we mainly consider properties of solution mixtures.[50-51] We have been using Kirkwood-Buff theory to help obtain force field parameters.[51] Nowadays, this theory is widely applied to study solution mixtures and we can obtain proper relationships between KB integrals and solution activities, which is a very important factor that we can use in force field development.[50] Our previous studies have shown that the KB derived force fields can reproduce the experimental behavior of many solution mixtures more correctly than other force fields.[50]

3.2 Methods

In this study, several acetate mixtures were simulated using classical molecular dynamics simulations. Mainly, we have used methyl acetate-water, methyl acetate-methanol, methyl acetate-ethanol, ethyl acetate-methanol, methyl propionate-methanol systems to optimize the partial atomic charges for the ester linkage. We have used alcohol solvents because most of these esters are immiscible in aqueous medium. On the other hand alcohols are polar solvents. In this study, all the simulations were carried out at 298 K and 1 atm unless stated otherwise.

The KB integrals (G_{ij}) are defined by Equation (3.1),

$$G_{ij} = 4\pi \int_0^{\infty} \left[g_{ij}^{\mu VT}(r) - 1\right] r^2 \, dr \approx 4\pi \int_0^{R} \left[g_{ij}^{NpT}(r) - 1\right] r^2 dr \tag{3.1}$$

where g_{ij} is the corresponding radial distribution function (rdf). For a closed system, KB integrals can be calculated using above approximation. In this study, we have used different approach to obtain the simulated KB integrals rather than the traditional approach. Here, an expression for the finite-volume KB integrals are used, and then these integrals are linearly extrapolated to obtain a value corresponding to an infinite system.[52-54] The main reason that we used this method was due to convergence problems associated with the traditional expression. In particular, we observed possible convergence problems with the water-methyl acetate system. Using this new approach we could obtain the same KB integrals using a different method. Schnell and coworkers have used this method to obtain the KB integrals according to,

$$G_{ij} = V\left(\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_i\rangle}\right) \tag{3.2}$$

where N_i is the number of i particles inside V, <...> denotes grand canonical ensemble averages, and δ_{ij} is the Kronecker delta function. Hence, particle number fluctuations have been used to obtain the KB integrals.[52-54]

A Kirkwood-Buff analysis of the experimental data for all the acetate mixtures was carried out as explained by Ben-Naim and in our previous studies.[1-4, 7, 9] Experimental activities and densities were taken from the literature for all acetate mixtures.[55-59] A simple mixture rule based on volume fractions was used to obtain the compressibilities.[7, 60-61] Partial molar volumes were determined from the experimental density data by calculating the excess molar volume,

$$X_m^E = X_m - x_2 X_{m,2}^0 - x_1 X_{m,1}^0 \tag{3.3}$$

where X is the volume (V). $V_m = V/(N_1 + N_2)$ is the molar volume of the solution and $V_{m,1}^0$ is the molar volume of pure component 1.

The excess volume and excess molar Gibbs free energy values of acetate-alcohol and acetate-water systems were fitted to the Redlich-Kister equation or NRTL equation.[62] The Redlich-Kister fitting equation is shown in Equation (3.4),

$$X_m^E = x_1 x_2 \sum_{i=0}^{n} a_i \, (x_1 - x_2)^i \tag{3.4}$$

where a_i are fitting constants, x_i are mole fractions, and X is either the volume or Gibbs free energy. Partial molar quantities at any composition are then given by the standard relationship,

$$Y_1 = X_m^E - x_2 \left[\frac{\partial X_m^E}{\partial X_2}\right]_{p,T} \tag{3.5}$$

where X can be V or βG giving rise to the properties (Y) corresponding to the partial molar volume ($\bar{V}$) and excess chemical potential ($\beta\mu^E = lnf$) respectively. The NRTL equation is shown in Equation (3.6),

$$\frac{g^E}{RT} = x_1 x_2 \left(\frac{\tau_{21} G_{21}}{x_1 + x_2 G_{21}} + \frac{\tau_{12} G_{12}}{x_2 + x_1 G_{12}}\right) \tag{3.6}$$

$$ln\gamma_1 = x_2^2 \left[\tau_{21} \left(\frac{G_{21}}{x_1 + x_2 G_{21}}\right)^2 + \frac{\tau_{12} G_{12}}{(x_2 + x_1 G_{12})^2}\right] \tag{3.7}$$

$$ln\gamma_2 = x_1^2 \left[\tau_{12} \left(\frac{G_{12}}{x_2 + x_1 G_{12}}\right)^2 + \frac{\tau_{21} G_{21}}{(x_1 + x_2 G_{21})^2}\right] \tag{3.8}$$

where $\tau_{12} = (g_{12} - g_{22})/RT$; $\tau_{21} = (g_{21} - g_{11})/RT$; $G_{12} = \exp(-\alpha_{12}\tau_{12})$; $G_{21}\exp(-\alpha_{12}\tau_{21})$ Here, we have indicated the standard notation of the NRTL equation (Equation 3.6-Equation 3.8). G_{12} and G_{21} values are determined using the above expressions and are not KB integrals.

Chemical potentials, partial molar volumes, and compressibilities of the solution mixtures can be related to the KB integrals according to following equations,[63]

$$G_{12} = RT\kappa_T - \frac{\bar{V}_1 \bar{V}_2}{(1 + f_{22})V_m} \tag{3.9}$$

$$G_{11} = G_{12} + \frac{1}{x_1}\left(\frac{\bar{V}_2}{(1 + f_{22})} - V_m\right) \tag{3.10}$$

where R is the gas constant, x_1 is the mole fraction of component 1, V_m is the molar volume, and

$$\beta \left(\frac{\partial \mu_2}{\partial lnx_2}\right)_{p,T} = 1 + \left(\frac{\partial lnf_2}{\partial lnx_2}\right)_{p,T} = 1 + f_{22} \tag{3.11}$$

with ($\beta = 1/RT$) and f_2 equal to the solute activity coefficient on the mole fraction scale with the pure solute as the standard state. Using this approach suggested by Ben-Naim, experimental densities, compressibilities, and activity coefficients were used to determine the experimental KB integrals.[63] Hence, we use experimental KB integrals, obtained using common thermodynamic data, as the target data for developing the force field parameters.

In this study we have used the SPC/E[64] water model together with the KBFF methanol and ethanol models.[6, 65] All the Lennard-Jones parameters were taken from previous studies.[5-6] Bonded parameters were taken from the GROMOS force field.[66] Acetate dihedral parameters were obtained using Equation (3.12).

$$V_\psi = k_\psi[1 + \cos(n\psi - \delta)] \tag{3.12}$$

where V_ψ is dihedral interaction energy, k_ψ is force constant, n is periodicity and δ is phase. The carbonyl O-C-O-CH_3 dihedral parameters were obtained from fitting to the rotational potential curves obtained from QM calculations.[67] The corresponding rotational barrier around 90° is about 14.6 kcal/mol. For the ethyl acetate molecule the C-O-CH_2-CH_3 dihedral angle was also determined from fitting to QM calculations.[68] The corresponding rotational barrier around 0° is then about 6.84 kcal/mol. For methyl propionate molecule C-C-C-O dihedral angles were obtained from our previous work considering the glutamine molecule. The partial atomic charge on acetate atoms were then adjusted to reproduce the experimental KB integrals.

All the simulations were performed in the isothermal isobaric (NpT) ensemble at 300 K and 1 atm unless stated otherwise using Gromacs simulation package (version 4.6).[69] Berendsen[70] pressure coupling and v-rescale[71] temperature coupling techniques were used with relaxation times of 5 and 0.1 ps, respectively. All bonds were constrained using Lincs and a relative tolerance of

10^{-4}, allowing a 2 fs time step for integration of equation of motions.[72] The particle mesh Ewald technique was used to calculate the electrostatic interactions.[73] Random initial configurations were generated in a 10 nm cubic simulation box. All the systems were equilibrated (2 ns) and then performed a production run of 20 ns.

In this study we have determined self diffusion coefficients using the mean square displacement approach.[74] Dielectric constant values were calculated using the dipole moment fluctuations, using a reaction field permittivity of $\varepsilon_{RF} = \infty$ corresponding to the Ewald conducting boundary conditions.[75] Furthermore, enthalpy of mixing values were determined using the following equation,

$$\Delta H_m = H_{sol} - x_1 H_1 - x_2 H_2 \tag{3.13}$$

where ΔH_m is the enthalpy of mixing, H_{sol} is the molar enthalpy of the solution, H_1 and H_2 are the molar enthalpy of the pure components 1 and 2, respectively.

The Kirkwood Buff derived force field (KBFF) is a non polarizable force field. Bonded parameters were taken from the GROMOS96 force field. Non bonded parameters are listed in Table 3.1. Bonded parameters are listed in Table 3.2. Non bonded van der Waals interactions were calculated using a Lennard-Jones (LJ) potential and electrostatic interactions were calculated using a Coulombic potential. The combination rules were applied in order to determine cross terms; $\varepsilon_{ij} = (\varepsilon_{ii}\varepsilon_{jj})^{1/2}$ and $\sigma_{ij} = (\sigma_{ii}\sigma_{jj})^{1/2}$.

Table 3.1 Nonbonded parameters for the models

Model	Atom	ε (kJ/mol)	σ (nm)	q (e)
methyl acetate				
KBFF	C	0.330	0.336	0.585
	Carbonyl O	0.560	0.310	-0.535
	CH_3 (C)	0.867	0.374	0.000
	O	0.650	0.319	-0.4
	CH_3 (O)	0.867	0.374	0.35
ethyl acetate				
KBFF	$CH_3(CH_2)$	0.867	0.374	0
	$CH_2(0)$	0.410	0.407	0.35
methyl propionate				
KBFF	CH_2	0.410	0.407	0
	$CH_3(CH_2)$	0.867	0.374	0
methanol				
KBFF	O	0.650	0.319	-0.82
	H	0.088	0.158	0.52
	CH_3	0.867	0.374	0.3
ethanol				
KBFF	CH_2	0.410	0.407	0.3
	O	0.650	0.319	-0.82
	H	0.088	0.158	0.52
	CH_3	0.867	0.374	0.0
water				
SPC/E	O	0.6506	0.3166	-0.8476
	H	0.0	0.0	0.4238

After significant trial and error the best simulated KB integrals were obtained when using the charge distribution displayed in Table 3.1.

Table 3.2 Bonded parameters for KBFF

Bonds	r (nm)		
C=O	0.123		
C-CH_3	0.153		
C-O	0.136		
O-CH_3	0.143		
C-CH_2	0.153		
CH_2-CH_3	0.153		
Angles	k_θ (kJ/mol/rad)	θ_0 (degrees)	
Carbonyl O-C-CH_n	685	121	
Carbonyl O-C-O	730	124	
CH_n-C-O	610	115	
C-O-CH_n	450	109	
C-C-C	520	109.5	
Dihedrals	k_ψ (kJ/mol/rad)	δ (degrees)	n
O=C-O-CH_n	2.9	180	1
O=C-O-CH_n	24.81	180	2
C-O-CH_2-CH_n	-4	0	1
C-O-CH_2-CH_n	2.09	0	2
C-O-CH_2-CH_n	8	0	3
CH_n-CH_2-C-O	2.75	0	1
CH_n-CH_2-C-O	-4.75	0	2
Impropers	k_ω (kJ/mol/rad)	ω_0	
C-CH_n-O-O	334.8	0.0	

Potential functions are: Angles, $V_\theta = 1/2\, k_\theta(\theta - \theta_0)^2$; Dihedrals $V_\psi = k_\psi[1+\cos(n\psi-\delta)]$; Improper dihedrals $V_\omega = 1/2\, k_\omega(\omega - \omega_0)^2$

3.3 Results

The simulated KB integrals and the experimental KB integrals for acetate alcohol mixtures are compared in Figure 3.1. When we increase the acetate mole fraction, we observe a significantly higher association of methanol molecules. This is to be expected, considering the higher methanol-methanol interaction when we increase the number of acetate molecules in the system. For instance, methanol molecules contain both hydrogen bond donor and acceptor properties while acetate molecules contain the hydrogen bond accepter properties. Hence, methanol molecules can

make stronger hydrogen bond networks compared to the acetate molecules in the solution mixture and this leads to self-aggregation. Our models well reproduce the mole fraction dependent KB integrals. For all of these acetate alcohol systems the largest deviation from the experimental values were obtained for low alcohol mole fraction. Generally, uncertainty of both simulation and experimental data is higher at extreme mole fractions.

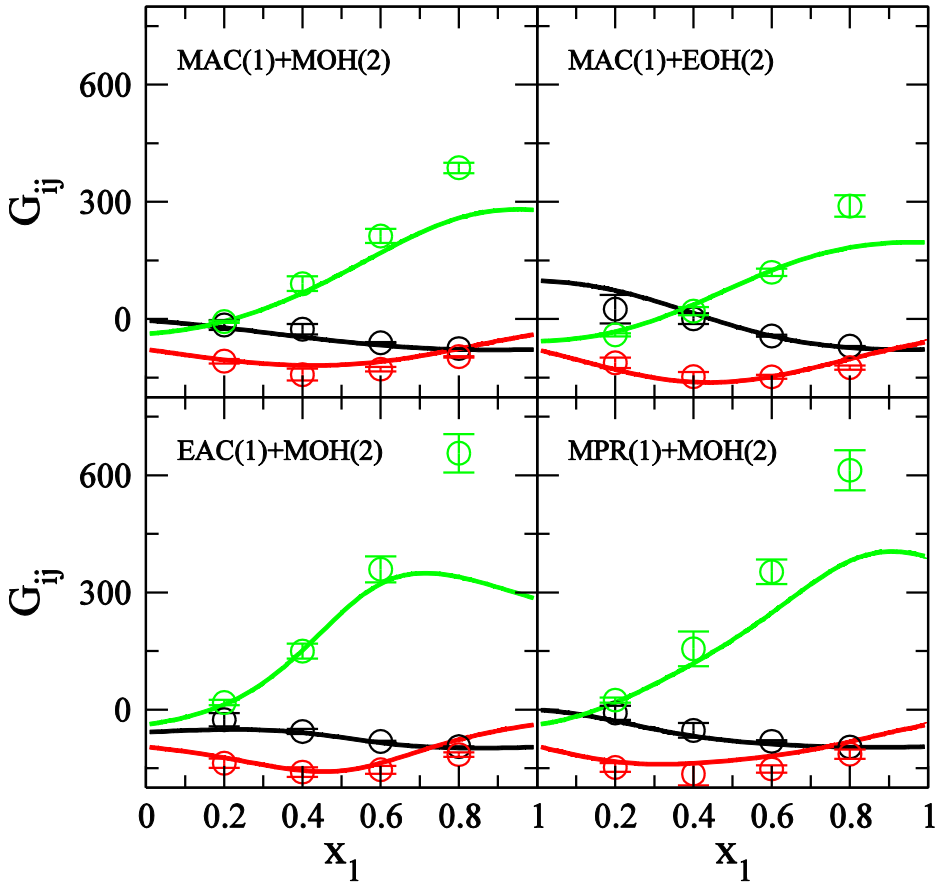

Figure 3.1 Kirkwood-Buff integrals (G_{ij}, cm^3/mol) as a function of acetate mole fraction (x_1). Solid lines represent the experimental data. All experimental data correspond to 298.15 K. The circles are the results for the KBFF parameterization. Black, red, green colors correspond to acetate-acetate, acetate-alcohol, alcohol-alcohol KB integrals respectively.

In both the ethyl acetate-methanol and methyl propionate-methanol systems we observe that the simulated methanol-methanol KB integrals are somewhat off from the experimental values at higher mole fractions of acetate. We have used several values for the partial atomic charges and

compared with the experimental KB integrals several times. This was the best charge distribution that we could find that reproduce KB integrals reasonably for all the systems including the water-methyl acetate system.

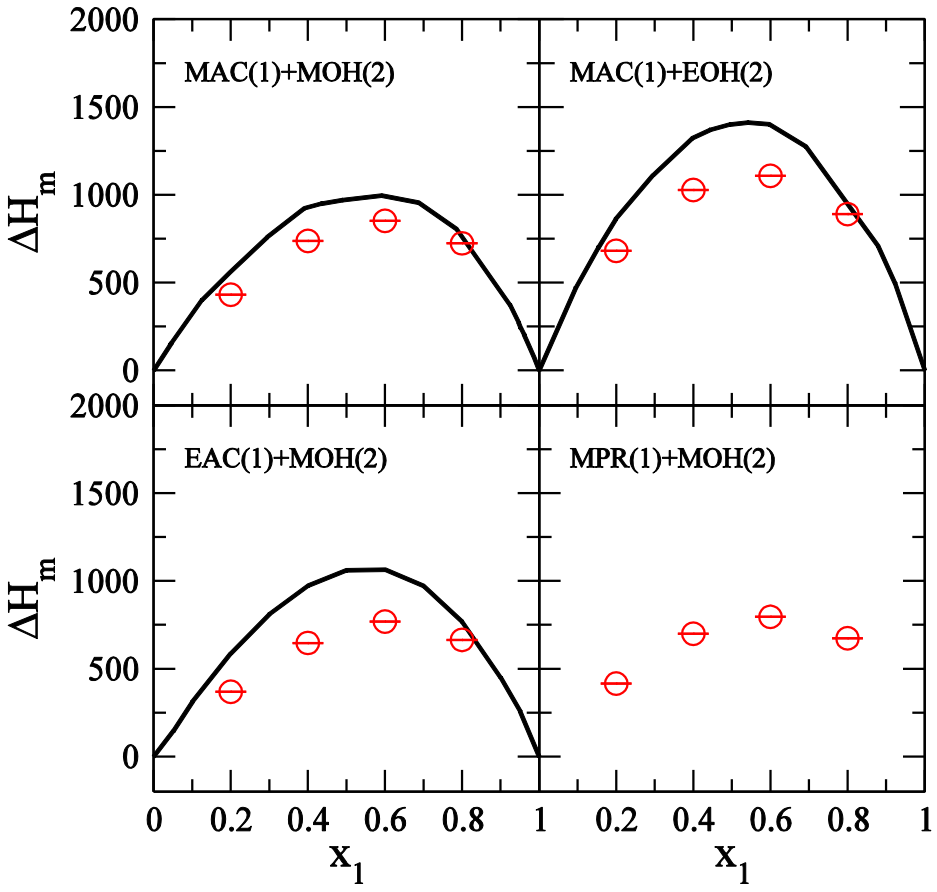

Figure 3.2 Enthalpy of mixing (ΔH_m, J/mol) values as a function of acetate mole fraction (x_1). Lines represent the experimental data[56, 58] and circles represent the KBFF model. All experimental data corresponds to 298.15 K.

The enthalpy of mixing values as a function of acetate mole fraction are displayed in Figure 3.2. We can see unfavorable enthalpy of mixing values for these acetate-alcohol systems from both experiment and simulation. For methyl propionate-methanol system we do not have experimental data. The simulated enthalpy of mixing values are more favorable than the experimental values for all the other three systems. However, the simulation data well reproduce the correct trends in the

experimental enthalpy of mixing. The methyl acetate-ethanol system shows a relatively positive experimental enthalpy of mixing value and that behavior is well reproduced in our models. Generally, it is difficult to get the enthalpy of mixing values correct for most force fields.[74]

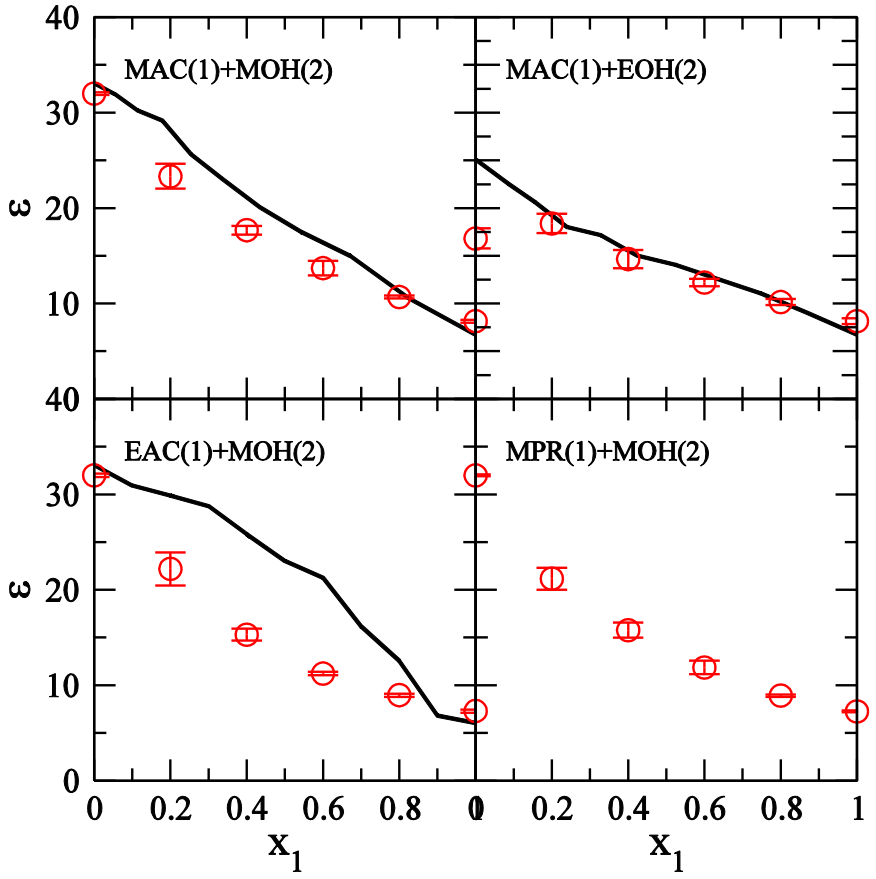

Figure 3.3 Relative permittivity (ε) values as a function of acetate mole fraction (x_1). Lines represent the experimental data[76,77] and circles represent the KBFF model. All the experimental data correspond to 298.15 K.

Relative permittivity values are shown in Figure 3.3. Although our models slightly underestimate the experimental dielectric constant values, these models well reproduce the trends of variation of dielectric constants with mole fraction. For methyl propionate-methanol system we do not have experimental dielectric constant values.

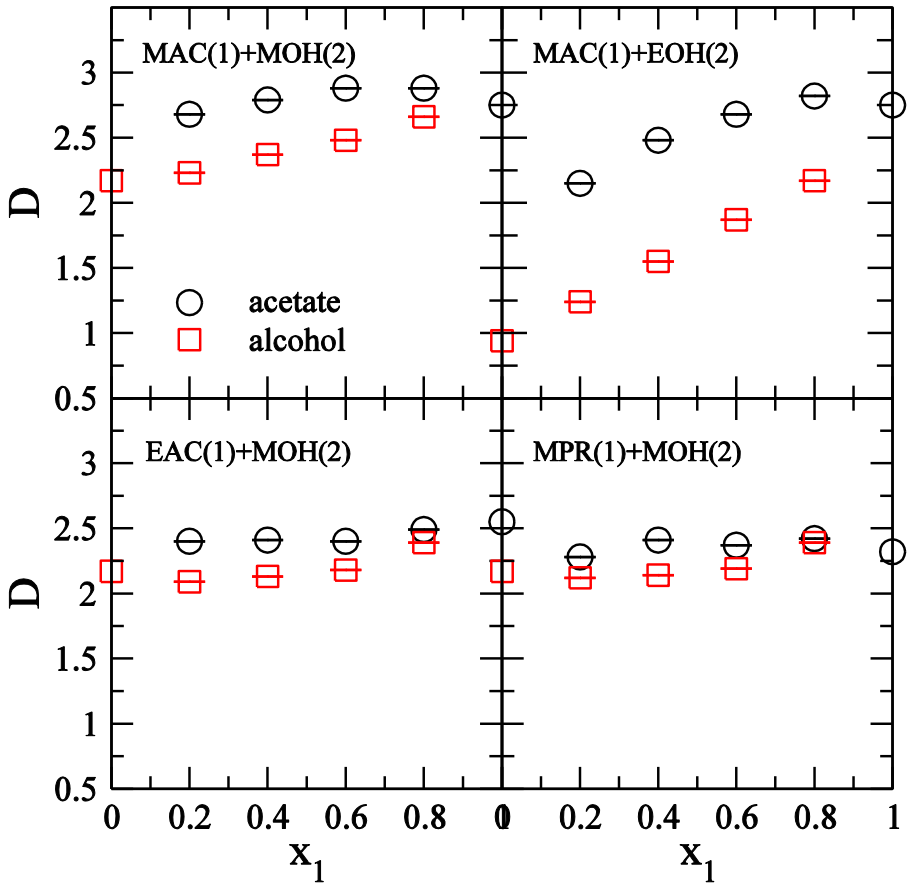

Figure 3.4 Self diffusion constants (D, 10^{-5} cm^2/s) as a function of acetate mole fraction (x_1). Squares and circles represent alcohol and acetate self diffusion coefficients respectively.

Self diffusion constant values of acetate and alcohol molecules are shown in Figure 3.4. Unfortunately, we do not have experimental diffusion constant values. For all the mixtures, we observe that the alcohol diffusion constant values are less compared to the acetate diffusion constant values.

The KB integrals provide an overall description of the relative distribution of molecules over all solvation shells. Most of the important interactions such as hydrogen bonds are dominant within first solvation shell. Thus, we have also calculated first shell coordination numbers for all the mole fractions. The main idea was to determine whether there is a significant possibility to form hydrogen bonds between acetate molecules and alcohol molecules. The calculated first shell

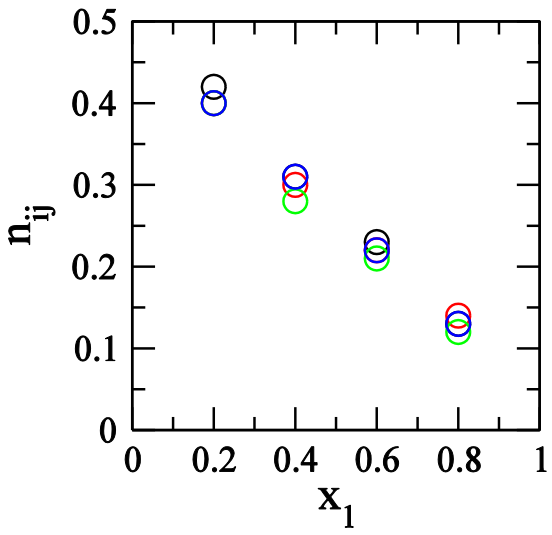

Figure 3.5 First shell coordination number (r_{min}≈ 0.3 nm) as a function of acetate mole fraction (x_1). i is carbonyl oxygen atom and j is methanol oxygen atom. Black, red, green, blue circles represent methyl acetate-methanol, methyl acetate-ethanol, ethyl acetate-methanol, methyl propionate-methanol systems respectively.

coordination numbers are displayed in Figure 3.5. To calculate first shell coordination numbers we have selected the carbonyl oxygen atom and the "alcohol" O atom. By obtaining atom-atom radial distribution function we can find the R_{min} corresponding to first solvation shell. Then we have obtained coordination numbers correspond to the R_{min} values. We do not see any strong hydrogen bonds between acetate and alcohol molecules. For all the systems, we do see a decrease in the first shell coordination numbers with an increase in the methyl acetate mole fraction.

3.3.1 Water-methyl acetate System

The water-methyl acetate system is one of the most important systems as many biological molecules are in an aqueous environment. One of the major challenges with this water-methyl acetate system is that these two components are not fully miscible over the whole composition range. There is a relatively large immiscible region. So we had to use a limited range of data for the parameterization process. In addition, there is a relatively low number of experimental studies

that have investigated water acetate solution mixtures. In this study we have analyzed the water-methyl acetate system. Here, we could reproduce the experimental KB integrals at higher methyl acetate mole fractions. But, unfortunately, at lower methyl acetate mole fraction we couldn't obtain the correct simulated KB integrals to match with experimental KB integrals. At lower methyl acetate mole fraction (x_1=0.025) we observe significant aggregation of methyl acetate molecules with G_{11} about 2000 cm^3/mol (but not converged). We do not see this behavior in the experimental

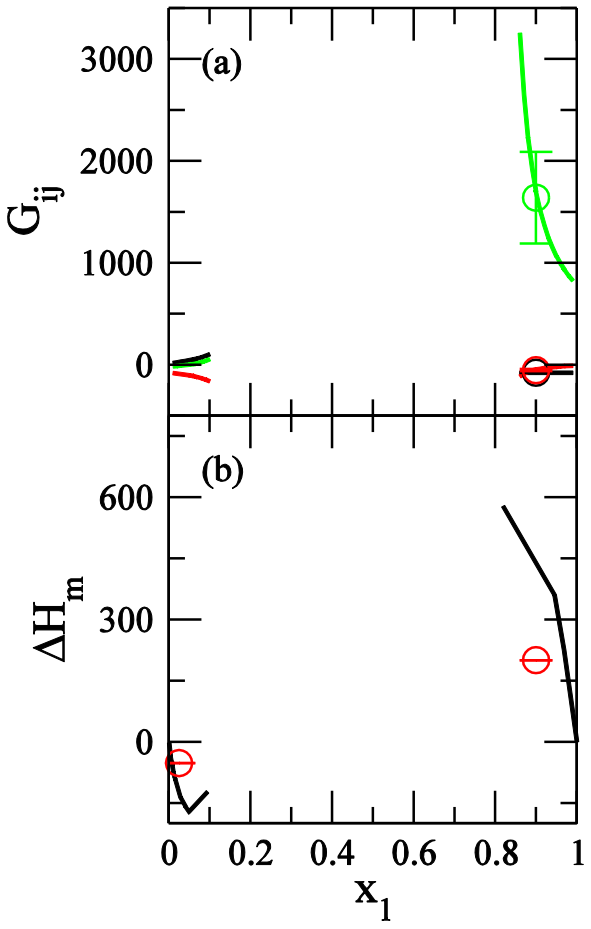

Figure 3.6 Panel (a) shows Kirkwood-Buff integrals (G_{ij}, cm^3/mol) as a function of acetate mole fraction (x_1). Black, red, green colors correspond to acetate-acetate, acetate-water, water-water KB integrals respectively. Lines represent the experimental data and circles represent the KBFF model. Panel (b) shows Water - methyl acetate enthalpy of mixing (ΔH_m, J/mol) values as a function of acetate mole fraction (x_1). Lines represent the experimental data[78] and circles represent the KBFF model. All the experimental data correspond to 298.15 K.

KB integrals.

Figure 3.6 (a) shows the water methyl acetate KB integrals. Water and methyl acetate molecules are immiscible in the middle of the composition range. At higher mole fractions of methyl acetate all the KB integrals are well reproduced. Figure 3.6 (b) shows the enthalpy of mixing values for a water methyl acetate mixture. Enthalpy of mixing values are also very sensitive to the partial atomic charges. With our model we could reproduce reasonable enthalpy of mixing values at both end of the composition range.

The main problem associated with the water-methyl acetate solution mixture is that at low mole fractions methyl acetate shows very high self association. Here, the methyl acetate-methyl acetate radial distribution functions are not converged to one. Therefore, if we use the traditional approach to calculate KB integrals it can be inaccurate. Consequently, we have used the particle particle fluctuation approach to calculate the KB integrals to ensure the higher aggregation was not due to the use of finite system sizes. Therefore, we have performed simulations increasing the box size up to 200 Å. However, still the observed radial distribution functions were not converged to one. Then, we have simulated several water methyl acetate systems with 4 different box sizes. Figure 3.7 shows the KB integrals of a 0.025 mole fraction of methyl acetate using 4 different box sizes. Here, we wanted to extrapolate these curves to predict the KB integrals for an infinitely large box. But still we do not obtain the correct trend.

We have also studied these water methyl acetate systems using OPLSAA, AMBER and CHARMM force fields because we wanted to investigate whether these force fields can reproduce the KB integrals correctly.[79-81] The results are shown in Table 3.3. For both lower and higher methyl acetate mole fractions we could not obtain reasonable values for KB integrals using OPLSAA force field. The AMBER and CHARMM force fields resulted in comparatively lower

values for methyl acetate – methyl acetate KB integrals than the KBFF model at x_1=0.025. However, at x_1=0.9 the water-water KB integrals were underestimated. Hence, none of the force fields provide perfect results for both the KB integrals and enthalpy of mixing. The KBFF model reproduces the sign of the enthalpy of mixing correctly at both composition extremes.

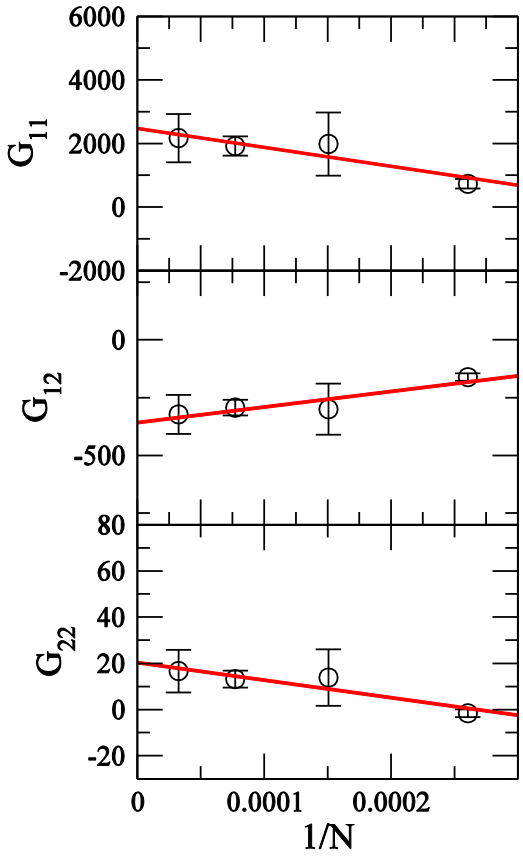

Figure 3.7 KB integrals (G_{ij}, cm^3/mol) as a function of the inverse total number of molecules in the box for methyl acetate-water system. The methyl acetate mole fraction was 0.025. Total number of molecules ($N=N_1+N_2$) are corresponding to 4 simulation box sizes;10 nm, 7.5 nm, 6 nm, 5 nm.

Recently, significant research has been performing concerning mesoscale inhomogeneities of aqueous solutions.[82-88] Sedlak and coworkers have studied the kinetics and long time stability of these large scale supra molecular structures.[82-84] They have studied about 100 different solute-solvent pairs using static and dynamic light scattering techniques.[84] For example, a variety of

people have studied about mesoscale inhomogeneity in aqueous tertiary butyl alcohol system. [85-88] However, according to our knowledge, there is no evidence for methyl acetate-water system aggregation on the mesoscale. However, we doubt the mesoscale inhomogeneities can be seen in methyl acetate-water system. Because, from the simulations, it is possible to observe methyl acetate clusters (at lower mole fraction of methyl acetate), that contain dimensions of several nano meters which agrees with the experimental observed mesoscopic structure dimensions for other systems. According to the experimental evidence, these long lived mesoscopic structures should have dimensions on the order of several hundred of nanometers.[82, 85]

Table 3.3 Comparison of the KB integrals (G_{ij}, cm^3/mol) and enthalpy of mixing (ΔH_m, J/mol) values for KBFF, OPLSAA,[80] AMBER,[81] CHARMM[79] force fields. Methyl acetate molecules are denoted as molecule type 1 and water molecules denoted as molecule type 2

	x_1=0.025				x_1=0.9			
Force field	G_{11}	G_{12}	G_{22}	ΔH_m	G_{11}	G_{12}	G_{22}	ΔH_m
KBFF	2166*	-322	16	-52	-78	-53	1639	290
OPLSAA	41990*	-4750	508	97	-58	-730	25977*	439
AMBER	486	-135	-4	-87	-76	-27	312	-216
CHARMM	119	-97	-8	-227	-79	-22	239	-299
Experiment	24	-85	-10	-136	-77	-44	1701	360

All simulations were performed at 300 K and 1 atm in the NpT ensemble. We have used SPC/E water model with KBFF and TIP3P water model with OPLSAA, AMBER, CHARMM. * indicates the KB integrals are not converged.

Table 3.4 Properties of the pure acetate models. ρ, E_{pot}, $E_{pot}^{(intra)}$, $E_{pot}^{(inter)}$, D, ε, η indicate density, potential energy per molecule, intramolecular energy per molecule, intermolecular energy per molecule, diffusion coefficient, relative permittivity and viscosity, respectively.

Molecule	Property	Experiment	Simulation	Units
Methyl acetate	ρ	0.926[57]	0.921	g/cm^3
	E_{pot}		-92.75	kJ/mol
	$E_{pot}^{(intra)}$		14.67	kJ/mol
	$E_{pot}^{(inter)}$		-107.42	kJ/mol
	D		2.75	10^{-5} cm^2/s
	ε	6.74[76]	8.14	
	η	0.336[89]	0.387	cp
Ethyl acetate	ρ	0.894[57]	0.883	g/cm^3
	E_{pot}		-84.41	kJ/mol
	$E_{pot}^{(intra)}$		21.53	kJ/mol
	$E_{pot}^{(inter)}$		-105.94	kJ/mol
	D		2.55	10^{-5} cm^2/s
	ε	6.06[77]	7.29	
	η	0.305[90]	0.306	cp
Methyl propionate	ρ	0.909[59]	0.887	g/cm^3
	E_{pot}		-88.30	kJ/mol
	$E_{pot}^{(intra)}$		16.25	kJ/mol
	$E_{pot}^{(inter)}$		-104.55	kJ/mol
	D		2.32	10^{-5} cm^2/s
	ε		7.26	
	η		0.345	cp

The simulation results for the pure liquid esters are presented in Table 3.4. The density, relative permittivity and viscosity data are in reasonable agreement with the available experimental

values. Unfortunately, we could not find experimental relative permittivity and viscosity data for the methyl propionate system, or diffusion coefficients for all three acetates.

3.4 Conclusions

Force field parameters for esters have been developed using KB theory. The overall goal was to provide an accurate description of lipids. In this work we have used methyl acetate, ethyl acetate, methyl propionate as the model compounds for the ester linkage of the lipid molecule. To derive a suitable charge distribution for acetate molecules we have studied acetate-alcohol and methyl acetate-water mixtures and attempted to reproduce the experimental KB integrals. Our model reasonably well reproduce the experimental data for all the acetate-alcohol mixtures. Furthermore, these models well reproduce the enthalpy of mixing values and dielectric constant values. For the methyl acetate-water system, we could reproduce the experimental KB integrals at higher mole fractions of methyl acetate. But, we couldn't obtain reasonable values for the methyl acetate-methyl acetate KB integrals at lower mole fractions of methyl acetate. However, this is also appears to be a problem with several other force fields.

3.5 References

1. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *118* (23), 10663-10670.

2. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003,** *107* (16), 3891-3898.

3. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003,** *119* (21), 11342-11349.

4. Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004,** *121* (5), 2180-2186.

5. Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006,** *27* (13), 1477-1485.

6. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005,** *109* (31), 15080-15086.

7. Bentenitis, N.; Cox, N. R.; Smith, P. E. *J. Phys. Chem. B* **2009,** *113* (36), 12306-12315.

8. Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011,** *13* (40), 18154-18167.

9. Gee, M. B.; Cox, N. R.; Jiao, Y.; Bentenitis, N.; Weeerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011,** *7* (5), 1369-1380.

10. Weerasinghe, S.; Gee, M. B.; Kang, M.; Bentenitis, N.; Smith, P. E. In *Developing force fields from the microscopic structure of solutions: the Kirkwood-Buff approach*, Wiley-VCH Verlag GmbH & Co. KGaA: 2010; pp 55-76.

11. Davis, J. E.; Rahaman, O.; Patel, S. *Biophys. J.* **2009,** *96* (2), 385-402.

12. Lundbæk, J. A.; Collingwood, S. A.; Ingólfsson, H. I.; Kapoor, R.; Andersen, O. S. *J. R. Soc. Interface.* **2010,** *7* (44), 373-395.

13. Schlenkrich, M.; Brickmann, J.; MacKerell, A., Jr.; Karplus, M. An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications. In *Biological Membranes*, Merz, K., Jr.; Roux, B., Eds. Birkhäuser Boston: 1996; pp 31-81.

14. Biggin, P. C.; Bond, P. J. *Methods Mol. Biol.* **2015,** *1215*, 91-108.

15. Poger, D.; Caron, B.; Mark, A. E. *Biochim. Biophys. Acta*. In Press

16. Bovigny, C.; Tamo, G.; Lemmin, T.; Maino, N.; Dal Peraro, M. *J. Chem. Inf. Model.* **2015,** *55* (12), 2491-2499.

17. Ashkar, R.; Nagao, M.; Butler, P. D.; Woodka, A. C.; Sen, M. K.; Koga, T. *Biophys. J.* **2015,** *109* (1), 106-112.

18. White, P. J.; Davenport, R. J. *Plant Physiol.* **2002,** *130* (3), 1386-1395.

19. Kukol, A. *Methods Mol. Biol.* **2015**, *1215*, 73-90.

20. Cranfield, C.; Carne, S.; Martinac, B.; Cornell, B. *Methods Mol. Biol.* **2015,** *1232*, 45-53.

21. Lensink, M. F. *Methods Mol. Biol.* **2015,** *1215*, 109-124.

22. Yesylevskyy, S. O.; Demchenko, A. P. *Methods Mol. Biol.* **2015,** *1232*, 291-306.

23. Tan, J.; Ye, S.; Luo, Y. *J. Phys. Chem. C* **2015,** *119* (51), 28523-28529.

24. Agrawal, A.; Steigmann, D. J. *Z. Angew. Math. Phys.* **2011,** *62* (3), 549-563.

25. Kamath, G.; Robinson, J.; Potoff, J. J. *Fluid Phase Equilib.* **2006,** *240* (1), 46-55.

26. Cui, S. T.; Siepmann, J. I.; Cochran, H. D.; Cummings, P. T. *Fluid Phase Equilib.* **1998,** *146* (1,2), 51-61.

27. Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2001,** *105* (15), 3093-3104.

28. Stubbs, J. M.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2004,** *108* (45), 17596-17605.

29. Briggs, J. M.; Nguyen, T. B.; Jorgensen, W. L. *J. Phys. Chem.* **1991,** *95* (8), 3315-3322.

30. Jorgensen, W. L.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1988,** *110* (6), 1657-66.

31. Pranata, J.; Wierschke, S. G.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1991,** *113* (8), 2810-2819.

32. Briggs, J. M.; Matsui, T.; Jorgensen, W. L. *J. Comput. Chem.* **1990,** *11* (9), 958-971.

33. Tirado-Rives, J.; Jorgensen, W. L. *J. Am. Chem. Soc.* **1990,** *112* (7), 2773-2781.

34. Jorgensen, W. L.; Laird, E. R.; Nguyen, T. B.; Tirado-Rives, J. *J. Comput. Chem.* **1993,** *14* (2), 206-215.

35. Jorgensen, W. L.; Nguyen, T. B. *J. Comput. Chem.* **1993,** *14* (2), 195-205.

36. Jorgensen, W. L.; Tirado-Rives, J. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (19), 6665-6670.

37. Chhiba, M.; Tristram, F.; Vergoten, G. *J. Mol. Struct.* **1997,** *405* (2–3), 113-122.

38. Ferrando, N.; Lachet, V.; Boutin, A. *J. Phys. Chem. B* **2012,** *116* (10), 3239-3248.

39. Ungerer, P.; Beauvais, C.; Delhommelle, J.; Boutin, A.; Rousseau, B.; Fuchs, A. H. *J. Chem. Phys.* **2000,** *112* (12), 5499-5510.

40. Bourasseau, E.; Ungerer, P.; Boutin, A.; Fuchs, A. H. *Mol. Simul.* **2002,** *28* (4), 317-336.

41. Bourasseau, E.; Haboudou, M.; Boutin, A.; Fuchs, A. H.; Ungerer, P. *J. Chem. Phys.* **2003,** *118* (7), 3020-3034.

42. Contreras-Camacho, R. O.; Ungerer, P.; Boutin, A.; Mackie, A. D. *J. Phys. Chem. B* **2004,** *108* (37), 14109-14114.

43. Mackerell, A. D. *J. Comput. Chem.* **2004,** *25* (13), 1584-1604.

44. Hatcher, E.; Guvench, O.; Mackerell, A. D. *J. Phys. Chem. B* **2009,** *113* (37), 12466-76.

45. Foloppe, N.; Mackerell, A. D. *J. Comput. Chem.* **2000,** *21* (2), 86-104.

46. Mackerell, A. D.; Banavali, N. K. *J. Comput. Chem.* **2000,** *21* (2), 105-120.

47. Kamath, G.; Guvench, O.; MacKerell, A. D. *J. Chem. Theory Comput.* **2008,** *4* (5), 765-778.

48. Hatcher, E. R.; Guvench, O.; MacKerell, A. D. *J. Chem. Theory Comput.* **2009,** *5* (5), 1315-1327.

49. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. *J. Phys. Chem. B* **2010,** *114* (23), 7830-7843.

50. 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.

51. Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010,** *290* (1–2), 43-47.

52. Schnell, S. K.; Vlugt, T. J. H.; Simon, J.-M.; Bedeaux, D.; Kjelstrup, S. *Chem. Phys. Lett.* **2011,** *504* (4–6), 199-201.

53. Schnell, S. K.; Liu, X.; Simon, J.-M.; Bardow, A.; Bedeaux, D.; Vlugt, T. J. H.; Kjelstrup, S. *J. Phys. Chem. B* **2011,** *115* (37), 10911-10918.

54. Krüger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J.-M. *J. Phys. Chem. Lett.* **2013,** *4* (2), 235-238.

55. Iglesias, M.; Marino, G.; Orge, B.; Piñeiro, M. M.; Tojo, J. *Phys. Chem. Liq.* **1999,** *37* (3), 193-213.

56. Nagata, I.; Ohta, T.; Takahashi, T. *J. Chem. Eng. Jpn.* **1972,** *5* (3), 227-231.

57. González, B.; Calvar, N.; Gómez, E.; Domínguez, Á. *J. Chem. Thermodyn.* **2007,** *39* (12), 1578-1588.

58. Murti, P.; Van Winkle, M. *Industrial & Engineering Chemistry Chemical & Engineering Data Series* **1958,** *3* (1), 65-71.

59. Polak, J.; Lu, B. C. Y. *J. Chem. Eng. Data* **1972,** *17* (4), 456-458.

60. Oswal, S. L.; Oswal, P.; Modi, P. S.; Dave, J. P.; Gardas, R. L. *Thermochim. Acta* **2004,** *410* (1-2), 1-14.

61. Easteal, A. J.; Woolf, L. A. *J. Chem. Thermodyn* **1985,** *17* (1), 49-62.

62. *Physical and chemical equilibriums for chemical engineers*. John Wiley & sons: New Jersey, 2012.

63. Ben-Naim, A. *Molecular theory of solutions*. Oxford university press, 2006.

64. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

65. Jiao, Y. The development of accurate force fields for protein simulation. Kansas State University, 2012.

66. van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I.; Tironi, G. *Biomolecular Simulation: The GROMOS96 Manual and User Guide*. Zürich, Switzerland: Vdf Hochschulverlag at the ETH Zürich, 1996.

67. Nagy, P. I.; Tejada, F. R.; Sarver, J. G.; Messer, W. S. *J. Phys. Chem. A* **2004,** *108* (46), 10173-10185.

68. Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.

69. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013,** *29* (7), 845-54.

70. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

71. Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007,** *126* (1), 014101.

72. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

73. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

74. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2000,** *104* (24), 5854-5864.

75. Allen, M. P.; Tildesley D. J. *Computer Simulations of Liquids*. Oxford University Press: 1987.

76. Shirke, R. M.; Chaudhari, A.; More, N. M.; Patil, P. B. *J. Chem. Eng. Data* **2000,** *45* (5), 917-919.

77. Shirke, R. M.; Chaudhari, A.; More, N. M.; Patil, P. B. *J. Mol. Liq.* **2001,** *94* (1), 27-36.

78. Hoch, M. *Calphad* **1997,** *21* (3), 359-379.

79. Fischer, N. M.; van Maaren, P. J.; Ditz, J. C.; Yildirim, A.; van der Spoel, D. *J. Chem. Theory Comput.* **2015,** *11* (7), 2938-2944.

80. Price, M. L. P.; Ostrovsky, D.; Jorgensen, W. L. *J. Comput. Chem.* **2001,** *22* (13), 1340-1352.

81. Virtualchemistry.org. http://virtualchemistry.org/index.php.

82. Sedlák, M. *J. Phys. Chem. B* **2006,** *110* (9), 4329-4338.

83. Sedlák, M. *J. Phys. Chem. B* **2006,** *110* (9), 4339-4345.

84. Sedlák, M. *J. Phys. Chem. B* **2006,** *110* (28), 13976-13984.

85. Sedlák, M.; Rak, D. *J. Phys. Chem. B* **2013,** *117* (8), 2495-2504.

86. Sedlák, M.; Rak, D. *J. Phys. Chem. B* **2014,** *118* (10), 2726-2737.

87. Gupta, R.; Patey, G. N. *J. Chem. Phys* **2012,** *137* (3), 034509.

88. Subramanian, D.; Anisimov, M. A. *J. Phys. Chem. B* **2011,** *115* (29), 9179-9183.

89. González, B.; Domínguez, A.; Tojo, J. *J. Chem. Eng. Data* **2004,** *49* (6), 1590-1596.

90. Modarress, H.; Mohsen-Nia, M. *Phys. Chem. Liq.* **2006,** *44* (1), 67-76.

Chapter 4 - Preferential Solvation in Binary and Ternary Mixtures

4.1 Abstract

Preferential solvation has become a useful tool to help characterize and understand the properties of liquid mixtures. Here, we provide a new quantitative measure of preferential solvation in binary and ternary mixtures that uses Kirkwood-Buff integrals as input, but differs from traditional measures. The advantages of the new measure are highlighted and compared with established literature approaches. Molecular dynamics simulations are performed to further investigate the nature of binary mixtures, as described by the new and existing measures of preferential solvation. It is shown that the new measure of preferential solvation is rigorous, has a simple physical interpretation, can be easily related to the underlying thermodynamic properties of the mixture, and naturally leads to zero values for ideal mixtures.

4.2 Introduction

Most solution mixtures are homogeneous on the macroscopic scale. All but ideal solutions, however, can be significantly inhomogeneous on the microscopic scale.[1] The local distribution of molecules around a central molecule can deviate from the bulk distribution due to variations in both the molecule sizes (packing) and the interactions between the different molecules. The mutual net attraction (or repulsion) between molecules leading to an inhomogeneous distribution over short length scales is generally referred to as preferential solvation (*PS*). *PS* has significant consequences for the observed physical and thermodynamic properties of a mixture.[2-7]

While the concept of *PS* is relatively simple, attempts to define and quantify *PS* in liquid mixtures have been more difficult. For favorable systems, such as host-guest molecules or strong hydrogen bond donor-acceptor pairs in apolar solvents, the effect of solvent composition on the

equilibrium constant for association can be used very effectively.[8-9] Alternatively, for sparingly soluble solutes the effects of cosolvents on the solubility of solutes can be cast in terms of preferential interactions and thereby *PS*.[10-14] However, most simple fully miscible liquid mixtures, such as alcohol and water mixtures do not fall into either of these categories. The interactions and deviations from the bulk distribution are much more subtle in these cases and are not amenable to analysis using equilibrium constants or solubilities. This type of system is the focus of the present study. For these latter systems, and some of the former, the most common and general quantitative approaches have involved the use of Kirkwood-Buff (KB) theory.[3, 6, 15-17]

Kirkwood-Buff theory is an exact theory of solution mixtures.[18] It relates the thermodynamics of any stable multicomponent mixture to the relative distribution of molecules within the liquid *via* a series of Kirkwood-Buff integrals (KBIs) between all molecule pair types. These integrals can be obtained from experimental thermodynamic (activity, density, and compressibility) data,[19] and can then be used to quantify the mutual affinity between the different components within a mixture. Hence, information concerning *PS* should be available from these integrals, and the solution "structure" can then be related to the thermodynamic properties.[1, 20] Ben-Naim provided the first rigorous framework for studying *PS* in binary and ternary mixtures using the experimental KBIs.[3, 21-22] This has since been extended by others. Unfortunately, these extensions have resulted in some controversy regarding the properties of reference or ideal solution mixtures, leading to new measures of *PS* that use corrected expressions, that is outlined in detail below.

The aim of the this work is to provide a general measure of *PS* using KBIs that is valid for all stable mixtures, has a rigorous definition and physical meaning, adopts desirable behavior for ideal mixtures, and simplifies many of the previous KBI based approaches. We illustrate this

approach using both experimental and simulation data, and then compare and contrast with existing measures of *PS*. In addition, using the simulated KBIs we also examine some of the approximations and issues inherent to the previous approaches.

4.3 Theory

4.3.1 General

Here we outline the basic theory leading to the most common quantitative descriptions of *PS*. The general approach is due to Ben-Naim,[1, 3] and we retain the notation of A, B, C, *etc.,* as referring to the different types of species present in solution. First, let us define a distance dependent KBI,

$$G_{AB}(R) \equiv 4\pi \int_0^R [g_{AB}(r) - 1] r^2 \, dr \tag{4.1}$$

where g_{AB} is the radial distribution function (rdf) between molecules A and B, defined in the grand canonical ensemble, and r is the intermolecular (center of mass) distance. We note that the KBIs obtained from experimental data, using the usual KB inversion approach,[19] correspond to the limit $R \to \infty$. However, it is often presumed a local region of solution exists around each central molecule such that beyond this local region the distribution of other molecules resembles the bulk distribution.[1, 23-24] In this case $g_{AB}(r)$ is essentially unity when r is large (but not necessarily infinite) and the integral may be considered converged. The size of this local region is generally unknown and will depend on the components, the composition, and the state point. However, simulation results on many mixtures under ambient conditions indicate that, while deviations from

bulk behavior can extend over many nanometers, the rdfs are indeed structureless beyond the first few solvation shells and may be safely assumed to be unity.[23]

Using the above integral the local average number of *B* molecules observed within the spherical volume *V* defined by a distance *R* around a central *A* molecule can then be written,

$$< N_B >_A = 4\pi\rho_B \int_0^R g_{AB}(r) r^2 dr = \rho_B G_{AB}(R) + \rho_B V \tag{4.2}$$

where ρ_B is the bulk number density (molarity) of *B* molecules. Consequently, the local mole fraction of *B* molecules around a central *A* molecule within a sphere of radius *R* can then be expressed as,

$$x_{BA}^L(R) = \frac{< N_B >_A}{\sum_\alpha < N_\alpha >_A} = x_B \frac{V + G_{AB}(R)}{V + Y_A(R)} \tag{4.3}$$

where the sum is over all components in the mixture. The right hand side was obtained after using Equation (4.2) followed by the substitution $Y_A(R) \equiv \sum x_\alpha G_{A\alpha}(R)$. Alternatively, the local number density of *B* molecules around a central *A* molecule is given by,

$$\rho_{BA}^L(R) = \frac{< N_B >_A}{\sum_\alpha < N_\alpha >_A \bar{V}_\alpha} = \rho_B \frac{V + G_{AB}(R)}{V + Y_A^\phi(R)} \tag{4.4}$$

where $\bar{V}_\alpha$ is the partial molar volume of species α. The right hand side was obtained after using Equation (4.2) followed by the substitution $Y_A^\phi(R) \equiv \sum \phi_\alpha G_{A\alpha}(R)$, where $\phi_\alpha = \rho_\alpha \bar{V}_\alpha$ is the volume fraction of species *α* in the bulk mixture. Finally, the local volume fraction of *B* molecules around a central *A* molecule can also be determined *via*,

$$\phi_{BA}^{L}(R) = \rho_{BA}^{L}(R)\bar{V}_B = \frac{< N_B >_A \bar{V}_B}{\sum_\alpha < N_\alpha >_A \bar{V}_\alpha} = \phi_B \frac{V + G_{AB}(R)}{V + Y_A^{\phi}(R)} \tag{4.5}$$

Note that $\rho_{BA}^{L}(R) \neq < N_B >_A / V$ due to the excluded volume of the central molecule.

The partial molar volumes appearing in the above equations could be expanded in terms of KBIs.[18] However, it is simpler to retain the partial molar volumes themselves. Furthermore, the partial molar volumes are physically easy to understand, and are available from any analysis for which the KBIs are also obtained.

4.3.2 Existing Measures of Preferential Solvation

The above expressions can be used to determine the deviation of the local solution composition from that of the bulk composition. Traditionally, this has focused on changes in the mole fraction composition. However, here it will be extended to include volume fractions, and thereby number densities, in the following sections. A measure for the *PS* of a central *A* molecule by *B* molecules can be defined as,

$$\delta_{BA}(R,V) \equiv x_{BA}^{L}(R) - x_B = x_B \frac{G_{AB}(R) - Y_A(R)}{V + Y_A(R)} \tag{4.6}$$

Note that we have written the *PS* as a function of two variables. The first is the integration distance for the KBIs (R), while the second is the volume of the local region (V). According to Equation (4.2) these are required to be consistent ($V = 4\pi R^3/3$). However, some of the approaches discussed later will relax this condition. Using volume fractions one finds,

$$\delta_{BA}^{\phi}(R) \equiv \phi_{BA}^{L}(R) - \phi_B = \phi_B \frac{G_{AB}(R) - Y_A^{\phi}(R)}{V + Y_A^{\phi}(R)} \tag{4.7}$$

We will not allow R and V to vary independently in this case. All the above quantities tend to zero as the local volume increases. The expressions are useful as they quantify deviations in the local solution composition as a function of distance from a central molecule of interest. They are problematic, however, as the spatial dependence of the KBIs, and the extent of the local volume of interest, are unknown. There are two general solutions to this problem. Both are described below.

Before discussing these approaches it is informative to develop and examine the expressions in more detail. A general equation for the partial molar volume of a solute in any multicomponent mixture is available and can be used to provide,[25-27]

$$V_A^* \equiv \bar{V}_A - RT\kappa_T = \sum_\alpha \phi_\alpha \, G_{A\alpha}(\infty) = -Y_A^\phi(\infty) \tag{4.8}$$

where V^* is known as the pseudo volume, κ_T is the isothermal compressibility, R the Gas constant (not to be confused with the integration distance), and T the absolute temperature. This expression only holds for the fully integrated KBIs. It can be used to simplify the limiting form of the volume fraction based *PS* expressions. For example,

$$\delta_{BA}^\phi(R \to \infty) = \frac{\phi_B[G_{AB}(\infty) + V_A^*]}{V - V_A^*} \tag{4.9}$$

In this form, the measure of *PS* only involves one explicit KBI. This seems logical as the KBIs indeed directly quantify the deviation from the bulk distribution. However, this is less apparent using the mole fraction approach in Equation (4.6). Rearrangement of the above expression followed by the use of the definition presented in Equation (4.7) gives,

$$\rho_B G_{AB}(\infty) = \rho_{BA}^{L}(V - V_A^*) - \rho_B V \tag{4.10}$$

which provides a rigorous and concise meaning for the fully integrated KBIs and the local density. Specifically, $\rho_B G_{AB}(\infty)$ is the difference between the average number of *B* molecules around a central *A* molecule in a large volume *V* of solution compared to the average number of *B* molecules that would be found in the same large volume of bulk solution. The exact size of the "large" volume of solution is irrelevant as long as it is big enough that the bulk distribution ($g_{AB} = 1$) is encountered. Note that the local number density does not include the volume occupied by the central *A* molecule.

All the above expressions are exact for any number of components. For binary solutions the expressions are typically presented in an alternative form,

$$\delta_{BA}(R,V) = -\delta_{AA}(R,V) = x_A x_B \frac{G_{AB}(R) - G_{AA}(R)}{V + x_A G_{AA}(R) + x_B G_{AB}(R)} \tag{4.11}$$

for the mole fraction based expressions and,

$$\delta_{BA}^{\phi}(R) = -\delta_{AA}^{\phi}(R) = \phi_A \phi_B \frac{G_{AB}(R) - G_{AA}(R)}{V + \phi_A G_{AA}(R) + \phi_B G_{AB}(R)} \tag{4.12}$$

for the volume fraction based expressions. Expressions for the additional measures $\delta_{AB}(R,V) = -\delta_{BB}(R,V)$, and their volume fraction counterparts, can be obtained from a simple index change. It is clear from these expressions that all values approach zero when $R \to \infty$, and also when the concentration of either *A* or *B* tends to zero. Less obvious, but generally true, is that the denominator in the above expressions is positive for volumes larger than the excluded volume of the central molecule. Consequently, the sign of the above *PS* measures in binary systems is determined by the numerators, and hence just by the difference in two KBIs.

4.3.3 Ben-Naim Limiting Approach

As mentioned above, the primary disadvantage of the previous expressions for quantifying *PS* in liquid mixtures is that the distance dependence of the KBIs is unknown except for when the local volume approaches infinity, at which point the measures themselves are zero. The first solution to this problem was presented by Ben-Naim.[3] In this approach only the limiting (large volume) behavior is used. The limiting *PS* expressions for binary mixtures are then given by,

$$\delta_{BA}^{o} \equiv \left.\frac{\partial \delta_{BA}}{\partial V^{-1}}\right|_{V^{-1}=0} = x_A x_B [G_{AB}(\infty) - G_{AA}(\infty)] \tag{4.13}$$

for the mole fraction approach and,

$$\delta_{BA}^{\phi,o} \equiv \left.\frac{\partial \delta_{BA}^{\phi}}{\partial V^{-1}}\right|_{V^{-1}=0} = \phi_A \phi_B [G_{AB}(\infty) - G_{AA}(\infty)] = \rho^2 \overline{V}_A \overline{V}_B \delta_{BA}^{o} \tag{4.14}$$

for the volume fraction approach, where ρ is the total number density. Note that these limiting values no longer depend on *R* (or *V*). Hence, the evaluation of the KBIs for a specific local volume is avoided. The limiting quantities then describe the change in the *PS*, or local solution composition, as one approaches the central molecule from the bulk solution region ($g_{AB} = 1$). The required KBIs are the values extracted from experiment, using the usual KB inversion approach,[19, 28] and may be used directly and without approximation. The main advantage of this approach is that the rigorous link to the solution thermodynamics is retained. The disadvantage of this approach lies in the absence of any spatial information, such as changes in the first solvation shell composition, that may be more relevant for explaining many experimental (spectroscopic) observations.[29]

4.3.4 Solvation Shell Approach

The second type of approach for determining *PS* attempts to explicitly evaluate the volume in Equation (4.6) or Equation (4.11) that corresponds to a particular solvation shell around each central molecule.[15, 30] Hence, the value of V is restricted to that of a series of solvation shells, or correlation volumes (V_{cor}'s), around each molecule. In doing so it is also implicitly assumed that the finite KBIs can be replaced by the infinite limit KBIs. This leads to expressions of the form,

$$\delta_{BA}\left(\infty, V_{cor,A}\right) = x_A x_B \frac{G_{AB}(\infty) - G_{AA}(\infty)}{V_{cor,A} + x_A G_{AA}(\infty) + x_B G_{AB}(\infty)} \tag{4.15}$$

for binary mixtures. However, the use of the infinite limit KBIs is only strictly valid when $V_{cor,A} \rightarrow \infty$, as indicated in Equation (4.2), and not for an intermediate correlation volume. Whether this approximation, $\delta_{BA}\left(\infty, V_{cor,A}\right) = \delta_{BA}(R_{cor,A}, V_{cor,A})$, is reasonable requires a comparison with the results from Equation (4.11). This suggests a simulation based approach to provide the partially integrated KBIs as illustrated below.

The final step involves the determination of the correlation volumes around each species. This has been achieved in a number of different ways.[7] However, the results are very similar and hence we only describe the approach used by Marcus.[15] Here, the correlation volume is related to the composition and volumes of the molecules representing the solvation shell in question. For binary mixtures the correlation volume(s) are expressed in terms of the molar volumes of the pure liquids (indicated by a zero superscript) and the local compositions according to,

$$\begin{aligned} V_{cor,A} = 2522.7[-0.085m + 0.5 * 0.1363(V_A^o)^{\frac{1}{3}} + 0.1363(m \\ - 0.5)(x_{AA}^L V_A^o + x_{BA}^L V_B^o)^{1/3}]^3 \end{aligned} \tag{4.16}$$

where *m* is an indicator of the solvation shell of particular interest, and all volumes are in cm^3/mol ($1\ nm^3 = 602.3\ cm^3/mol$). As the correlation volume equation actually uses the local compositions that it is intended to determine, the correlation volumes and local compositions have to be determined in an iterative manner at each bulk composition. The main advantage of this approach lies in the possible insights provided by the various solvation shells that may lead to a deeper physical picture of the solution structure. The main disadvantage is an inability to directly relate these measures to the underlying solution thermodynamics. We will only consider the first solvation shell correlation volumes in this work.

4.3.5 Ideal Solutions

Ideal solutions are a useful reference frame for understanding the behavior of real liquid mixtures. The KBIs for ideal mixtures are neither zero nor independent of composition. A general expression for the KBIs in any multicomponent symmetric ideal (SI) mixture has been provided,[28, 31]

$$G_{AB}^{SI} = RT\kappa_T - V_A^o - V_B^o + \sum_\alpha \rho_\alpha (V_\alpha^o)^2 \tag{4.17}$$

This leads to the following relationships that can be used to help simplify the corresponding *PS* parameters,

$$Y_A^{SI} = RT\kappa_T - V_A^o - V_m + \sum_\alpha \rho_\alpha (V_\alpha^o)^2 = G_{A\alpha}^{SI} + V_\alpha^o - V_m \tag{4.18}$$

$$Y_A^{\phi,SI} = RT\kappa_T - V_A^o = -V_A^{o,*} \tag{4.19}$$

where $V_m = 1/\rho$ is the molar volume. There is no spatial dependence associated with any of these integrals or functions.

Using the SI results in Equations (4.11) - Equation (4.15) produces finite *PS* parameters when the molecules involved possess different molar volumes. This is particularly evident in biomolecular systems where the protein volume is typically large compared to other components in the system. Consequently, there have been attempts to ensure that the *PS* parameters are zero for SI mixtures.[32-34] One approach is to modify the original expressions to give,

$$\Delta\delta_{BA}(\infty, V_{cor,A}) = -\Delta\delta_{AA}(\infty, V_{cor,A}) = x_A x_B \frac{\Delta G_{AB} - \Delta G_{AA}}{V_{cor,A} + x_A \Delta G_{AA} + x_B \Delta G_{AB}} \qquad (4.20)$$

where $\Delta G_{AB} = G_{AB}(\infty) - G_{AB}^{SI}$, *etc*. Alternative modifications have also been suggested but will not be discussed further here.[35] These measures are often referred to as "volume corrected" *PS* parameters and are intended to highlight the role of molecular interactions over the effects of different molecular volumes. This procedure has been the cause of some controversy. In particular, the exact meaning of the volume corrected expressions has been the subject of significant debate.[36-38] Nevertheless, there is clearly a desire for measures of *PS* that are zero for ideal mixtures.

4.3.6 Conservation of Volume Relationship

One of the major reasons for correcting the KBIs relates to the proposed expression,[16, 35]

$$\rho_A \Delta G_{AA} \bar{V}_A + \rho_B \Delta G_{AB} \bar{V}_B = 0 \qquad (4.21)$$

written as an equality for binary mixtures. This relationship is based on the suggestion that if there is an excess of *A* molecules around a central *A* molecule then there must be an equal volume deficit of *B* molecules that are replaced, *i.e.* there is a conservation of volume condition on perturbing the

molecule distributions from that of an equivalent SI mixture to that of the real mixture. The argument is essentially physical in nature and seems reasonable. Indeed, Equation (4.8) for binaries can be written,

$$[1 + \rho_A G_{AA}(\infty)]\bar{V}_A + \rho_B G_{AB}(\infty)\bar{V}_B = RT\kappa_T \tag{4.22}$$

which, after inclusion of the volume due to the central molecule, agrees with the spirit of the above condition, to within a term related to the compressibility, but with the reference system being a randomly distributed set of molecules.[26] Equation (4.22) is exact and, as one has $-V\kappa_T = (\partial V/\partial P)_{T,N} = (\partial^2 G/\partial p^2)_{T,N} = -(\partial^2 A/\partial V^2)_{T,N}^{-1}$, applies to any closed isothermal binary system.

By subtracting the analogous version of Equation (4.22) for ideal mixtures from Equation (4.22) for a real mixture one can obtain Equation (4.21) as an equality, but only when the excess volume of mixing for the real mixture is zero and independent of pressure.[35] Hence, Equation (4.21) must be viewed as an approximation,[36] albeit a very reasonable one for many systems where the excess volume of mixing is small. This is true even for a closed isothermal system at constant volume. A particle fluctuation view of Equation (4.22) and the volume conservation condition is discussed elsewhere.[39-41]

4.3.7 New Measure of Preferential Solvation in Liquid Mixtures

Some of the issues raised in the previous sections have prompted us to re-examine *PS* in multicomponent mixtures. First, we note that the concept of *PS*, while rigorously defined above, is in essence subjective. Consequently, many reasonable choices are possible. Here we choose the

following definition that quantifies the difference between the local distribution of A molecules around B and C molecules according to,

$$PS_{A|B-C}(R) \equiv \rho_{AB}^{L}(R) - \rho_{AC}^{L}(R) \tag{4.23}$$

using the local number density approach. This expression can be used to describe *PS* in any multicomponent mixture, although there will clearly be more possibilities as the number of components increases.[30] In binary mixtures one simply replaces C with A. The general expression for $PS_{A|B-C}(R)$ in terms of KBIs, obtained using Equation (4.4) or Equation (4.7), is somewhat cumbersome to use. However, by employing the Ben-Naim approach described above we find the limiting *PS* of B and C molecules by A molecules can be written quite simply as,

$$PS_{A|B-C}^{o} \equiv \left.\frac{\partial(\rho_{AB}^{L} - \rho_{AC}^{L})}{\partial V^{-1}}\right|_{V^{-1}=0} = \rho_A[G_{AB}(\infty) + V_B^* - G_{AC}(\infty) - V_C^*] \tag{4.24}$$

for any multicomponent mixture after using Equation (4.8). This dimensionless measure of *PS* tends to zero when $\phi_A \to 0$, and also when $\phi_A \to 1$; as Equation (4.8) indicates that $G_{AB}(\infty) \to -V_B^*$ and $G_{AC}(\infty) \to -V_C^*$ for any number of components under the latter conditions. The limiting *PS* is also zero for SI mixtures at all compositions as the presence of the pseudovolumes naturally accounts for the different excluded volumes of the two central molecules. Hence, we have maintained the rigor of the Ben-Naim limiting approach while also capturing the desired behavior that led to the development of the previous volume corrected quantities. The "trick" is to investigate the *change* in the distribution of a particular molecule around two *different* central molecules, rather than two different molecules around the *same* central molecule. If $PS_{A|B-C}^{o} > 0$ then one concludes that the A molecules prefer to accumulate around B molecules more than they do around C molecules, and *vice versa*, at a distance close to where $g_{AB} = g_{AC} = 1$ is satisfied.

The above definition of *PS* can be developed further to provide some insight into the exact meaning of the volume corrected quantities. Using the SI results in Equation (4.24) provides,

$$PS^{o}_{A|B-C} = \rho_A[\Delta G_{AB} - \Delta G_{AC} + (\bar{V}_B - V^{o}_B) - (\bar{V}_C - V^{o}_C)] \tag{4.25}$$

The volume differences in the above equation are simply the excess partial molar volumes for both central molecules. The magnitude of typical excess partial molar volumes is small (0-2 cm^3/mol) compared to the magnitude of the KBI values observed for mixtures under ambient conditions (0-10,000 cm^3/mol) and so one can write,

$$\rho_A(\Delta G_{AB} - \Delta G_{AC}) \approx PS^{o}_{A|B-C} = \left.\frac{\partial(\rho^{L}_{AB} - \rho^{L}_{AC})}{\partial V^{-1}}\right|_{V^{-1}=0} \tag{4.26}$$

to a very good approximation. Consequently, the volume corrected KBIs describe the limiting change in the local number density difference of A molecules around a central B and C molecule. Hence, by correcting the KBIs the measures of *PS* appear to correspond most closely to the new definition used in Equation (4.23), and not the original definition provided in Equations (4.13) or Equation (4.14).

Not all the possible measures of *PS* in binary and ternary systems are unique. Indeed there are several relationships between the various measures for binary and ternary mixtures. In addition to the obvious relationship that $PS^{o}_{A|A-B} + PS^{o}_{A|B-A} = 0$, using Equation (4.8) in Equation (4.24) provides,

$$\bar{V}_A PS^{o}_{A|B-A} + \bar{V}_B PS^{o}_{B|B-A} = 0 \tag{4.27}$$

for binary mixtures. Consequently, there is only one unique *PS* measure for binary mixtures. For ternary mixtures these relationships take the form,

$$PS^o_{A|A-B} + PS^o_{A|B-C} + PS^o_{A|C-A} = 0 \tag{4.28}$$

$$\bar{V}_A PS^o_{A|B-C} + \bar{V}_B PS^o_{B|B-C} + \bar{V}_C PS^o_{C|B-C} = 0 \tag{4.29}$$

where a permutation of indices provides a total of three expressions of each type. The first set of expressions can be proved by direct use of Equation (4.24). The second type of expression can be obtained from the definition of $PS^o_{A|B-C}$ and the subsequent expansion of the pseudovolumes using Equation (4.8) followed by simple rearrangement. Physically, the second set of relationships simply state that the sum of the local volume fractions is unity and therefore the same around both the B and C molecules. Consequently, there are only three unique measures of *PS* for a ternary mixture.

Before leaving this section we note that a similar approach using the mole fraction based δ's leads to,

$$\delta^o_{AB} - \delta^o_{AC} = \left.\frac{\partial(\delta_{AB} - \delta_{AC})}{\partial V^{-1}}\right|_{V^{-1}=0} = x_A[G_{AB}(\infty) - Y_B(\infty) - G_{AC}(\infty) - Y_C(\infty)] \tag{4.30}$$

which is also zero at $x_A = 0$, $x_A = 1$, and for any multicomponent SI solution. However, the volume fraction based expressions in Equation (4.24) provide a more general and simpler practical form, especially for ternary mixtures, due to the use of Equation (4.8).

4.3.8 Thermodynamics of Binary Mixtures

The most common application of *PS* studies is to binary mixtures. Here, there is only one unique measure of *PS* and we have,

$$\bar{V}_B PS^o_{B|A-B} = -\phi_A \phi_B \Delta_{AB} \tag{4.31}$$

The $\Delta_{AB} \equiv G_{AA}(\infty) + G_{BB}(\infty) - 2G_{AB}(\infty)$ term is instantly recognizable as a measure of the deviation from ideality provided by the KB theory of binary mixtures.[1] Hence, the proposed new measure of *PS* can be easily related to the solution thermodynamics to give,[1, 18]

$$\left(\frac{\partial ln\gamma_B}{\partial lnx_B}\right)_{p,T} = \frac{PS^o_{B|A-B}}{\rho\bar{V}_A - PS^o_{B|A-B}} \tag{4.32}$$

where γ is the mole fraction scale activity coefficient. This link to both solution "structure" and thermodynamics is possible,[21] but significantly less clear, using the traditional measures of *PS*. At low concentrations of *B* this simplifies to,

$$\left(\frac{\partial ln\gamma_B}{\partial lnx_B}\right)_{p,T,x_{B\to 0}} = PS^o_{B|A-B} \tag{4.33}$$

Consequently, if the *B* molecules prefer to associate with other *B* molecules, rather than *A* molecules, then the derivative is negative and the activity of *B* will decrease on increasing the *B* concentration. Alternatively, if the *B* molecules prefer to associate with *A* molecules, rather than other *B* molecules, then the derivative is positive and the activity of *B* will increase on increasing the *B* concentration. This type of behavior is well known, but can now be quantified and related, in a simple manner, to the new definition of *PS* and the experimentally available KBIs.

The new measure of *PS* is also intimately linked to the excess molar Gibbs free energy of mixing (G_m^E) for binary mixtures. Manipulation of Equation (4.32) using standard thermodynamic derivatives provides,

$$PS^o_{B|A-B} = \rho\bar{V}_A \frac{x_A x_B \beta(\partial^2 G_m^E/\partial x_B^2)_{p,T}}{1 + x_A x_B \beta(\partial^2 G_m^E/\partial x_B^2)_{p,T}} = \rho\bar{V}_A \frac{(\partial^2 G_m^E/\partial x_B^2)_{p,T}}{(\partial^2 G_m/\partial x_B^2)_{p,T}} \tag{4.34}$$

where $\beta = 1/RT$. The denominators in the above equation must be positive for stable (miscible) binary mixtures. Hence, except for the rare occasions that $\bar{V}_A$ may be negative, the sign of the new *PS* measure is determined by the curvature of the excess molar Gibbs free energy of mixing.

4.3.9 Thermodynamics of Ternary Mixtures

The application of *PS* in ternary mixtures is more complicated as it involves additional KBIs to completely characterize the mixture.[30] Nevertheless, the new definition of *PS* provided in Equation (4.23) can still be used. Here, we investigate the relationship between the new measure of *PS* and the solution thermodynamics in more detail. A common situation where this arises involves protein thermodynamics. Unfortunately, the traditional notation is different in this case. Typically, the index *1* is used to denote the primary solvent, index *2* is used for the (infinitely dilute) biomolecule solute and index *3* (or higher) is used for any additional cosolvents that may appear in the solution. The thermodynamics of ternary (protein) solutions provided by KB theory has been outlined in detail elsewhere.[13, 42-44] There are many expressions that can be used depending on the measure of concentration one adopts. Only selected (most common) examples are provided here.

The chemical potential of an infinitely dilute biomolecule depends on the cosolvent activity (a_3) according to,[27]

$$\beta\left(\frac{\partial\mu_2}{\partial ln a_3}\right)_{p,T,m_2\to 0} = -m_3 - \rho_3\left[\frac{PS^o_{3|2-1}}{\rho_3} - \frac{PS^o_{1|2-1}}{\rho_1}\right] = -m_3 - \frac{PS^o_{3|2-1}}{\phi_1} \qquad (4.35)$$

where $m_\alpha = \rho_\alpha / \rho_1$ is the molality of species α. The final step was achieved using the relationships provided in Equation (4.28) - Equation (4.29). This expression involves bulk solution properties and the difference between the cosolvent and water distributions around the biomolecule and

water. If the cosolvent displays a stronger affinity for the biomolecule over water than water displays for the biomolecule over itself, then the biomolecule chemical potential will decrease on addition of more cosolvent.

Another traditional measure of cosolvent "binding" is the preferential binding or interaction parameter. This takes slightly different forms depending on the thermodynamic constraints and the concentration scale adopted.[27, 45-46] The measure most commonly provided from equilibrium dialysis studies is given by,[47]

$$\frac{m_2}{m_3}\Gamma_{23} \equiv \left(\frac{\partial ln m_3}{\partial ln m_2}\right)_{T,\mu_1,\mu_3,m_2\to 0} = PS^o_{2|3-1} + \rho_2(\bar{V}_1 - \bar{V}_3) \tag{4.36}$$

and indicates that an affinity of the biomolecule for the cosolvent over the solvent results in an increase of the cosolvent molality in the vicinity of the protein on increasing the protein concentration. This, of course, is logical but can now also be quantified in a simple and direct manner using the new measure of *PS*. Both Equation (4.35) and Equation (4.36) contain additional terms unrelated to the biomolecule in question. However, these are properties of the bulk solution and will cancel when comparing different protein forms (native or denatured) and/or different proteins.

Finally, the effect of a cosolvent on the equilibrium ($K = \rho_D/\rho_N$, with $\rho_2 = \rho_N + \rho_D$) between a denatured protein (*D*) and the native protein (*N*) can be quantified according to,[13]

$$\left(\frac{\partial \ln K}{\partial ln a_3}\right)_{p,T,m_2\to 0} = \rho_3\left[\frac{PS^o_{D|3-1}}{\rho_D} - \frac{PS^o_{N|3-1}}{\rho_N}\right] \tag{4.37}$$

Hence, if the denatured protein displays a stronger affinity for the cosolvent (over water) than the native protein, then the equilibrium constant will increase on addition of more cosolvent. Again, these ternary system results are not so clear with some of the other *PS* approaches.

It seems logical that one could directly express the new *PS* measures in terms of the thermodynamic properties as illustrated in Equation (4.34) for binary mixtures. We attempted such a process, using our previous results in this area (specifically Equations 4.7- Equation 4.9),[41] but were ultimately unsuccessful.

4.4 Methods

Four binary mixtures were chosen for study as they display a range of *PS* behavior. These were: methanol (MOH) and water (HOH) at 300 K; methyl acetate (MAC) and methanol at 300 K; 1,2-dichloroethane (DCE) and methanol at 323 K; and isopropanol (POH) and water at 300 K. All systems correspond to a pressure of 1 bar. The infinite limit KBIs were obtained from the experimental data in the usual manner using the expression,[19, 41]

$$x_\alpha\left[\delta_{\alpha\beta} + \rho_\beta G_{\alpha\beta}(\infty)\right] = x_\alpha x_\beta \rho RT\kappa_T + \frac{(1-\phi_\alpha)(1-\phi_\beta)}{(1-x_\beta)\mu_{\alpha\beta}} \tag{4.38}$$

where $\delta_{\alpha\beta}$ is the Kronecker delta function and $\mu_{\alpha\beta} = \beta\left(\partial\mu_\alpha/\partial x_\beta\right)_{T,p}$ is a composition derivative of the chemical potential μ_α. The composition dependent chemical potential derivatives, partial molar volumes and isothermal compressibility were obtained from the composition derivatives of the excess Gibbs free energy and volume of mixing, and the pressure derivative of the density, respectively, as outlined elsewhere.[19, 48] Excess molar Gibbs free energies of mixing were taken directly from the literature,[49-52] as were the density and excess molar volume data.[53-56] The

compressibility data was taken to be ideal, which is a minor approximation,[57] with pure liquid values taken from the literature.[58-61]

Classical molecular dynamics simulations of each mixture were performed using the Gromacs simulation package (version 4.6).[62] The majority of force field models were taken from the literature: MOH/HOH;[63-64] MAC/HOH;[64] DCE(Model MP2/ExpSQ-Q)/MOH;[65] and POH/HOH.[64] The models for MAC and POH were developed by us and will be published shortly. The simulations were performed using a time step of 2 fs with bond lengths constrained using the LINCS and SETTLE algorithms.[66-67] Electrostatic interactions were determined using the particle mesh Ewald approach,[68] with a 1.0 nm cutoff for electrostatics and a twin range 1.0 and 1.5 nm cutoff for van der Waals interactions, except for the Amber based models (DCE/MOH) for which the traditional van der Waal and electrostatic cutoffs of 1.0 and 1.2, respectively, were used together with a van der Waals switch starting at 0.9 nm. All systems involved initial random placement of molecules in cubic boxes of length 10 nm. The systems were equilibrated from 5-10 ns and followed with 10 ns of production. Temperature and pressure coupling was achieved using the v-rescale and Berendsen algorithms at the experimental temperature and pressure,[69-70] respectively.

Both partially and fully integrated KBIs were obtained from direct integration. Other approaches to determine the KBIs are available, and their advantages and disadvantages have been discussed in detail.[23, 71] However, we found direct integration to be accurate enough for the present application. Comparison of the fully integrated KBIs obtained from direct integration with those determined *via* the local particle number fluctuations did not indicate any meaningful differences outside the deviations observed between individual block averages. Integration was performed to a distance ranging from 1.5 to 2.0 nm.

4.5 Results

The four systems studied here were chosen as they display a range of *PS* behavior indicated by the magnitude of the KBIs. In Figure 4.1 the experimental and simulated fully integrated KBIs are compared. Large positive values for the KBIs indicated a tendency for those species to

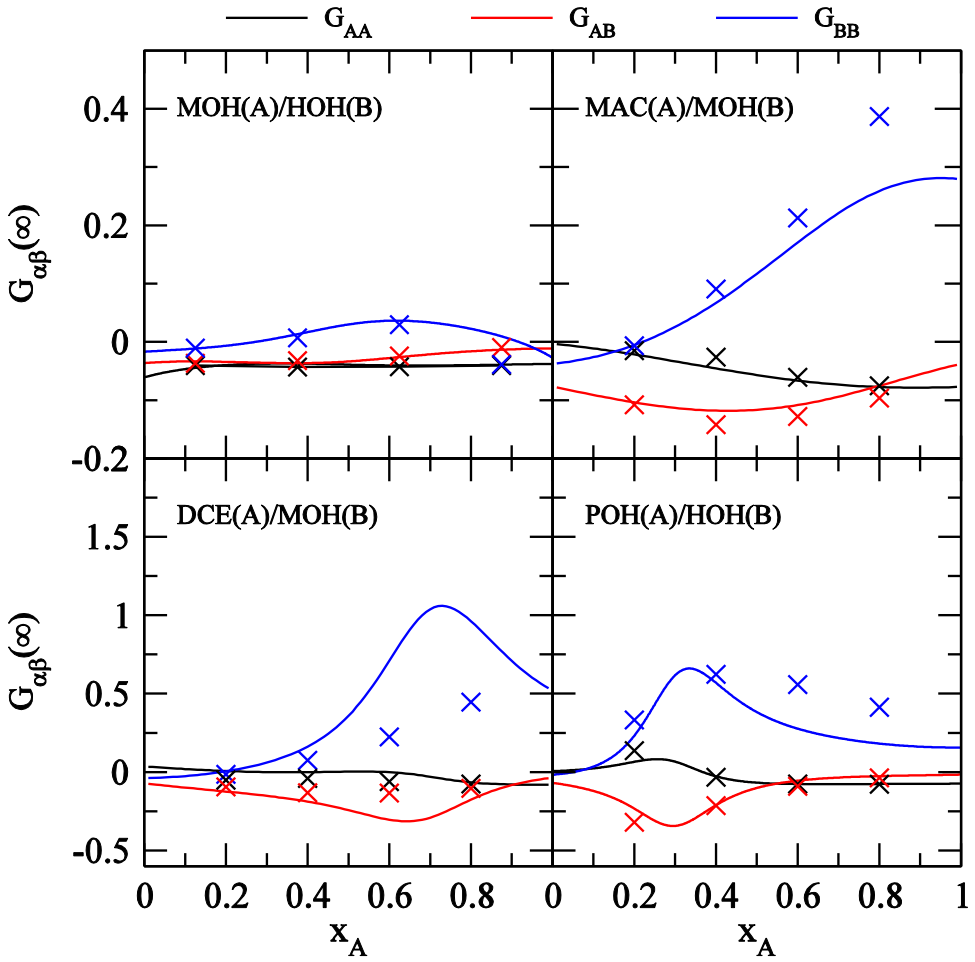

Figure 4.1 Experimental (lines) and simulated (symbols) fully integrated KBIs (L/mol) as a function of composition for four binary mixtures.

associate, thereby inferring deviations from the bulk distribution for some or all of the solvation shells involved. The trends in the composition dependent KBIs were reasonably well reproduced by the simulations, although there were some regions where the agreement was not quantitative.

For example, in mixtures of DCE/MOH at high DCE mole fractions the tendency for methanol self-association was reproduced but significantly underestimated. The MOH/HOH system displayed rather subtle changes in the KBIs in comparison with the other three systems where the KBIs were significantly larger in magnitude. This can be attributed to the inability of the more polar molecule (*B*) to satisfy its hydrogen bonding requirements as the volume fraction of the less polar molecule (*A*) started to dominate the mixture.[72] We note that the degree of agreement with experiment is primarily determined by the quality of the force fields used in the simulations, as sampling is not usually a concern. Hence, the agreement between experimental and simulated KBIs can be used to measure the quality of the models employed.[73] However, this was not the focus of the current study and none of the concepts or results presented here require agreement with experiment in order to be valid.

In Figure 4.2 the results of a *PS* analysis of the mixtures using the experimental KBIs are presented. The degree of *PS* varied substantially with composition for all but the MOH/HOH system. Quantitatively, most of the *PS* measures were different, however, they were qualitatively very similar and all but the MOH/HOH system suggested a large depletion of *B* molecules around a central *A* molecule that was characteristic of a preference for self-association at most compositions. The volume corrected *PS* measures differed only slightly from the uncorrected values due to the relatively small molecular volumes involved, except for the MOH/HOH mixtures where the correction led to a change in sign of the *PS* due to the relatively small initial values. The Ben-Naim and uncorrected solvation shell *PS* measures displayed the same features. This included similar changes in sign and also identical compositions for which their value was zero, as determined by the condition $G_{AB}(\infty) = G_{BB}(\infty)$. Conversely, the corrected solvation shell measures

and the values of $PS^{o}_{B|A-B}$ displayed essentially identical features, as expected from Equation (4.26), and little or no variation in sign with composition compared to the other *PS* measures. This is a consequence of the fact that the new measure of *PS* is related to the underlying thermodynamics through Equation (4.32), and that all four mixtures studied here are characterized by only positive deviations from ideality, *i.e.* by positive values of the excess molar Gibbs free energy of mixing. Clearly, measures of *PS* that display less variation in sign with composition are logically more attractive as multiple sign changes are significantly more difficult to explain.

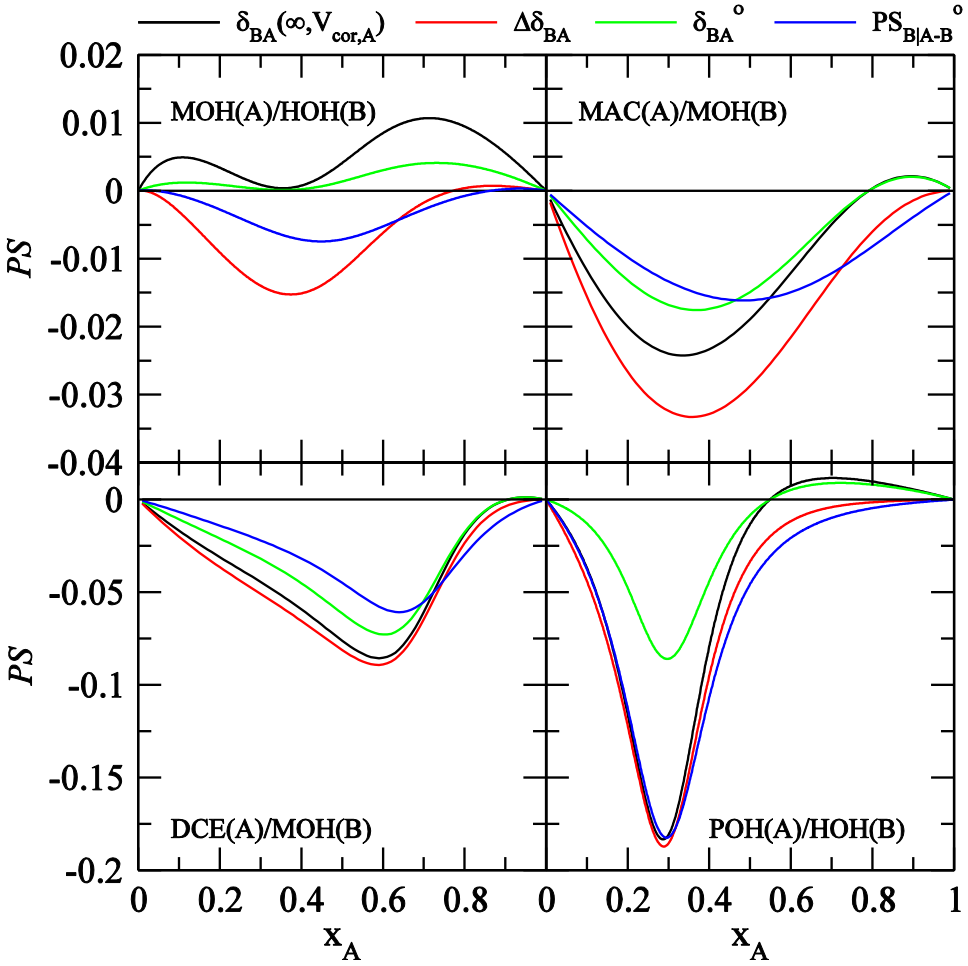

Figure 4.2 Experimental measures of *PS* as a function of composition for four binary mixtures. The values of δ^{o}_{BA} have units of L/mol, while the values of $PS^{o}_{B|A-B}$ have been scaled down by a factor of 100 for ease of comparison

The experimental and simulated *PS* measures are compared in Figure 4.3. Only two of the measures are presented due to the similarities mentioned above. Again, the simulations reproduced

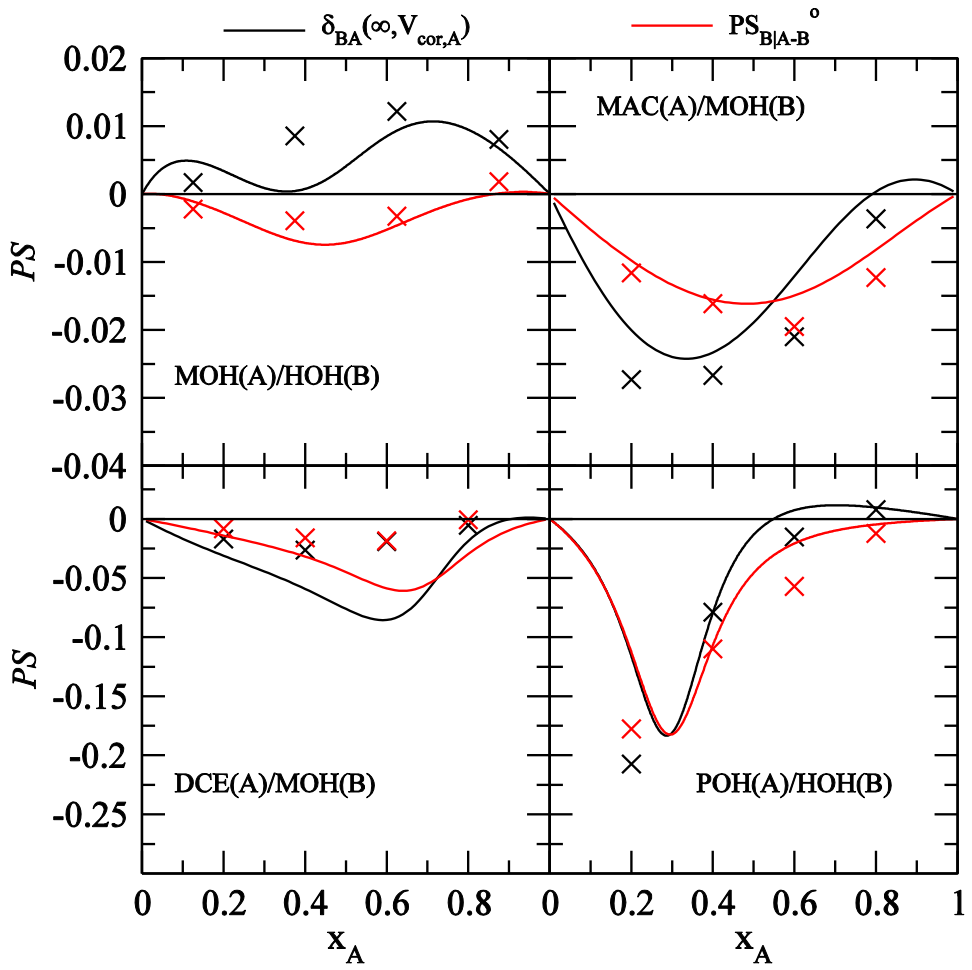

Figure 4.3 Experimental (solid lines) and simulated (symbols) *PS* measures as a function of composition for four mixtures. The values of $PS^{o}_{B|A-B}$ have been scaled down by a factor of 100 for ease of comparison

the trends observed in the experimental *PS* measures with composition. All but the MOH/HOH mixtures indicated negative *PS* values for most compositions. The two *PS* measures differed in sign for the MOH/HOH system, and this difference was reproduced by the simulation data. Hence, when a system displays relatively small deviations from ideality the definition of *PS* adopted for the analysis can affect the sign of the results. This did not occur with the new measure described

above as both the sign and the magnitude of the *PS* are closely related to the sign and magnitude of the excess molar Gibbs free energy; Equation (4.34).

The data in the previous figures only involved quantities related to the fully integrated KBIs. Using the simulation data one can also obtain the partially integrated equivalents. Selected examples are provided in Figure 4.4 and Figure 4.5. In Figure 4.4 the results for the POH/HOH

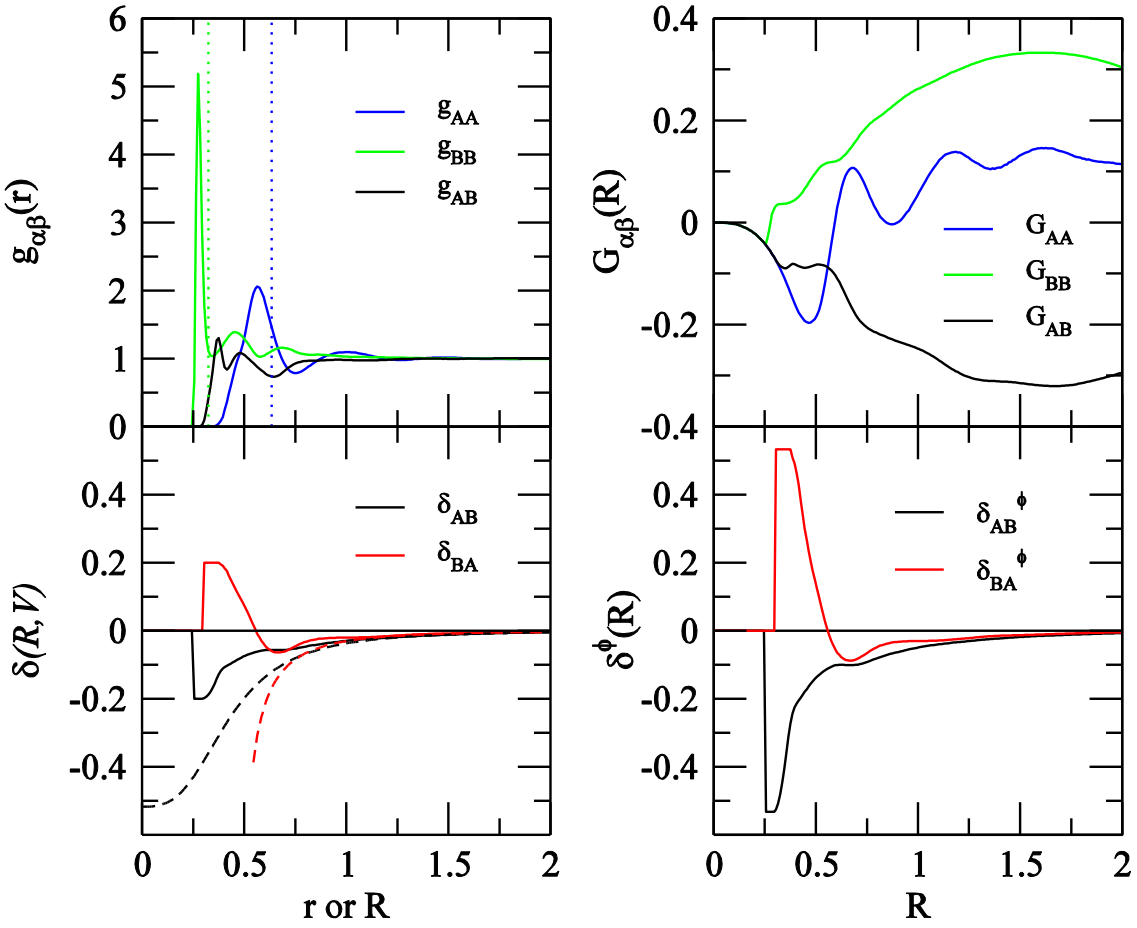

Figure 4.4 Center of mass based radial distribution functions (top left), KBIs (top right), and distance dependent *PS* measures (bottom) for a simulated mixture of isopropanol (*A*) and water (*B*) at an alcohol mole fraction of 0.2. KBIs are in L/mol and distances are in nm. Vertical dotted lines in the top left panel correspond to the correlation volume radii obtained from Equations (4.15) and (4.16) using the simulated KBIs. Dashed lines in the bottom left panel correspond to the *PS* measures, $\delta(\infty,V)$, provided by Equation (4.15) using the simulation data.

system at an alcohol mole fraction of 0.2 are displayed. At this composition there were significant deviations in the local composition. Analysis of the rdfs indicated a series of solvation shells surrounding each central molecule (as expected). The enhanced first shell peaks for the POH-POH and HOH-HOH rdfs suggested an increase in the tendency to self-associate at this composition. This was accompanied by rather small magnitude solvation shells in the POH-HOH rdf. For comparison, we have included the first shell correlation volume radii obtained from the iterative

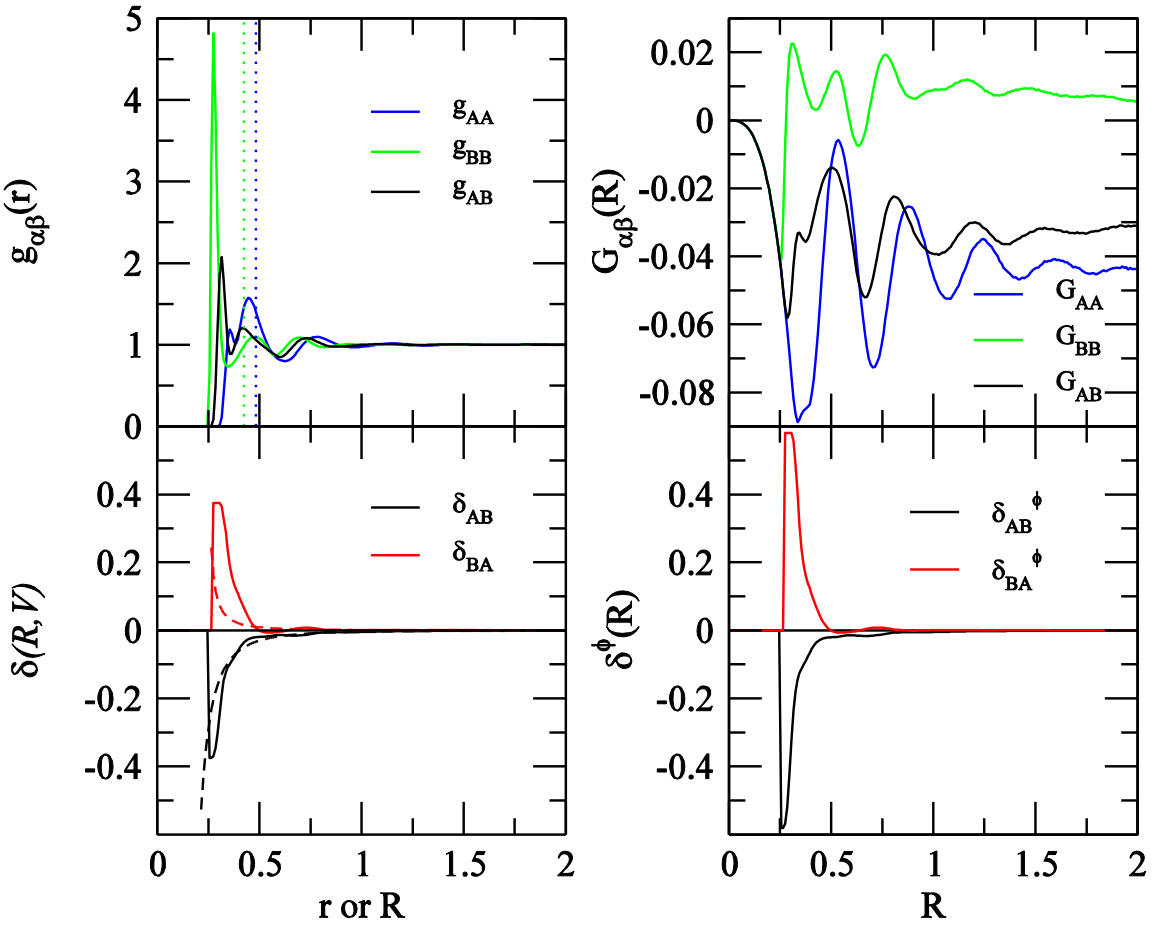

Figure 4.5 Center of mass based radial distribution functions (top left), KBIs (top right), and distance dependent *PS* measures (bottom) for a simulated mixture of methanol (*A*) and water (*B*) at an alcohol mole fraction of 0.375. KBIs are in L/mol and distances are in nm. Vertical dotted lines in the top left panel correspond to the correlation volume radii obtained from Equations (4.15) and (4.16) using the simulated KBIs. Dashed lines in the bottom left panel correspond to the *PS* measures, $\delta(\infty,V)$, provided by Equation (4.15) using the simulation data

solution of Equations (4.15) and (4.16). The HOH radius matched well with the observed first minimum in the HOH-HOH rdf, but at this distance the simulations also suggested the solvation shell contained essentially no POH molecules. The POH radius was larger and also close to the first minimum in the POH-POH rdf. Clearly, the first correlation volume radius cannot coincide with the first minima displayed by multiple rdfs. Indeed, the correlation volume radius is defined as a weighted mean involving the distribution and size of all the possible species in solution.[15] Nevertheless, the very short radii suggested for a central water molecule did not appear to agree with the simulation results presented here.

The partially integrated KBIs are also provided in Figure 4.4 and appeared to converge reasonably well beyond 1.5 nm certainly sufficient for the present analysis. The distance dependent δ's that form the basis of all the traditional *PS* measures are displayed in Figure 4.4. At short distances there was a clear excluded volume effect where the smallest molecule (in this case HOH, *B*), dominated the composition at the contact distance. Hence, $x_{AB}^{L} = 0$ and therefore δ_{AB} = - x_A, while $x_{BA}^{L} = 1$ and therefore δ_{BA} = 1 - x_B, at small distances. The same was true for the volume fraction based measures where $\phi_A \approx 0.5$. After this initial region (0.7 nm) the δ values were consistently negative and rise slowly with distance to zero. At intermediate distances the value of δ changed sign in some cases. The results obtained using Equation (4.15) are also presented in Figure 4.4 (and Figure 4.5). As expected, the results displayed a consistent sign and tended to the results obtained using Equation (4.11). This occurred at distances larger than 1 nm suggesting the differences between the KBIs was negligible beyond this distance (see Figure 4.4 and Figure 4.5).

Similar data for the MOH/HOH mixture at a methanol mole fraction of 0.375 is provided in Figure 4.5. Here the measures of *PS* beyond the excluded volume distance were much more subtle and resulted in δ values that were small, oscillated in sign with distance, and appeared

shorter ranged in nature. The first shell correlation volume radii obtained from the iterative solution of Equations (4.15) and (4.16) are also included in Figure 4.5. There appeared to be little if any relationship between the correlation volume radii and the solvation shells indicated by the rdfs. Furthermore, the correlation volume around water in this system was substantially larger than that for the POH/HOH systems, even though POH has a larger volume than MOH.

In Figure 4.6 we compare and contrast the results provided by Equations (4.11) and (4.15)

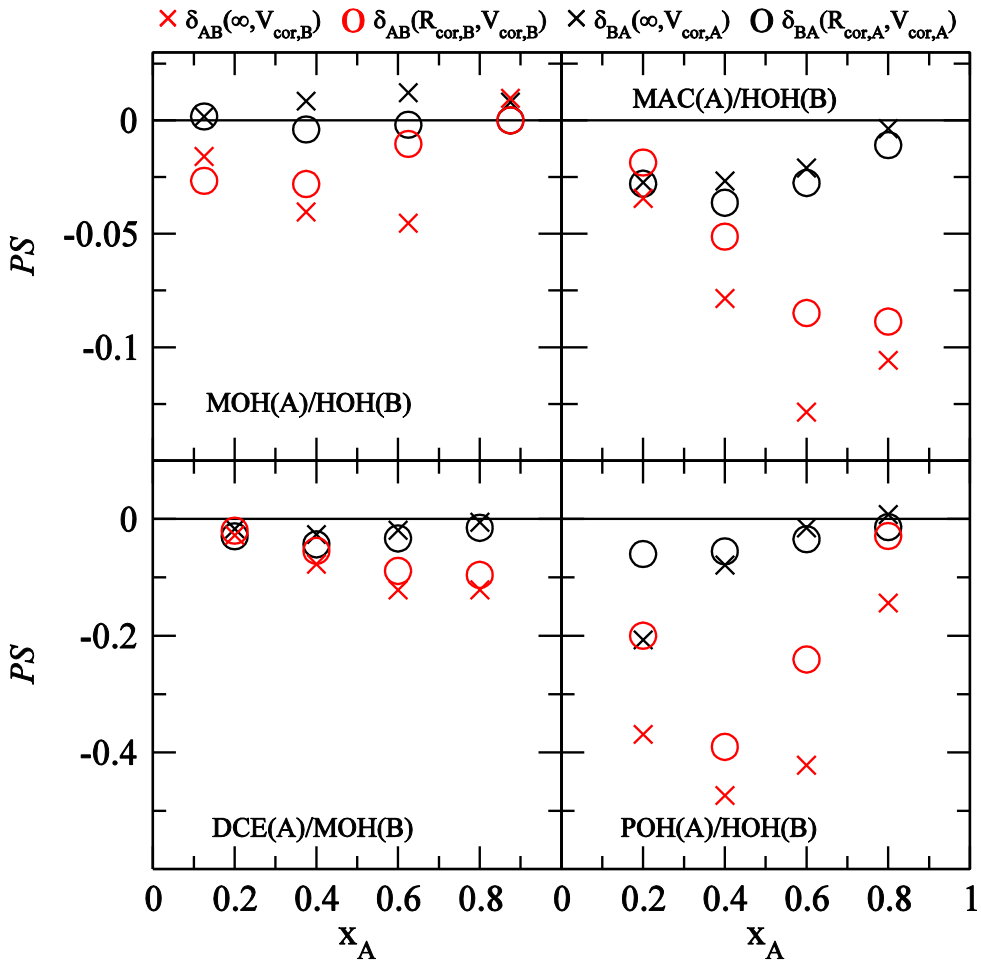

Figure 4.6 Simulated measures of *PS* corresponding to the first solvation shell for four binary mixtures. The simulated values of $\delta_{BA}(\infty,V_{cor,A})$ obtained using Equation (4.15) and Equation (4.16) are compared to simulated values of $\delta_{BA}(R_{cor,A},V_{cor,A})$ as provided by Equation (4.11)

using the simulated integrals. This directly examines the approximation that $\delta_{BA}(\infty, V_{cor,A}) = \delta_{BA}(R_{cor,A}, V_{cor,A})$. It was observed that, for the systems with strong *PS* measures, the two approaches gave the same sign and were semi-quantitative in nature. The exception was the MOH/HOH mixtures where the *PS* measures were more subtle and opposite signs were observed for the different δ values. Consequently, it appears that for many systems the questionable approximation used in developing Equation (4.15) might actually be reasonable. Unfortunately, without the additional insight from simulation one will never know for sure exactly how good the approximation might be for a particular system of interest.

It may appear somewhat strange that the results of the solvation shell approach, obtained assuming fully integrated KBIs, appeared to agree reasonably well with the results obtained using the partially integrated KBIs. A possible explanation is provided in Figure 4.7. Here, one observes a reasonable linear correlation between the fully integrated KBIs and the corresponding partially integrated quantities for all the systems, with the possible exception of the water-water KBIs in the POH/HOH system. The correlation was observed whether one integrates to the correlation volume radius or the first minimum in the corresponding rdf. Similar results for other systems have appeared previously.[12] Assuming a simple linear relationship between the KBIs such that $G_{\alpha\beta}(\infty) = aG_{\alpha\beta}(R_{cor,\alpha}) + c$ suggests that,

$$\delta_{BA}(R_{cor,A}, V_{cor,A}) = x_A x_B \frac{G_{AB}(\infty) - G_{AA}(\infty)}{V_{cor,A}^{app} + x_A G_{AA}(\infty) + x_B G_{AB}(\infty)} = \delta_{BA}(\infty, V_{cor,A}^{app}) \quad (4.39)$$

where $V_{cor,A}^{app} = aV_{cor,A} - c$ can be considered an apparent correlation volume. Consequently, under these conditions the expressions provided in Equations (4.11) and (4.15) are almost identical, differing only in the precise value of the correlation volume.

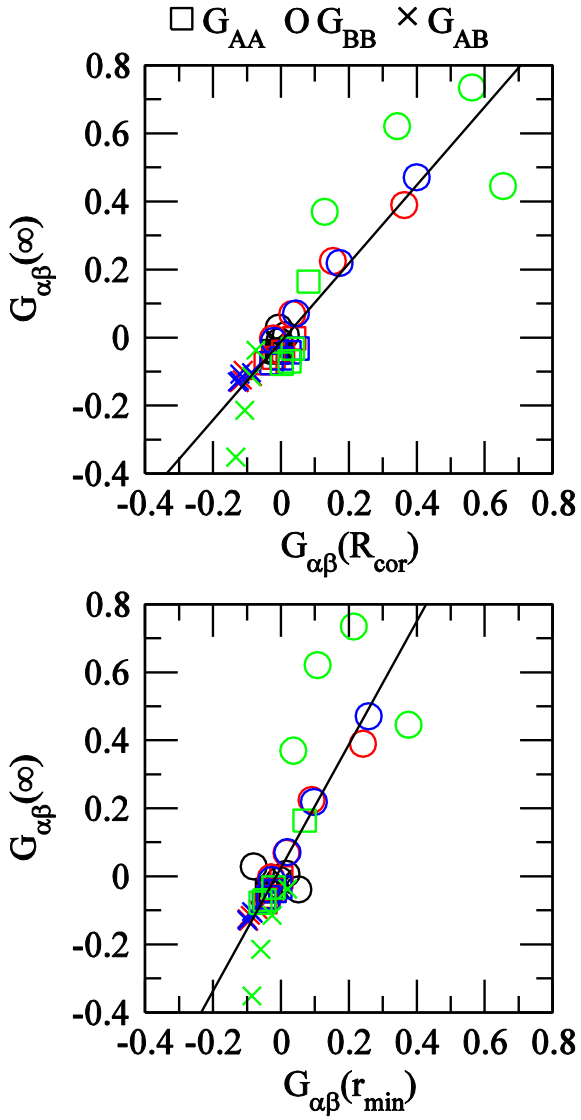

Figure 4.7 A comparison of partially integrated KBIs (L/mol) with the corresponding fully integrated values obtained from the simulation of four mixtures. Partial integration was performed to the correlation volume radius ($R_{,cor}$), as provided by Equations (4.15) and (4.16), and to the first major minimum in the corresponding rdf (r_{min}). The results include all compositions and are color coded according to the particular mixture: MOH(*A*)/HOH(*B*) in black; MAC(*A*)/MOH(*B*) in red; DCE(*A*)/MOH(*B*) in blue; and POH(*A*)/HOH(*B*) in green.

The distance dependent behavior of the new PS measure obtained from the simulations is presented in Figure 4.8. The values of $PS_{B|A-B}(R)$ describe the preference of water (B) for either POH or MOH molecules (A) compared to other water molecules as a function of distance. Both

mixtures displayed a negative PS beyond the excluded volume region, with the POH/HOH mixture values significantly larger in magnitude. The only exception was the MOH/HOH mixture at high MOH mole fractions. Here, the PS was small and positive and oscillated in sign with distance. The composition dependence of the PS measures obtained for various solvation shells are also displayed in Figure 4.8. The distances chosen corresponded to the first three solvation shells as described by the major minima in $g_{AA}(r)$; see Figure 4.4 and Figure 4.5. As expected, the measures decreased in magnitude as the solvation shell number increased, with the largest changes observed on going from the first to the second solvation shell. Interestingly, the minimum in the PS measures occurred close to the compositions where both components occupy equal volumes, i.e. at $x_A = 0.2$ for POH/HOH and $x_A = 0.3$ for MOH/HOH. However, the significance of this is unknown at present. The magnitude of the deviations from the bulk distribution at contact were significant for the POH/HOH system varying from 40% to 100% of the bulk water molarity on going from $x_A = 0.2$ to 0.8. The percentage changes at contact for the MOH/HOH mixture were generally less than 2% of the bulk water molarity.

The new measure of *PS* described here can also be applied to ternary mixtures. We have not provided examples here as the number of fully miscible ternary systems studied by KB related approaches is relatively low.[16] We do, however, expect similar results to the binary systems described above. The *PS* in ternary protein systems has also been the subject of much study,[11, 13, 42, 74-77] and the relationship between the new *PS* measure and the thermodynamics of proteins in mixed solvents is clearly outlined in the thermodynamics of ternary mixtures section. Equation (4.11) or (4.20) is not typically used in these application. Hence, the new measure of *PS* suggested here is related in a very simple manner, see Equation (4.36) for example, to the traditional measures of *PS* already in use.

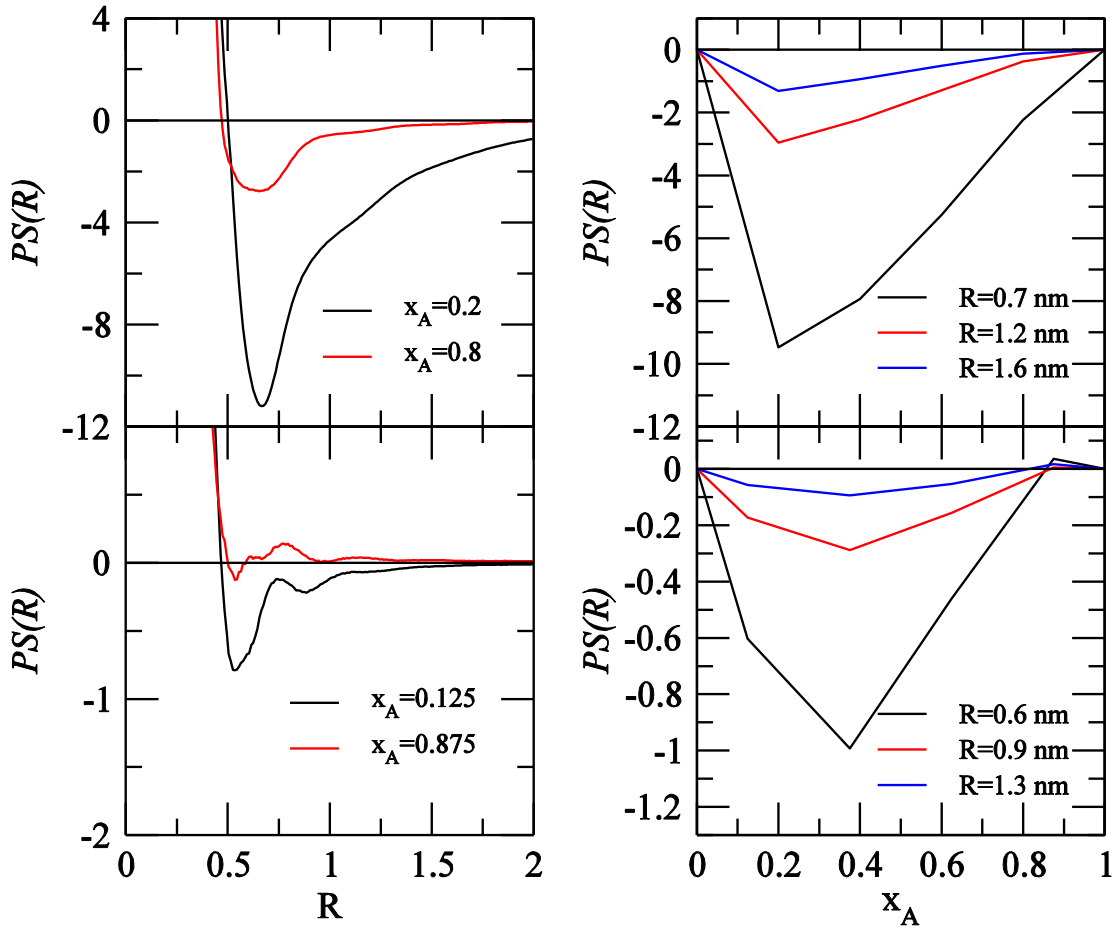

Figure 4.8 Simulated distance dependent preferential solvation measures, $PS_{B|A-B}(R)$ in mol/L, as given by Equations (4.23) and Equation (4.4). The top two panels refer to the POH(*A*)/HOH(*B*) system, while the bottom two panels refer to the MOH(*A*)/HOH(*B*) system. For comparison, the bulk water densities (ρ_B) were 26.9, 14.5, 7.5 and 3.1 mol/L at compositions of x_A = 0.2, 0.4, 0.6, and 0.8, respectively, for the POH/HOH system, with 42.1, 23.7, 11.7, and 3.3 mol/L at compositions of x_A = 0.125, 0.375, 0.625, and 0.875, respectively, for the MOH/HOH system.

4.6 Conclusions

A new measure of *PS* in binary and ternary mixtures has been presented. The new measure combines the rigor of the Ben-Naim limiting approach with the desired properties for ideal mixtures, the absence of which led to the development of the (criticized) volume corrected quantities. The main change from the Ben-Naim approach to *PS* is the focus on the difference in

the distribution of a *single* species around two *different* central molecules, rather than two *different* molecules around the *same* central molecule. The use of local volume fractions or densities also helps to simplify the expressions for ternary systems. The new measure can be easily related to the solution thermodynamics and thereby changes in the activity coefficients. Molecular dynamics simulations were used to investigate the nature of existing and new *PS* measures. It was observed that the partially and fully integrated KBIs are related, to a reasonable approximation, in a simple linear manner. Hence, it is plausible that existing approximate measures of *PS* provide realistic semi-quantitative data concerning the original precise measures. However, for truly accurate work the spatial dependence of a *PS* measure will require simulation data as input.

4.7 References

1. Ben-Naim, A. *Molecular Theory of Solutions*. Oxford University Press: New York, 2006.
2. Grunwald, E.; Baughman, G.; Kohnstam, G. *J. Am. Chem. Soc.* **1960,** *82* (22), 5801-5811.
3. Ben-Naim, A. *Cell Biophys.* **1988,** *12*, 255-269.
4. Banerjee, D.; Laha, A. K.; Bagchi, S. *J. Chem. Soc., Fara. Trans.* **1995,** *91* (4), 631-636.
5. Hunter, C. A. *Angew. Chem. Int. Ed. Engl.* **2004,** *43* (40), 5310-5324.
6. Zielkiewicz, J. *Phys. Chem. Chem. Phys.* **2000,** *2* (13), 2925-2932.
7. 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.
8. Amenta, V.; Cook, J. L.; Hunter, C. A.; Low, C. M. R.; Vinter, J. G. *J. Phys. Chem. B* **2012,** *116* (49), 14433-14440.
9. Chatterjee, P.; Laha, A. K.; Bagchi, S. *J. Chem. Soc., Fara. Trans.* **1992,** *88* (12), 1675-1678.
10. Ben-Naim, A. *J. Chem. Phys.* **1975,** *63* (5), 2064-2073.
11. Shulgin, I. L.; Ruckenstein, E. *Biophys. Chem.* **2005,** *118* (2-3), 128-134.
12. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2001,** *105* (46), 11513-11522.
13. Smith, P. E. *J. Phys. Chem. B* **2004,** *108* (48), 18716-18724.
14. Smith, P. E.; Mazo, R. M. *J. Phys. Chem. B* **2008,** *112* (26), 7875-7884.
15. Marcus, Y. *J. Chem. Soc., Fara. Trans. I* **1989,** *85*, 3019-3032.
16. Matteoli, E.; Lepori, L. *J. Chem. Soc., Fara. Trans.* **1995,** *91* (3), 431-436.
17. Jimenez, D. M.; Cardenas, Z. J.; Delgado, D. R.; Pena, M. A.; Martinez, F. *Fluid Phase Equilib.* **2015,** *397*, 26-36.
18. Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951,** *19* (6), 774-777.
19. Ben-Naim, A. *J. Chem. Phys.* **1977,** *67* (11), 4884-4890.
20. Ploetz, E. A.; Smith, P. E. *Adv. Chem. Phys.* **2013,** *153*, 311-372.
21. Ben-Naim, A. *J. Phys. Chem.* **1989,** *93* (9), 3809-3813.

22. Ben-Naim, A. *Pure Appl. Chem.* **1990,** *62* (1), 25-34.

23. Ploetz, E. A.; Smith, P. E. *J. Phys. Chem. B* **2015,** *119* (25), 7761-7777.

24. Ben-Naim, A. Global and Local Properties of Mixtures: An Expanded Paradigm for the Study of Mixtures. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics.*, Smith, P. E.; Matteoli, E.; O' Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 35-64.

25. Hall, D. G. *Trans. Faraday Soc.* **1971,** *67* (585), 2516-2524.

26. Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2002,** *106* (6), 1491-1500.

27. Smith, P. E. *Biophys. J.* **2006,** *91* (3), 849-856.

28. Smith, P. E. *J. Chem. Phys.* **2008,** *129* (12), 124509.

29. Karunaweera, S.; Gee, M. B.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2012,** *8* (10), 3493-3503.

30. Matteoli, E.; Gianni, P.; Lepori, L. Kirkwood-Buff Integrals in Fully Miscible Ternary Systems: Thermodynamic Data, Calculation, Representation, and Interpretation. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering, and Biophysics*, Smith, P. E.; Matteoli, E.; O' Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 93-116.

31. Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *J. Chem. Phys.* **2010,** *132* (16), 164501.

32. Lepori, L.; Matteoli, E. *J. Phys. Chem.* **1988,** *92* (24), 6997-7001.

33. Matteoli, E. *J. Phys. Chem. B* **1997,** *101* (47), 9800-9810.

34. Marcus, Y. *Monatsh. Chem.* **2001,** *132* (11), 1387-1411.

35. Shulgin, I.; Ruckenstein, E. *J. Phys. Chem. B* **1999,** *103* (13), 2496-2503.

36. Ben-Naim, A. *J. Phys. Chem. B* **2007,** *111* (11), 2896-2902.

37. Matteoli, E.; Lepori, L. *J. Phys. Chem. B* **2007,** *111* (11), 3069-3071.

38. Ben-Naim, A. *J. Phys. Chem. B* **2007,** *111* (11), 3072-3072.

39. Mazo, R. M. *J. Chem. Phys.* **2008,** *129* (15), 154101.

40. Mazo, R. M.; Matteoli, E.; Smith, P. E. *J. Chem. Phys.* **2009,** *130* (23), 234508.

41. Ploetz, E. A.; Smith, P. E. *J. Chem. Phys.* **2015,** *142*, 094504.

42. Shimizu, S. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (5), 1195-1199.

43. Shulgin, I. L.; Ruckenstein, E. *J. Chem. Phys.* **2005,** *123* (5), 054909.

44. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochem. Biophys.* **2008,** *50* (1), 1-22.

45. Anderson, C. F.; Courtenay, E. S.; Record, M. T. *J. Phys. Chem. B* **2002,** *106* (2), 418-433.

46. Schurr, J. M.; Rangel, D. P.; Aragon, S. R. *Biophys. J.* **2005,** *89* (4), 2258-2276.

47. Smith, P. E. *J. Phys. Chem. B* **2006,** *110* (6), 2862-2868.

48. Smith, P. E.; Matteoli, E.; O' Connell, J. P. Fluctuation Solution Theory: A Primer. In *Fluctuation Theory of Solution: Applications in Chemistry, Chemical Engineering and Biophysics*, Smith, P. E.; Matteoli, E.; O' Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 1-34.

49. Butler, J. A. V.; Thomson, D. W.; Maclennan, W. H. *J. Chem. Soc.* **1933**, 674-686.

50. Nagata, I.; Ohta, T.; Takahashi, T. *J. Chem. Eng. Jpn.* **1972,** *5* (3), 227-231.

51. Chaudhari, S. K.; Katti, S. S. *Fluid Phase Equilib.* **1989,** *50* (3), 329-338.

52. Hu, J. H.; Haynes, C. A.; Wu, A. H. Y.; Cheung, C. M. W.; Chen, M. M.; Yee, E. G. M.; Ichioka, T.; Nishikawa, K.; Westh, P.; Koga, Y. *Can. J. Chem.* **2003,** *81* (2), 141-149.

53. Douheret, G.; Khadir, A.; Pal, A. *Thermochim. Acta* **1989,** *142* (2), 219-243.

54. Gonzalez, B.; Calvar, N.; Gomez, E.; Dominguez, A. *J. Chem. Thermodyn.* **2007,** *39* (12), 1578-1588.

55. Prakash, D. J.; Lakshmi, D. S.; Rao, M. V.; Prasad, D. H. L. *Phys. Chem. Liq.* **1996,** *33* (4), 249-254.

56. Washburn, E. W. *International Critical Tables of Numerical Data, Physics, Chemistry and Technology*. Knovel: Online, 2003; Vol. 3.

57. Matteoli, E.; Lepori, L. *J. Chem. Phys.* **1984,** *80* (6), 2856-2863.

58. Easteal, A. J.; Woolf, L. A. *J. Chem. Thermodyn.* **1985,** *17* (1), 49-62.

59. Oswal, S. L.; Oswal, P.; Modi, P. S.; Dave, J. P.; Gardas, R. L. *Thermochim. Acta* **2004,** *410* (1-2), 1-14.

60. Findenegg, G. H.; Kohler, F. *Trans. Faraday Soc.* **1967,** *63* (532P), 870-878.

61. Katti, P. K.; Shil, S. K. *J. Chem. Eng. Data* **1966,** *11* (4), 601-604.

62. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013,** *29* (7), 845-54.

63. Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005,** *109* (31), 15080-15086.

64. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

65. Mobley, D. L.; Liu, S.; Cerutti, D. S.; Swope, W. C.; Rice, J. E. *J. Comput. Aided Mol. Des.* **2012,** *26* (5), 551-562.

66. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

67. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992,** *13* (8), 952-962.

68. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

69. Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007,** *126* (1), 014101.

70. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

71. Abildskov, J.; Wedberg, R.; O' Connell, J. P. Fluctuation Solution Theory Properties from Molecular Simulation. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics*, Smith, P. E.; Matteoli, E.; O' Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 133-162.

72. Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilib.* **2010,** *290* (1-2), 43-47.

73. Ploetz, E. A.; Weerasinghe, S.; Kang, M.; Smith, P. E. Accurate Force Fields for Molecular Simulation. In *Fluctuation Theory of Solutions: Applications in Chemistry, Chemical Engineering and Biophysics* Smith, P. E.; Matteoli, E.; O' Connell, J. P., Eds. CRC Press: Boca Raton, 2013; pp 117-132.

74. Vagenende, V.; Yap, M. G. S.; Trout, B. L. *Biochemistry* **2009,** *48* (46), 11084-11096.

75. Auton, M.; Bolen, D. W.; Rosgen, J. *Proteins: Struct. Funct. Bioinform.* **2008,** *73* (4), 802-813.

76. Shukla, D.; Trout, B. L. *J. Phys. Chem. B* **2011,** *115* (41), 11831-11839.

77. Shimizu, S. *Chem. Phys. Lett.* **2011,** *514* (1-3), 156-158.

Chapter 5 - An Experimental Investigation of the Kirkwood Superposition Approximation for Fluid Water

5.1 Abstract

A combination of Fluctuation Solution Theory and experimental pair radial distribution functions are used to investigate the accuracy of the Kirkwood Superposition Approximation, as given by the integrals over the pair and triplet distribution functions, at a series of state points for pure water. A variety of additional approximate relationships between the pair and triplet correlations in fluids are also investigated and generally provide good agreement with the fluid thermodynamic results for regions of the phase diagram where the compressibility is small. A simple power law relationship between the pair and triplet fluctuations is observed for low to moderately high compressibilities.

5.2 Introduction

Statistical theories of fluids attempt to provide a rigorous link between the macroscopic thermodynamics properties and the microscopic molecular properties of fluids.[1-4] The distribution of molecules in space characterize the main differences between solid, liquid and gas states.[3] However, the characterization of liquids and liquid solution mixtures is much more complicated due to the strong intermolecular interactions and random motions of the molecules.[3] Liquids and liquid solution mixtures are typically characterized by relative probability distribution functions or correlation functions.[2-4] These n body distribution functions provide an approach to relate the structure to the thermodynamic properties.[2, 4]

Most of the theoretical treatments of liquids involve the n body distribution functions, $g^{n}_{\alpha\beta\ldots}(r_1 r_2 \ldots r_n)$.[1-2, 5] Although, the two body distribution functions can be obtained from scattering studies, information regarding the triplet or higher distribution functions are not readily available.[5-7] The main solution to this problem is the Kirkwood Superposition Approximation (KSA). This is an important approximation that appears in many liquid state theories and relates the two body distribution function to three body and higher order distribution functions.[2] For the three body distribution function this assumption indicates that the probability of finding three particles in a particular arrangement can be obtained from the product of individual pairwise probabilities.[8] For instance, Equation (5.1) indicates the relationship between three body distribution function $g^{(3)}_{111}$ and two body distribution function ($g^{(2)}_{11}$) via the KSA for a pure liquid,[8]

$$g^{(3)}_{111}(r_1, r_2, r_3) = g^{(2)}_{11}(r_1, r_2) g^{(2)}_{11}(r_1, r_3) g^{(2)}_{11}(r_2, r_3) \tag{5.1}$$

where r_1, r_2, r_3 are the positions of the particle one, two and three, respectively.

In general, the knowledge of one distribution function requires the knowledge of the other lower distribution functions. Therefore, with the intention of obtaining a link between n body distribution functions, Kirkwood proposed the superposition approximation in 1935.[9] In other words, the spirit of the KSA is that all the higher order distribution functions can be expressed in terms of the pair distribution function.[10] On the other hand, the physical meaning of the KSA can be further expressed using the potential of mean force.[10-12] For example, if we consider three particles located at positions r_1, r_2 and r_3, it is possible to define the mean force acting on the third particle (at position r_3) resulting from the other two particles at positions r_1 and r_2. This mean force is equal to the sum of the forces obtained if particles 1 and 2 affected particle 3 independently, as shown in Equation (5.2).[12-13]

$$W^{(3)}(r_1, r_2, r_3) = W^{(2)}(r_1, r_2) + W^{(2)}(r_1, r_3) + W^{(2)}(r_2, r_3) \quad (5.2)$$

where $W^{(n)}(r_1, r_2 \ldots . r_n)$ is the potential of mean force, i.e. the average force required to bring n particles to a particular configuration. The relationship between the potential of mean force and n body distribution functions is shown in Equation (5.3).

$$W^{(n)}(r_1, r_2 \ldots . r_n) = -k_B T ln g^{(n)}(r_1, r_2 \ldots . r_n) \quad (5.3)$$

where k_B is the Boltzmann constant.

The development of the theory of liquids is highly influenced by the KSA.[10] Integral equation methods, one of the most imperative and extremely attractive methods that has been used in the theoretical studies of liquids over years, have used the KSA to eliminate a chain of equations leading to much more simpler integral equations.[10,11] Consequently, the validity of the KSA has been studied for over 80 years.[10]

A large number of studies exist in the literature that have been conducted using various approaches to validate KSA.[10] Here, we will summarize some of the main studies and their conclusions. One of the most common approaches to study the KSA is based on the coefficients of the virial expansion for hard sphere gases.[10] Hart and coworkers have calculated virial coefficients, with and without the KSA, and obtained an error in the approximation.[14] Most of the studies have shown that the fourth virial coefficient gives 20-25% error.[14-15] However, using the KSA they could obtain a reasonable value for the third virial coefficient.[14] Boer has critically discussed theories of liquids and the inaccuracy of the KSA based on thermal and caloric quantity agreement in 1952.[16] Kirkwood and coworkers have studied the properties of fluid particles interacting according to a LJ potential using the KSA and compared with experimental data.[17]

They have concluded that the calculated equation of state is not in good agreement with the experimental data. The pressure of the system is overestimated and underestimated at small and high densities, respectively.[17] Therefore, they have concluded that the correlation obtained using the KSA is not suitable for all densities.[17] Several other workers have also discussed the validity of the KSA at different densities.[2] Moreover, the error of the KSA results in a significantly large error for the two and three body distribution function.[17] Salsburg and coworkers have studied the KSA in a strictly mathematical way and found that the approximation is accurate for one dimensional systems but it is impossible to extend this result to multi dimensional systems.[18]

Not only theoretical calculations but also number of molecular simulation approaches, such as Molecular Dynamics (MD) and Monte Carlo (MC) methods, have been used to study the KSA.[19-20] These methods also confirmed the low precision of the KSA. Moreover, these computational studies have investigated the validity of the KSA with respect to molecular distance. The KSA appears to be more accurate at shorter distances and inaccurate at large distances.[10] The work of Rowlinson and Alder has revealed that the KSA overestimates the quantity of particles when using the calculation for the triplet correlation function.[21-22] Moreover, Krumhansl and coworkers have stated that the KSA would result in a 15% error for the symmetric triplet configuration and a 35% for asymmetric ones.[23-24] Temperly and coworkers have stated that the KSA might be valid for very high and low densities, such as solid and gases, but not for the liquid state.[10] Egelstaff and coworkers have shown that the ternary distribution of solid spheres can be predicted reasonably accurately using the KSA. In contrast, Bildstein and coworkers have argued that the KSA is poor as they observed significant errors at direct contact using computer simulation studies.[25] Ben-Amotz and coworkers have shown that the KSA is least accurate near the contact separation.[26]

Dhabal and coworkers have recently studied the KSA approximation for liquid water under ambient conditions using MD and Reverse Monte Carlo (RMC) and showed that the triplet correlation from RMC datasets are in reasonable agreement with the KSA, but the data from MD simulations showed significant disagreement within the first two neighbor shells.[27] Piasecki and coworkers have revealed the absence of a critical point within the KSA.[28] Singer and coworkers have provided a different explanation of the KSA using a variational formulation for systems at equilibrium in the thermodynamic limit.[29] In this work they have used a maximum entropy formulation of the KSA.[29] They have shown that the KSA fails when the three particles are very close to each other.[29] Furthermore, a study based on a colloidal model liquid also indicated that the KSA would fail for strongly interacting systems.[30] In summary, extensive theoretical and simulations studies have been performed to examine the KSA using various model systems. However, the primary motivation for this study is the lack of information concerning the validity of the KSA for real liquids using experimental data.

In this study, we investigate a different approach to validate the KSA for pure water based on Fluctuation Solution Theory (FST). Fluctuation Solution Theory is an important theory of liquids which provides a rigorous link between particle number fluctuations and common thermodynamic properties.[31-32] Moreover, the particle number fluctuations can be expressed using two and higher body probability distribution functions.[31-32] Recently, Smith and coworkers have proposed an approach to link particle number fluctuations to higher order probability distribution functions.[32-34] In other words, using FST it is possible to obtain higher order distribution functions using common thermodynamic properties. There is scarce experimental information available regarding the distribution functions beyond the pair correlation function.[33] Therefore, it is beneficial to have an approach to obtain these higher order distribution functions from

thermodynamic data for liquids and liquid solution mixtures. Hence, this provide a route to validate the KSA using just experimental data for real liquids.

Buff and coworkers have used this type of approach to validate the KSA for liquid argon.[35] We have followed their approach to further explore the KSA for pure water at a series of state points. We use experimental pair distribution functions obtained from neutron and x-ray scattering studies in combination with FST to investigate the KSA. Triplet and pair distributions are then related using a variety of additional approximations.

5.3 Theory

Recently, we applied FST to study the properties of pure liquids.[33] In this approach the properties of the liquid are expressed in terms of the particle fluctuations for an equivalent system open to matter exchange. The main fluctuating quantities of interest here are provided by,[33]

$$b_{11} \equiv \frac{\langle (\delta N_1)^2 \rangle}{\langle N_1 \rangle} = 1 + \rho_1 G_{11} \tag{5.4}$$

$$c_{111} \equiv \frac{\langle (\delta N_1)^3 \rangle}{\langle N_1 \rangle} = 1 + 3\rho_1 G_{11} + \rho_1^2 G_{111} \tag{5.5}$$

$$d_{1111} = \frac{\langle (\delta N_1)^4 \rangle - 3\langle \delta {N_1}^2 \rangle^2}{\langle N_1 \rangle} = 1 + 7\rho_1 G_{11} + 6\rho_1^2 G_{111} + \rho_1^3 G_{1111} \tag{5.6}$$

where $\rho_1 = \langle N_1 \rangle / V$ is the average number density, V is volume, $\delta N_1 = N_1 - \langle N_1 \rangle$ denotes a fluctuation in the value of the instantaneous number of molecules N_1, and the angular brackets denote an ensemble average for the Grand Canonical Ensemble (GCE). The above fluctuating quantities are essentially the cumulants of the particle probability distribution for the equivalent

GCE. The second equalities in Equation (5.4) - Equation (5.6) express these fluctuations in terms of integrals over *n*-body distribution functions $g^{(n)}_{\alpha\beta\ldots}(r_1, r_2, \ldots, r_n)$ according to,[33]

$$G_{11} \equiv V^{-1} \int [g^{(2)}_{11} - 1] dr_1\, dr_2 \tag{5.7}$$

$$G_{111} \equiv V^{-1} \int [g^{(3)}_{111} - 1 - 3(g^{(2)}_{11} - 1)] dr_1\, dr_2 dr_3 \tag{5.8}$$

$$G_{1111} \equiv V^{-1} \int [g^{(4)}_{1111} - 1 - 4(g^{(3)}_{111} - 1) - 3(g^{(2)}_{11} - 1)(g^{(2)}_{11} - 1) + 6(g^{(2)}_{11} - 1)] dr_1\, dr_2 dr_3 dr_4 \tag{5.9}$$

Hence, the above equations relate the thermodynamics of the fluid to integrals over the corresponding distribution functions that describe the structure of the liquid. Note that the molecular orientations do not appear in these integrals as the distribution functions correspond to those between the molecular centers of mass after averaging over their molecular orientations, and after averaging over the positions and orientations of all other molecules in the system.

The above fluctuations can be expressed in terms of pressure derivatives of the density according to,[34]

$$b_{11} = k_B T \rho_1' = \rho_1 k_B T \kappa_T \tag{5.10}$$

$$c_{111} = (k_B T)^2 [\rho_1 \rho_1'' + (\rho_1')^2] \tag{5.11}$$

$$d_{1111} = (k_B T)^3 [\rho_1^2 \rho_1''' + 4\rho_1 \rho_1' \rho_1'' + (\rho_1')^3] \tag{5.12}$$

where the prime indicates an isothermal derivative with respect to pressure (p), T the absolute temperature, and κ_T the isothermal compressibility. Consequently, if the density derivatives are known – usually from an accurate Equation of State – then experimental values for the fluctuations can be determined, and subsequently provide the integrals described in Equation (5.7) - Equation (5.9). The fluctuations can also be expressed in terms of integrals over the analogous direct correlation functions, although we shall not use that approach here.

The above integrals can be used in pressure or density expansions for a high density fluid. The pressure derivatives of the above integrals are given by,[33]

$$G'_{11} = \beta(G_{111} - 2G_{11}^2) \tag{5.13}$$

$$G''_{11} = \beta^2(G_{1111} - 7G_{111}G_{11} + 8G_{11}^3) \tag{5.14}$$

$$G'_{111} = \beta(G_{1111} - 3G_{111}G_{11}) \tag{5.15}$$

where $\beta = 1/k_BT$. The pressure derivatives are clearly related to integrals over higher distribution functions for the fluid. The above expressions correspond to the integrated versions of the well-known relationships between the probability density distribution functions,[5, 36]

$$\begin{aligned} k_BT\left(\frac{\partial[\rho_1^n g^{(n)}]}{\partial p}\right)_T &= \rho_1 k_BT\kappa_T\left(\frac{\partial[\rho_1^n g^{(n)}]}{\partial \rho_1}\right)_T \\ &= n\rho_1^{n-1}g^{(n)} + \rho_1^n\int[g^{(n+1)} - g^{(n)}]dr_{n+1} \end{aligned} \tag{5.16}$$

which can also be written in terms of just the spatial distribution functions,

$$k_BT\left(\frac{\partial g^{(n)}}{\partial p}\right)_T = \rho_1 k_BT\kappa_T\left(\frac{\partial g^{(n)}}{\partial \rho_1}\right)_T = \int[g^{(n+1)} - g^{(n)}\left(ng^{(2)} - n + 1\right)]dr_{n+1} \tag{5.17}$$

where the integration is over the molecular positions.

The expressions in the previous sections indicate that integrals over the pair, triplet and quadruplet distributions can be obtained from experiment. This provides a route for testing the KSA using purely experimental data for real liquids, as established by Buff and Brout.[35] Here we extend their approach. The accuracy of the KSA can be characterized by the difference between the triplet and quadruplet correlations and their pairwise analogues. Hence, we define,

$$\Delta g_{111}^{(3)} \equiv g_{11}^{(2)}(r_1, r_2) g_{11}^{(2)}(r_1, r_3) g_{11}^{(2)}(r_2, r_3) - g_{111}^{(3)}(r_1, r_2, r_3) \tag{5.18}$$

$$\begin{aligned} \Delta g_{1111}^{(4)} \equiv\ & g_{11}^{(2)}(r_1, r_2) g_{11}^{(2)}(r_1, r_3) g_{11}^{(2)}(r_1, r_4) g_{11}^{(2)}(r_2, r_3) g_{11}^{(2)}(r_2, r_4) g_{11}^{(2)}(r_3, r_4) \\ & - g_{1111}^{(4)}(r_1, r_2, r_3, r_4) \end{aligned} \tag{5.19}$$

where the expressions would be zero when the KSA, and the analogous four body approximations, hold, respectively. The comparison is facilitated by using the same functional form for the virial coefficients used to describe imperfect gases or osmotic solutions.[33] The first few virial coefficients we require are given by the expressions,[33]

$$B_2^* = -\frac{1}{2} G_{11} \tag{5.20}$$

$$B_3^* = -\frac{1}{3} [G_{111} - 3G_{11}^2] \tag{5.21}$$

$$B_4^* = -\frac{1}{8} [G_{1111} - 12 G_{111} G_{11} + 20 G_{11}^3] \tag{5.22}$$

where we have used an asterisk to indicate these expressions are valid for any density and are therefore different from the usual low density applications for these expressions. Using the same manipulations as performed for the traditional virial coefficients, the third virial coefficient using the KSA can then be written,

$$B^*_{3,KSA} = -\frac{1}{3}I_0 = -\frac{1}{3}V^{-1}\int h^{(2)}_{11}(r_1,r_2)\, h^{(2)}_{11}(r_1,r_3)h^{(2)}_{11}(r_2,r_3)dr_1dr_2dr_3 \qquad (5.23)$$

in terms of the total correlation function $h^{(2)}_{11} = g^{(2)}_{11} - 1$. Hence, we can evaluate the difference between the real triplet correlations and the KSA via,

$$\Delta G_{111} \equiv G^{KSA}_{111} - G^{Exp}_{111} \equiv V^{-1}\int \Delta g^{(3)}_{111}\, dr_1dr_2dr_3 = -3\left(B^*_{3,KSA} - B^*_3\right) \qquad (5.24)$$

The above integrated quantity provides a single measure concerning the accuracy of the KSA. It does not provide information concerning the accuracy as a function of distance, and clearly there may be some cancellation upon integration. Nevertheless, this approach avoids the necessity for theoretical or simulation data and therefore also avoids the assumption that these approaches are sufficiently accurate for analysis.

To evaluate the third virial coefficient using the KSA one requires additional experimental data; specifically, the experimental pair distribution as a function of distance. This can be obtained from the experimental structure factor, $S(k) = 1 + \rho_1 H_{11}(k)$, or the total correlation function, via Fourier transforms valid for isotropic liquids,

$$H_{11}(k) = \int_0^\infty h^{(2)}_{11}(r)\sin(kr)kr^{-1}4\pi r^2 dr \qquad (5.25)$$

$$h_{11}^{(2)}(r) = (2\pi)^{-3} \int_0^{\infty} H_{11}(k) \sin(kr)(kr)^{-1} 4\pi k^2 dk \tag{5.26}$$

The KSA expression for the third virial coefficient is then given by the fact that,[35]

$$I_0 = (2\pi)^{-3} \int_0^{\infty} [H_{11}(k)]^3 4\pi k^2 dk \tag{5.27}$$

via the convolution-correlation theorem.[37-38] Hence, one can compare the third virial coefficient obtained experimentally from the thermodynamic data to that obtained via the KSA using only pair correlations to determine the integrated difference indicated in Equation (5.24).

We can also investigate the severity of the KSA for four body correlations. A simple way to evaluate this is to determine the pressure dependence of the KSA approximation for the three body correlations. To achieve this we define,

$$\Delta G_{1111} \equiv G_{1111}^{KSA} - G_{1111}^{Exp} \equiv V^{-1} \int [\Delta g_{1111}^{(4)} - 4\Delta g_{111}^{(3)}] \, dr_1 dr_2 dr_3 dr_4 \tag{5.28}$$

Then, if we take pressure derivatives of G_{111} for the experimental and KSA approximations given by the expressions provided in Equation (5.13) – Equation (5.15) one finds,

$$\Delta G_{1111} = k_B T \left(\frac{\partial \Delta G_{111}}{\partial p} \right)_T + 3\Delta G_{111} G_{11} \tag{5.29}$$

which provides access to ΔG_{1111}. The difference between the real and KSA approximations for the four (and three) body correlations is related to the difference in the fourth virial coefficients via,

$$\Delta G_{1111} - 12\Delta G_{111} G_{11} = -8(B_{4,KSA}^* - B_4^*) \tag{5.30}$$

Unfortunately, the fourth virial coefficient under the KSA approximations does not simplify as easily as the third virial coefficient. Indeed, the fourth virial coefficient involves several terms,

$$B_{4,KSA}^{*} = -\frac{3}{8}I_1 - \frac{6}{8}I_2 - \frac{1}{8}I_3 \tag{5.31}$$

$$I_1 = V^{-1}\int [h_{11}^{(2)}(r_1,r_2)h_{11}^{(2)}(r_2,r_3)h_{11}^{(2)}(r_3,r_4)h_{11}^{(2)}(r_4,r_1)]\, dr_1 dr_2 dr_3 dr_4 \tag{5.32}$$

$$I_2 = V^{-1}\int \left[h_{11}^{(2)}(r_1,r_2)h_{11}^{(2)}(r_2,r_3)h_{11}^{(2)}(r_3,r_4)h_{11}^{(2)}(r_1,r_4)h_{11}^{(2)}(r_1,r_3)\right]\, dr_1 dr_2 dr_3 dr_4 \tag{5.33}$$

$$I_3 = V^{-1}\int \left[h_{11}^{(2)}(r_1,r_2)h_{11}^{(2)}(r_2,r_3)h_{11}^{(2)}(r_3,r_4)h_{11}^{(2)}(r_1,r_4)h_{11}^{(2)}(r_1,r_3)h_{11}^{(2)}(r_2,r_4)\right] dr_1 dr_2 dr_3 dr_4 \tag{5.34}$$

only one of which can be easily obtained from the experimental structure factor,

$$I_1 = (2\pi)^{-3}\int_0^{\infty} [H_{11}(k)]^4 4\pi k^2\, dk \tag{5.35}$$

via the convolution-correlation theorem.

5.4 Methods

The experimental thermodynamic data for pure water as a function of pressure and temperature were determined using the IAPWS-95 Equation of State, developed by Wagner and Pruss,[39] as implemented in the National Institute of Standards and Technology (NIST) Standard Reference Database 10: NIST/American Society of Mechanical Engineers Steam Properties Database version 2.22.[40] The source code provides the required first and second density derivatives as a function of pressure and temperature via a simple subroutine call. Third density derivatives

were obtained numerically via a finite difference approach using the second derivatives and a value of $dp = \pm 10^{-20}$ bar. Calculations were performed in quadruple precision.

The scattering data for liquid and fluid water were taken from the neutron scattering studies of Soper and coworkers.[41] The radial distribution functions for a variety of pressures and temperatures have been provided, after refinement using computer simulation data in an iterative procedure.[41] In addition, raw X-ray scattering data were also used to provide an indication of the effects of possible experimental errors.[42] The state points considered here are indicated in Figure 5.1, where we also include the values of b_{11} for the liquid and super critical regions.

Experimental scattering data suffer from technical issues at low and high scattering amplitudes.[43-47] Scattering at high amplitudes is a relatively unimportant contribution to the thermodynamic properties, however scattering at low amplitudes plays a significant role. Fortunately, the limiting value of $H_{11}(k)$ can be checked for consistency by noting that $S(0) = 1 + \rho_1 H_{11}(0) = \rho_1 k_B T \kappa_T = b_{11}$.[48] To ensure consistency between the thermodynamic and scattering data we have modified the low scattering behavior to obey the Ornstein-Zernike approximation – which has been demonstrated to hold even close to the critical point[49] – as given by,

$$S(k)/S(0) = 1 + ak^2 + bk^4 \tag{5.36}$$

The a and b parameters were obtained after fitting k values between 10-20 nm^{-1}. This resulted in very good fits to the experimental structure factors for all but the T = 673 K and p = 500 bar state point which lies closest to the critical point. The integrals described in Equation (5.25), Equation (5.26), Equation (5.27) and Equation (5.35) were determined using discrete Fourier transforms.

The rdfs extend to 1.5 nm using 500 observations (n_{pt}) with intervals (dr) of 0.003 nm which provides a maximum k value of 333 nm^{-1} and intermediate k values that satisfy $n_{pt}\, dr\, dk = 1$.

To determine the value of ΔG_{1111} we need to evaluate the derivative indicated in Equation (5.29). This was obtained after fitting the ΔG_{111} values to a simple polynomial,

$$\Delta G_{111} = c_0 + c_1 p + c_2 p^2 \tag{5.37}$$

This provides reasonable derivatives for points not too close to the critical point. In particular, the state points located at T = 673 K and p = 500 bar and T = 573 K and p = 100 bar were dropped to ensure reasonable fits were obtained using the above low order polynomial.

Classical molecular dynamics simulations of pure water were performed in order to investigate the distance dependent accuracy of the KSA and related approximations. Simulations of the SPC/E[50] water model were performed using the Gromacs simulation package (version 4.6).[51] The simulations were performed at 300.15 K and 1 bar in a 6nm cubic simulation box, using a time step of 2 fs with bond lengths constrained using LINCS algorithm.[52] Electrostatic interactions were determined using the particle mesh Ewald approach,[53] with a 1.0 nm cutoff for electrostatic interactions and a twin range 1.0 and 1.5 nm cutoff for van der Waals interactions. The system were equilibrated from 0.1 ns and followed with 15 ns production run. Temperature and pressure coupling was achieved using the Berendsen algorithm. The normalized triplet distributions can be computed from the simulation by dividing the real triplet distribution by the equivalent ideal triplet distribution.

5.5 Results

The behavior of the pair fluctuations, b_{11}, in the liquid and super critical regions of the water p-T phase diagram is illustrated in Figure 5.1. Also included are the state points for which the rdfs used here have been determined. The magnitude of b_{11} increases as one approaches the critical point, whereas b_{11} decreases with increasing pressure along the isotherms and for the state points considered here. For comparison, $b_{11} = 1$ for an ideal gas and is approximately 0.01 for ice.[33, 54-55] A detailed FST analysis of the thermodynamics is provided in Table 5.1. The values for c_{111} indicate a negative skewness for the particle number distribution and therefore particle deletion is more favorable than addition on the average. Positive values for d_{1111} indicate the distribution is more peaked than a normal distribution and so small net deletions or insertions are favored over larger deletions or insertions. Both of these quantities decrease in magnitude as the pressure increases, i.e. the particle number distribution tends towards a normal distribution as T decreases and p increases. This is the same region of the phase diagram that tends to the incompressible limit (IL).

The experimental water oxygen-oxygen radial distribution functions used in this study are presented in Figure 5.2 as a function of temperature and pressure. As expected, they generally indicate less structure as the temperature and/or pressure increases. The effect of pressure, however, leads to relatively small changes to the rdfs, even for the states closest to the critical point. The rdfs displayed in Figure 5.2 were used to determine the corresponding structure factors. The small k behavior of the structure factors were then modified using Equation (5.36) to ensure thermodynamic consistency. These modified structure factors were then used in determining the integrals in Equation (5.27) and Equation (5.35). The results of this process are displayed in Figure 5.3 and Figure 5.4 for two of the state points investigated here. At 298 K and 1 bar the difference

between the original and modified structure factors and rdfs is completely negligible. The differences at 423 K and 1900 bar are more significant. However, even here changes to the low k behavior of the structure factor have only small effects on the recalculated rdf. The effect on the third virial coefficient is sizable, but we consider the thermodynamically consistent structure

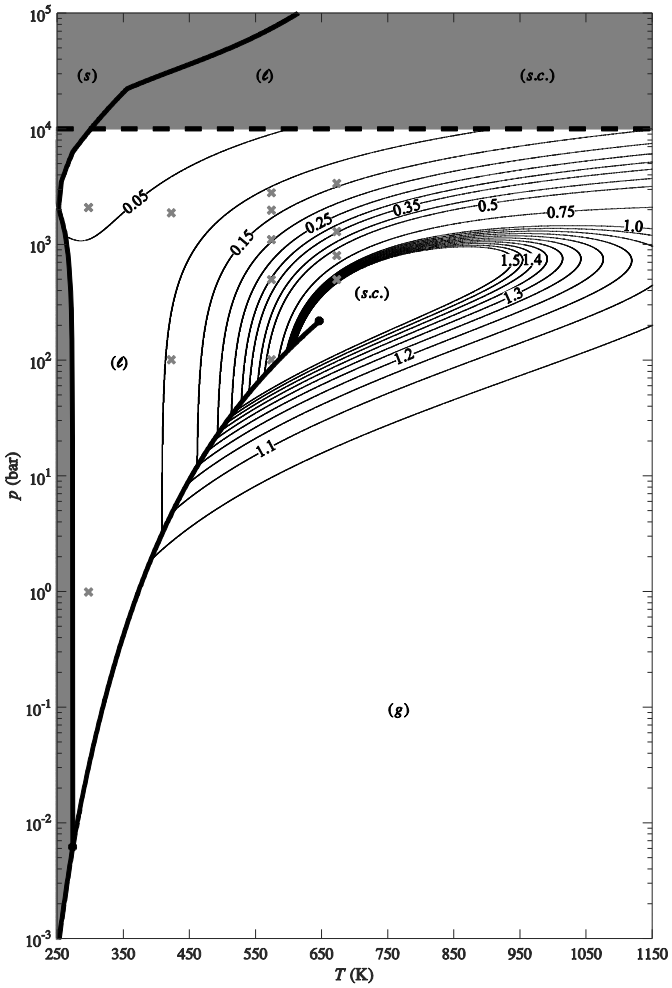

Figure 5.1 Contour plot of b_{11} as a function of temperature and pressure for the liquid (*l*), gas (*g*) and supercritical (*s.c.*) regions of pure water as given by the IAPWS-95 equation of state. The gray-filled regions were not contoured. Crosses indicate the state points considered here. Unlabeled contour values are as follows: 0.1, 0.2, 0.3, 0.425, 1.05, 1.15, 1.25, 1.35, 1.45. Contours above 1.5 were omitted for clarity, because b_{11} is increasing rapidly as the critical point is approached from any phase. The horizontal dashed line indicates the maximum valid pressure of the IAPWS-95 equation of state. The phase coexistence curves are shown as bold lines. The triple point and critical point are shown as filled black circles.

Table 5.1 Fluctuation Solution Theory Based Properties of Fluid Water

T	p	ρ_1	$10^5\kappa_T$	$\rho_1 G_{11}$	$\rho_1^2 G_{111}$	$\rho_1^3 G_{1111}$	b_{11}	c_{111}	d_{1111}	$\rho_1 B_2^*$	$\rho_1^2 B_3^*$	$\rho_1^3 B_4^*$
K	bar	M	bar^{-1}									
298	1	55.34	4.53	-0.94	1.80	-5.21	0.062	-0.014	0.006	0.468	0.279	0.184
298	2100	59.68	2.87	-0.96	1.86	-5.49	0.042	-0.008	0.003	0.479	0.295	0.203
423	100	51.20	5.93	-0.89	1.62	-4.33	0.107	-0.064	0.112	0.447	0.259	0.158
423	1900	55.44	3.42	-0.93	1.78	-5.13	0.067	-0.019	0.016	0.467	0.278	0.182
573	100	39.72	30.5	-0.42	-2.39	49.12	0.577	-2.655	32.834	0.212	0.974	-4.436
573	500	43.11	14.7	-0.70	0.54	3.28	0.303	-0.552	2.631	0.349	0.306	-0.126
573	1100	46.12	8.85	-0.81	1.22	-2.19	0.194	-0.195	0.506	0.403	0.242	0.104
573	1970	49.06	5.80	-0.86	1.51	-3.87	0.136	-0.084	0.137	0.432	0.244	0.141
573	2800	51.16	4.44	-0.89	1.63	-4.45	0.108	-0.049	0.060	0.446	0.253	0.155
673	500	32.10	71.3	0.28	-9.93	172.36	1.281	-8.087	115.737	-0.141	3.390	-25.791
673	800	36.62	29.2	-0.40	-1.44	21.12	0.598	-1.646	10.668	0.201	0.642	-1.609
673	1300	40.58	15.0	-0.66	0.47	2.53	0.341	-0.502	1.762	0.329	0.276	-0.070
673	3400	48.38	5.20	-0.86	1.50	-3.91	0.141	-0.075	0.099	0.430	0.237	0.136
IL			0	-1	2	-6	0	0	0	1/2	1/3	1/4

See the text for definitions

factors to be more reasonable. Also shown in Figure 5.3 and Figure 5.4 are the partially integrated analogues of Equation (5.27) where the integration is performed over all k values up to K. This plot illustrates the known large contributions to the third viral coefficient from the small k (or long r) behavior of the structure factor (or rdfs), typically providing 110-120% of the total contribution.

The results of using the KSA approximation are summarized in Table 5.2. The experimental $G_{111} - 3G_{11}^2$ values are always negative and vary systematically with pressure along each isotherm. The corresponding KSA values are also generally negative, but display significantly less variation with temperature and pressure, especially closer to the critical point. The ΔG_{111} values were typically negative – indicating an overall underestimation of the triplet correlations by the KSA – for state points away from the critical point. As the critical point is

approached the errors become positive and significantly larger in magnitude suggesting the real triplet correlations are increasing much slower than the pair correlations. The percentage errors for the triplet KSA suggest that this is a poor approximation for the majority of state points. The errors indicated by ΔG_{1111} generally display the opposite sign to that observed for ΔG_{111} suggesting that either the quadruplet correlations were overestimated for state points away from the critical point, or ΔG_{1111} is dominated by the error in the triplet correlations. The contribution of I_1, as obtained from Equation (5.35), to the fourth virial coefficient was typically of the same order of magnitude as the virial coefficient, although this contribution dropped to essentially zero as one approached the critical point.

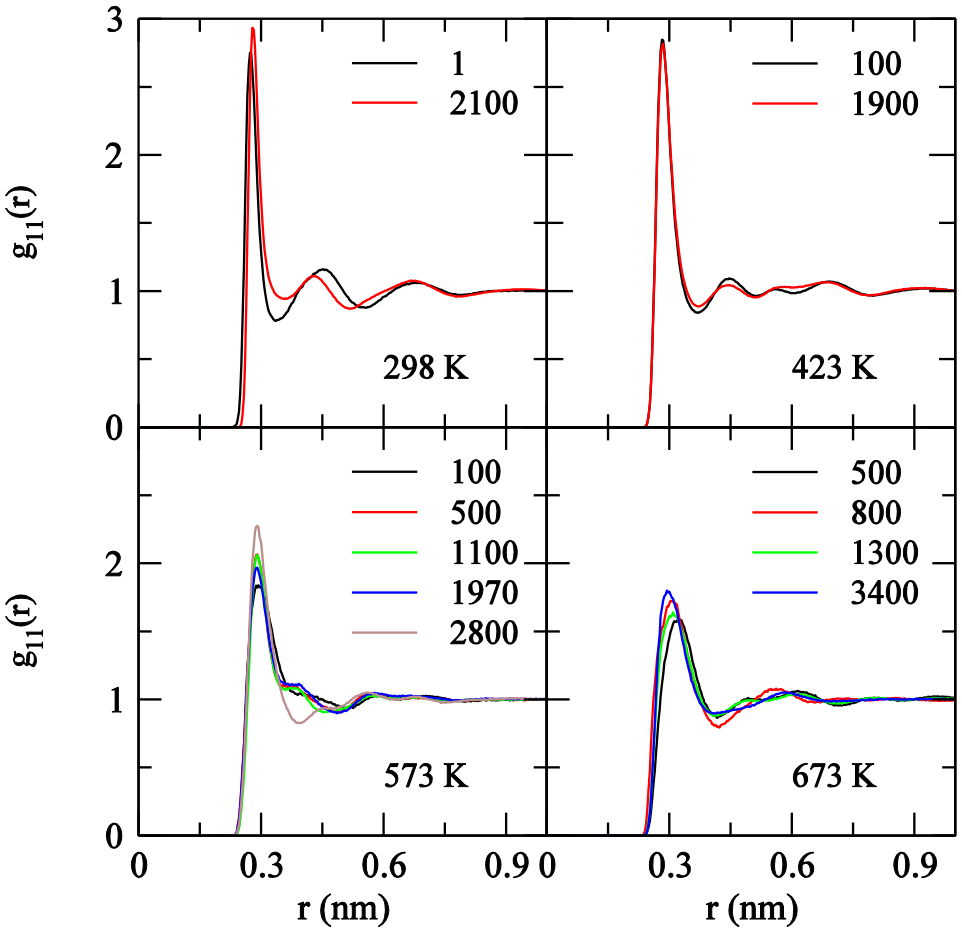

Figure 5.2 Oxygen-oxygen radial distribution functions obtained by Soper and coworkers as a function of temperature and pressure (bar).

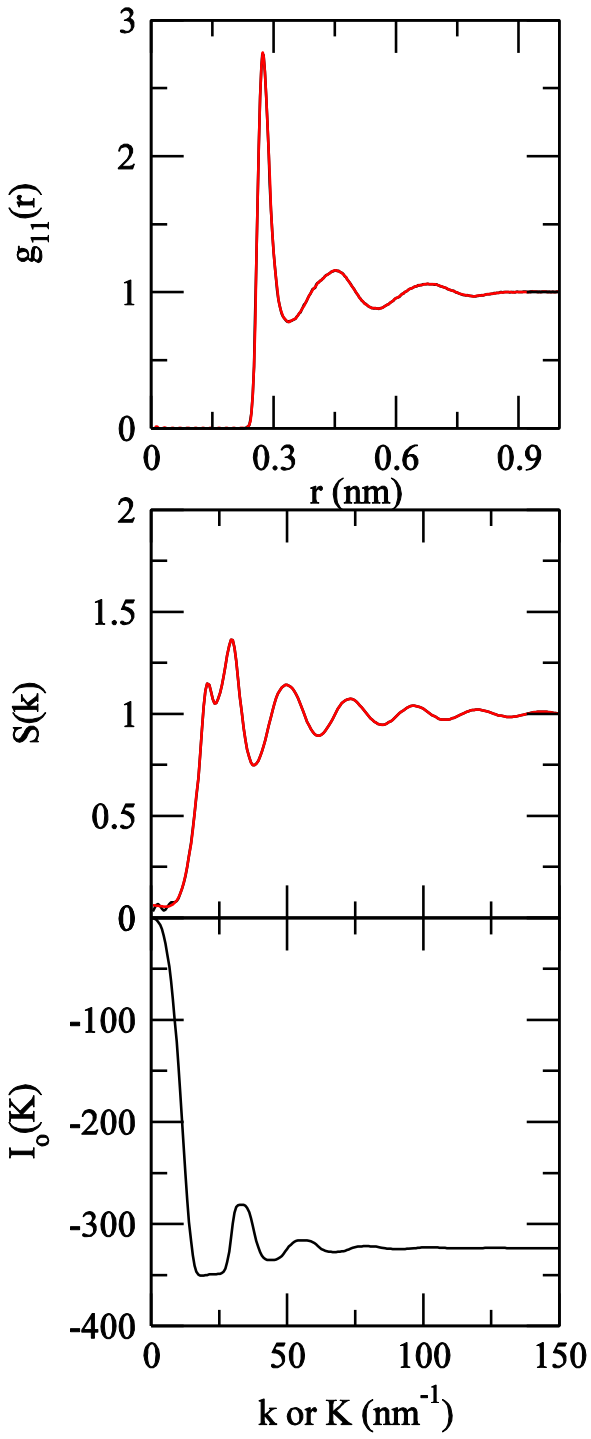

Figure 5.3 Radial distribution functions (top), structure factors (center), and integrals (bottom) for liquid water at 298 K and 1 bar. In the top two panels the original rdf and structure factor are shown in black, while the thermodynamically consistent rdf and structure factor are shown in red. The integral in the bottom panel is displayed as a function of total integration wavevector, K.

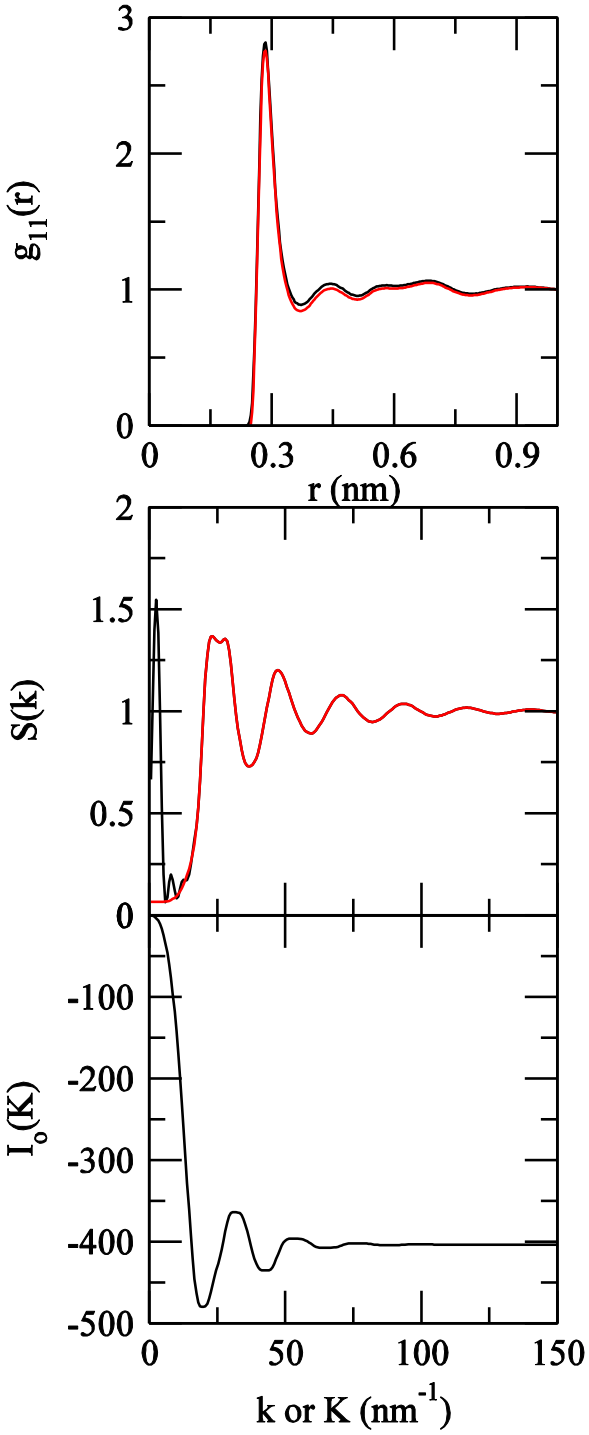

Figure 5.4 Radial distribution functions (top), structure factors (center), and integrals (bottom) for liquid water at 423 K and 1900 bar. In the top two panels the original rdf and structure factor are shown in black, while the thermodynamically consistent rdf and structure factor are shown in red. The integral in the bottom panel is displayed as a function of total integration wavevector, K.

Table 5.2 Triplet and Quadruplet Integrals using the Kirkwood Superposition Approximation.

T	p	ρ_1	G_{111}-$3G_{11}^2$		ΔG_{111}	%err	ΔG_{1111}	%err
K	bar	M	KSA	exp				
298	1	55.34	-324	-274	-50	-9	1313	4
298	2100	59.68	-403	-249	-155	-30	6210	24
423	100	51.20	-394	-297	-97	-16	4407	14
423	1900	55.44	-403	-271	-132	-23	6002	20
573	100	39.72	-443	-1853	1410	93		
573	500	43.11	-408	-494	86	30	-19907	-49
573	1100	46.12	-390	-341	-49	-9	-5967	-27
573	1970	49.06	-437	-304	-133	-21	8851	27
573	2800	51.16	-291	-290	-1	0	11774	35
673	500	32.10	78	-9872	9950	103		
673	800	36.62	-186	-1435	1250	116	-164655	-38
673	1300	40.58	-190	-502	312	109	-101687	-268
673	3400	48.38	-321	-304	-18	-3	69825	202

Units: G_{111} in $(cm^3/mol)^2$ and G_{1111} in $(cm^3/mol)^3$; %err was calculated as $100\%(G_{111}^{KSA} - G_{111}^{exp})/|G_{111}^{exp}|$ or $100\%(G_{1111}^{KSA} - G_{1111}^{exp})/|G_{1111}^{exp}|$, respectively.

It is noticeable that the KSA values for $G_{111} - 3G_{11}^2$ do not always vary systematically with pressure, although the subsequent values for ΔG_{111} are systematic. In an effort to determine how the above results may differ between different experimental determinations of the pair distribution function we have compared a variety of results obtained for water at 298 K and 1 bar. These are presented in Figure 5.5 and include the experimental neutron diffraction rdf data currently used here (Soper, 2000),[41] together with a series of refinement results for the structure factor also obtained from the Soper group but via X-ray scattering (Soper, 2013).[42] All data were made thermodynamically consistent by fitting to Equation (5.36). The data shown in Figure 5.5 indicate that, while there were some differences between the structure factor data, the final integrated values of $G_{111} - 3G_{11}^2 = I_0$ were very consistent with one notable exception. This exception corresponded

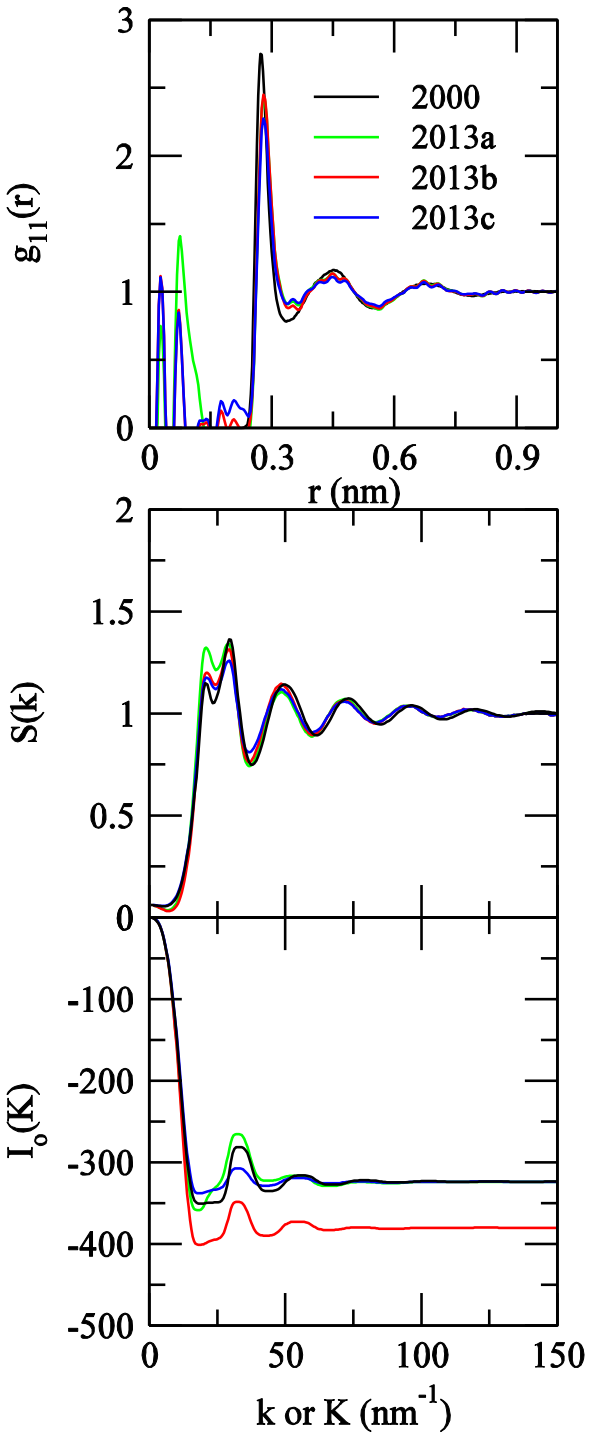

Figure 5.5 Radial distribution functions (top), structure factors (center), and integrals (bottom) for liquid water at 298 K and 1 bar. Rdfs were taken from Soper 2000 (current work) and a series of refinements by Soper 2013. The integral in the bottom panel is displayed as a function of total integration wavevector, K

to a data analysis procedure that was known to produce spurious results.[42] Hence, reasonable estimates for the structure factors produce consistent results for the KSA approximation, yet the results are not so insensitive to the structure factors that there is no information. We conclude that the KSA estimates for $G_{111} - 3G_{11}^2$ are sufficiently accurate, although the exact degree of accuracy might vary between state points, and therefore the results displayed in Table 5.2 are meaningful.

The results for the KSA are not particularly encouraging. This conclusion is in agreement with a variety of other studies.[10, 24, 35] The goal of the KSA was to relate the triplet and pair correlations in fluids. Here, we investigate a series of other approaches which can be used to achieve a similar goal. The first is due to Moelwyn-Hughes who observed that the change in the bulk modulus with pressure along an isotherm is essentially the same over a wide range of state points.[56] Hence, one can write,[56]

$$\left(\frac{\partial \kappa_T^{-1}}{\partial p}\right)_T \equiv \mu \tag{5.38}$$

where μ is a constant for a particular temperature. The value of μ does vary with temperature but only slightly.[33] Using the above definition in Equation (5-10) - Equation (5-12) provides,

$$c_{111} = (2-\mu)b_{11}^2 \tag{5.39}$$

$$d_{1111} = (2-\mu)(3-2\mu)b_{11}^3 - \rho_1 k_B T \mu' b_{11}^2 \tag{5.40}$$

Both expressions are exact if μ and μ' (= $\partial\mu/\partial p$) are evaluated at the state point of interest. The Moelwyn-Hughes approach assumes μ is indeed independent of pressure and so $\mu' = 0$. Consequently, the triplet and quadruplet fluctuations (correlations) are then simply proportional to

powers of the pair fluctuations (correlations). The corresponding relationships between the integrals are given by,

$$\rho_1^2 G_{111} = (1-\mu) + (1-2\mu)\rho_1 G_{11} + (2-\mu)\rho_1^2 G_{11}^2 \tag{5.41}$$

$$\begin{aligned}\rho_1^3 G_{1111} = &-(1-\mu)(1+2\mu) - \rho_1 k_B T\mu' + (5-9\mu+6\mu^2-2\rho_1 k_B T\mu')\rho_1 G_{11} \\ &+ [3(2-\mu)(1-2\mu) - \rho_1 k_B T\mu']\rho_1^2 G_{11}^2 + (2-\mu)(3-2\mu)\rho_1^3 G_{11}^3\end{aligned} \tag{5.42}$$

and implies triplet probability distribution of the form,

$$\begin{aligned}&g_{111}^{(3)}(r_1, r_2, r_3) - 1 \\ &\quad \approx 3h_{11}^{(2)}(r_1, r_2) + (2-\mu)\left[h_{11}^{(2)}(r_1, r_2)h_{11}^{(2)}(r_1, r_3)\right] + (1 \\ &\quad - 2\mu)h_{11}^{(2)}(r_1, r_2)/\langle N_1\rangle + (1-\mu)/\langle N_1\rangle^2\end{aligned} \tag{5.43}$$

Table 5.3 Triplet and Quadruplet Integrals using Moelwyn-Hughes Isotherms.

T	p	ρ_1	ΔG_{111}	%err	ΔG_{1111}	%err
K	bar	M				
298	1	55.34	0	0	28	0
298	2100	59.68	0	0	-8	0
423	100	51.20	9	1	-1554	-5
423	1900	55.44	1	0	-148	0
573	100	39.72	908	60	-566950	-72
573	500	43.11	116	40	-38360	-94
573	1100	46.12	26	5	-6275	-28
573	1970	49.06	7	1	-1337	-4
573	2800	51.16	2	0	-444	-1
673	500	32.10	1992	21	-1918704	-37
673	800	36.62	249	23	-124602	-29
673	1300	40.58	44	15	-14624	-39
673	3400	48.38	0	0	-189	-1

Units: G_{111} in $(cm^3/mol)^2$ and G_{1111} in $(cm^3/mol)^3$. Using Equation (5.39) – Equation (5.42) with $\mu = 5.68$ and $\mu' = 0$.

where the final two terms are only finite upon integration. It should be noted that $\mu = 1$ for an ideal gas resulting in a Poisson particle number distribution, $\mu = 2$ corresponds to a Gaussian particle number distribution, and $\mu = 5\text{-}11$ for common liquids.[57]

In Table 5.3 we investigate the accuracy of the above approximation for water as provided when adopting a single value of $\mu = 5.68$ (obtained at 298 K and 1 bar) for all temperatures and pressures. The results represent a substantial improvement over the KSA, although neither approach performs well on approaching the critical point. In Table 5.4 we provide the results for the related Gaussian ($\mu = 2$) approximation. The Gaussian limit is approached (but never reached) at low temperature and high pressure. Again, the results appear to be very reasonable for the liquid state away from the critical point, although the results for the critical fluid region are significant worse than in Table 5.3. Both approaches overestimate the magnitude of the triplet correlations as one nears the critical point.

Table 5.4 Triplet and Quadruplet Integrals using the Gaussian Approximation

T	p	ρ_1	ΔG_{111}	%err	ΔG_{1111}	%err
K	bar	M				
298	1	55.34	5	1	474	2
298	2100	59.68	2	0	213	1
423	100	51.20	25	4	1425	4
423	1900	55.44	6	1	500	2
573	100	39.72	1683	111	-563027	-72
573	500	43.11	297	102	-16525	-40
573	1100	46.12	92	16	2117	10
573	1970	49.06	35	6	1950	6
573	2800	51.16	19	3	1285	4
673	500	32.10	7852	81	-5794278	-111
673	800	36.62	1228	114	-256521	-60
673	1300	40.58	305	106	-12095	-32
673	3400	48.38	32	5	1932	6

Units: G_{111} in $(cm^3/mol)^2$ and G_{1111} in $(cm^3/mol)^3$. Using Equation (5.39) – Equation (5.42) with $\mu = 2$ and $\mu' = 0$.

An alternative approach is to assume that the rdfs are independent of pressure. This is clearly an approximation as indicated in Figure 5.2. Nevertheless, some of observed increases and decreases may cancel on integration leading to a reasonable approximation. The derivatives in Equation (5.13) – Equation (5.17) are then zero under these conditions and we find,

$$c_{111} = 1 + 3\rho_1 G_{11} + 2\rho_1^2 G_{11}^2 \tag{5.44}$$

$$d_{1111} = 1 + 7\rho_1 G_{11} + 12\rho_1^2 G_{11}^2 + 6\rho_1^3 G_{11}^3 \tag{5.45}$$

and,

$$G_{111} = 2G_{11}^2 \tag{5.46}$$

$$G_{1111} = 6G_{11}^3 \tag{5.47}$$

which implies a triplet distribution function of the form,

$$\begin{aligned} g_{111}^{(3)}(r_1, r_2, r_3) - 1 & \\ \approx h_{11}^{(2)}(r_1, r_2) h_{11}^{(2)}(r_1, r_3) + h_{11}^{(2)}(r_1, r_2) h_{11}^{(2)}(r_2, r_3) + h_{11}^{(2)}(r_1, r_2) & \\ + h_{11}^{(2)}(r_1, r_3) + h_{11}^{(2)}(r_2, r_3) & \end{aligned} \tag{5.48}$$

and corresponds to a situation where $\mu = 1/b_{11}$.

The results obtained from such an approximation are provided in Table 5.5. Again, the results are very good for the liquid region away from the critical point, with a small general underestimation of the triplet correlations. This corresponds to regions where the G_{111}/G_{11}^2 and G_{1111}/G_{11}^3 ratios approach the values predicted by Equation (5.46) – Equation (5.47). However,

the fluid region is unsatisfactory and the triplet correlations are again overestimated on approaching the critical point.

Table 5.5 Triplet and Quadruplet Integrals Assuming Pressure Independent rdfs.

T	P	ρ_1	G_{111}/G_{11}^2	G_{1111}/G_{11}^3	ΔG_{111}	%err	ΔG_{1111}	%err
K	Bar	M						
298	1	55.34	2.05	6.34	-13	-2	166	5
298	2100	59.68	2.03	6.25	-9	-2	1040	4
423	100	51.20	2.02	6.07	-8	-1	388	1
423	1900	55.44	2.04	6.31	-12	-2	1502	5
573	100	39.72	-13.33	-648.69	1739	115	-790896	-101
573	500	43.11	1.11	-9.66	233	80	-66257	-162
573	1100	46.12	1.88	4.18	36	6	-9687	-43
573	1970	49.06	2.02	5.99	-6	-1	-59	0
573	2800	51.16	2.04	6.28	-14	-2	1476	4
673	500	32.10	-125.55	7747.67	9795	102	-5209463	-100
673	800	36.62	-8.89	-324.29	1315	122	-438005	-102
673	1300	40.58	1.09	-8.86	239	83	-63514	-168
673	3400	48.38	2.04	6.16	-12	-2	891	3

Units: G_{111} in $(cm^3/mol)^2$ and G_{1111} in $(cm^3/mol)^3$. Using Equation (5.44) and Equation (5.47) corresponding to $\mu = 1/b_{11}$ and $\rho_1 k_B T \mu' = \mu - 1$.

The previous non-KSA approaches appear to work well for regions not too close to the critical point. We can make this statement more quantitative by analysis of the results in Table 5.1, Table 5.3-Table 5.5 and the data shown in Figure 5.1. It appears that reasonable results are obtained when $b_{11} = \rho_1 k_B T \kappa_T < 0.2$, i.e. low to moderate compressibility. This covers a large portion of the liquid and super critical regions of the phase diagram. Unfortunately, we do not know if this condition holds true for other systems. The performance is also shown graphically in Figure 5.6 for both ΔG_{111} and ΔG_{1111}. All the methods perform well for $b_{11} < 0.2$, while the KSA underestimates the three body correlations for low compressibilities but overestimates the correlations at moderate compressibilities.

The reason for the success of the non-KSA approaches for lower compressibilities appears to lie in the fact that when b_{11} is small then so are c_{111} and d_{1111}. Hence, it doesn't matter whether one uses $\mu = 2$ (Gaussian), $\mu = 5.68$ (water at 298 K and 1 bar), or $\mu = 1/b_{11} = 16.1$ (water at 298 K and 1 bar) one finds $\rho_1 G_{11} \approx -1$ and the expressions in Equations (5.41) – Equations (5.42) are essentially independent of μ as they are then close to the IL, or closed system, values.

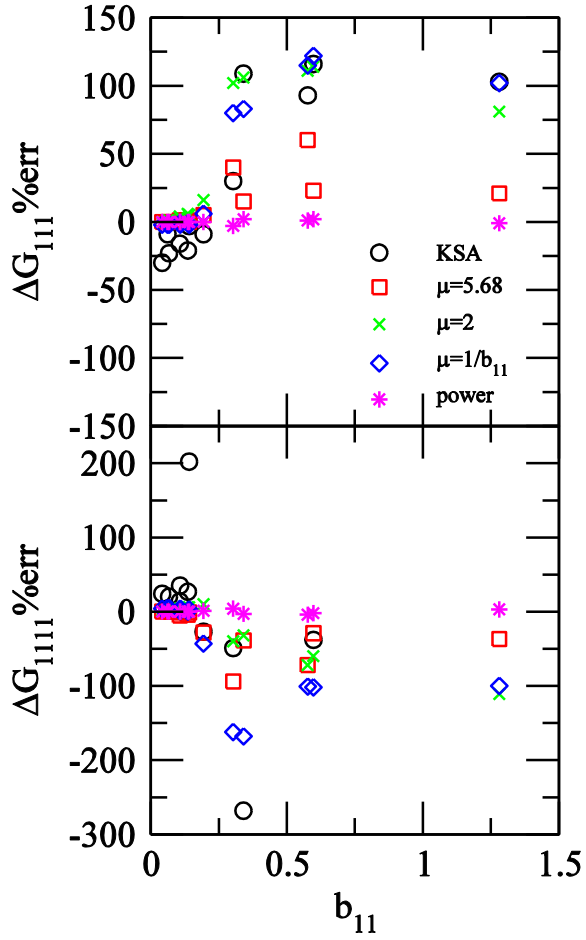

Figure 5.6 Observed errors obtained for the triplet and quadruplet integrals as obtained from the KSA and a series of approximate relationships between the pair and triplet fluctuations (see text for details) as a function of the reduced pair fluctuations.

The final approximation investigated here involves a simple power law dependence between the triplet and pair fluctuations along a particular isotherm. A plot of ln c_{111} vs ln b_{11} is displayed in Figure 5.7 and suggests the following simple function form,

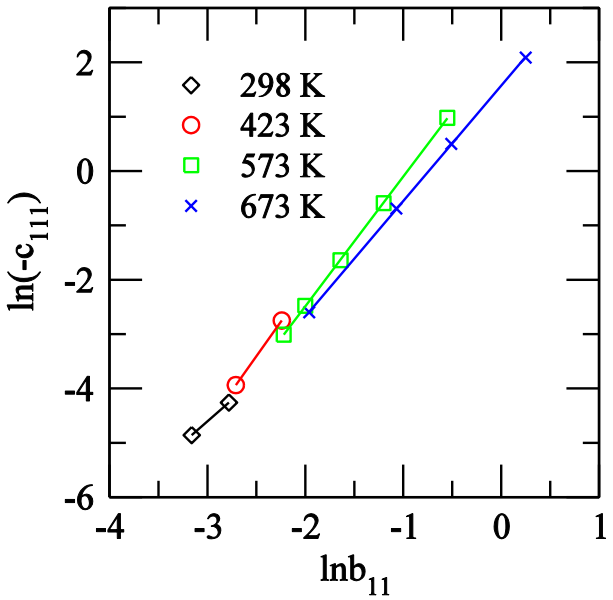

Figure 5.7 The correlation between triplet and pair fluctuations for different isotherms. The symbols represent the experimental data. The lines represent fits to the data using Equation (5.49) - Equation (5.50).

$$c_{111} = -yb_{11}^{m} \tag{5.49}$$

$$d_{1111} = yb_{11}^{m+1}[myb_{11}^{m-2} + m - 1] \tag{5.50}$$

where the second expression has been obtained from the isothermal pressure derivative of c_{111} using previous relationships.[33] Here, y and m are constants for a particular isotherm. The resulting fits provided in Figure 5.7 are excellent for the all the isotherms and pressures considered here. Further analysis presented in Table 5.6 also suggests that almost perfect agreement with experiment for the triplet and quadruplet correlations can be obtained for all state points considered here, i.e. for $b_{11} < 1.25$. Additional examination of states along the $T = 673$ K isotherm suggest that reasonable results (< 5% error) can be obtained for the triplet correlations up to $b_{11} \approx 2.0$.

Table 5.6 Triplet and Quadruplet Integrals Assuming a Power Law Dependence.

T	P	ρ_1	ΔG_{111}	%err	ΔG_{1111}	%err
K	bar	M				
298	1	55.34	0	0	35	0
298	2100	59.68	0	0	24	0
423	100	51.20	0	0	-23	0
423	1900	55.44	0	0	3	0
573	100	39.72	11	1	-30647	-4
573	500	43.11	-8	-3	1694	4
573	1100	46.12	-1	0	240	1
573	1970	49.06	0	0	12	0
573	2800	51.16	0	0	-5	0
673	500	32.10	-49	-1	173483	3
673	800	36.62	26	2	-7285	-2
673	1300	40.58	6	2	-1065	-3
673	3400	48.38	0	0	-37	0

Units: G_{111} in $(cm^3/mol)^2$ and G_{1111} in $(cm^3/mol)^3$
Using Equation (5.4) – Equation (5.6) and Equation (5.49) - Equation (5.50) corresponding to $\mu = 2 + y b_{11}^{m-2}$ and $\rho_1 k_B T \mu' = (m-2)(\mu-2)(1-\mu) b_{11}$. The values of m and y for each isotherm were as follows: 1.58 and 1.14 at 298 K, 2.53 and 18.57 at 423 K, 2.39 and 9.81 at 573 K, 2.12 and 4.81 at 673 K. The Moelwyn-Hughes isotherm would result in a slope of $m = 2$.

5.5.1 Analysis of the Molecular Dynamics Simulations

The simulated triplet correlations for pure water obtained with a series of different approximations are displayed in Figure 5.8-Figure 5.11. The contour plots were generated by fixing r_1 at the distances corresponding to first three consecutive peaks of $g^{(2)}$. When r_1 is fixed at the 1st peak in $g^{(2)}$ all the models and the real triplet correlation display a higher probability compared to when r_1 is fixed at the second or third peak. Moreover, all tested models show

relatively small deviations from the real $g^{(3)}(r_1, r_2, r_3)$ values for r_1 distances corresponding to the second and third peaks of $g^{(2)}$, compared to the first peak of $g^{(2)}$. Hence, significant uncertainty between the real and model values can be observed when r_1 is at the 1st peak in $g^{(2)}$. The white color regions represent values that are either above or below the range specified by the color bar. It is observed that the real figures have a upper left and lower right triangles (when r_1 at 1st peak in $g^{(2)}$) with the triplet distribution values corresponding to minus one. This is the most notable difference between the real triplet distribution and all the approximate models. In other words, the real triplet distribution has a greatly restricted area of space within which it is possible to find all three molecules. However, all the models were unable to capture this restricted distribution as they use only the pair distribution to predict the triplet distribution.

The non KSA approximations appear to work well from the experimental analysis discussed before. However, the previously explained experimental analysis does not provide any information regarding the distance dependent accuracy of these models. However, from the molecular dynamics simulations, it is possible to gain information regarding the distance dependent accuracies of these models. According to the simulation results it is observed that the asymptotic model and the Moelwyn-Hughes (μ=5.96) model display higher uncertainty compare to the KSA and the pressure independent model (Equation 5.48), especially at shorter distances.

In general, the greatest disagreement between the approximate models and the real triplet correlation can be observed when the distance between two particles is closed to the distance of the first peak in $g^{(2)}$. In particular, none of these approximate relations capture all of the features of real triplet distribution.

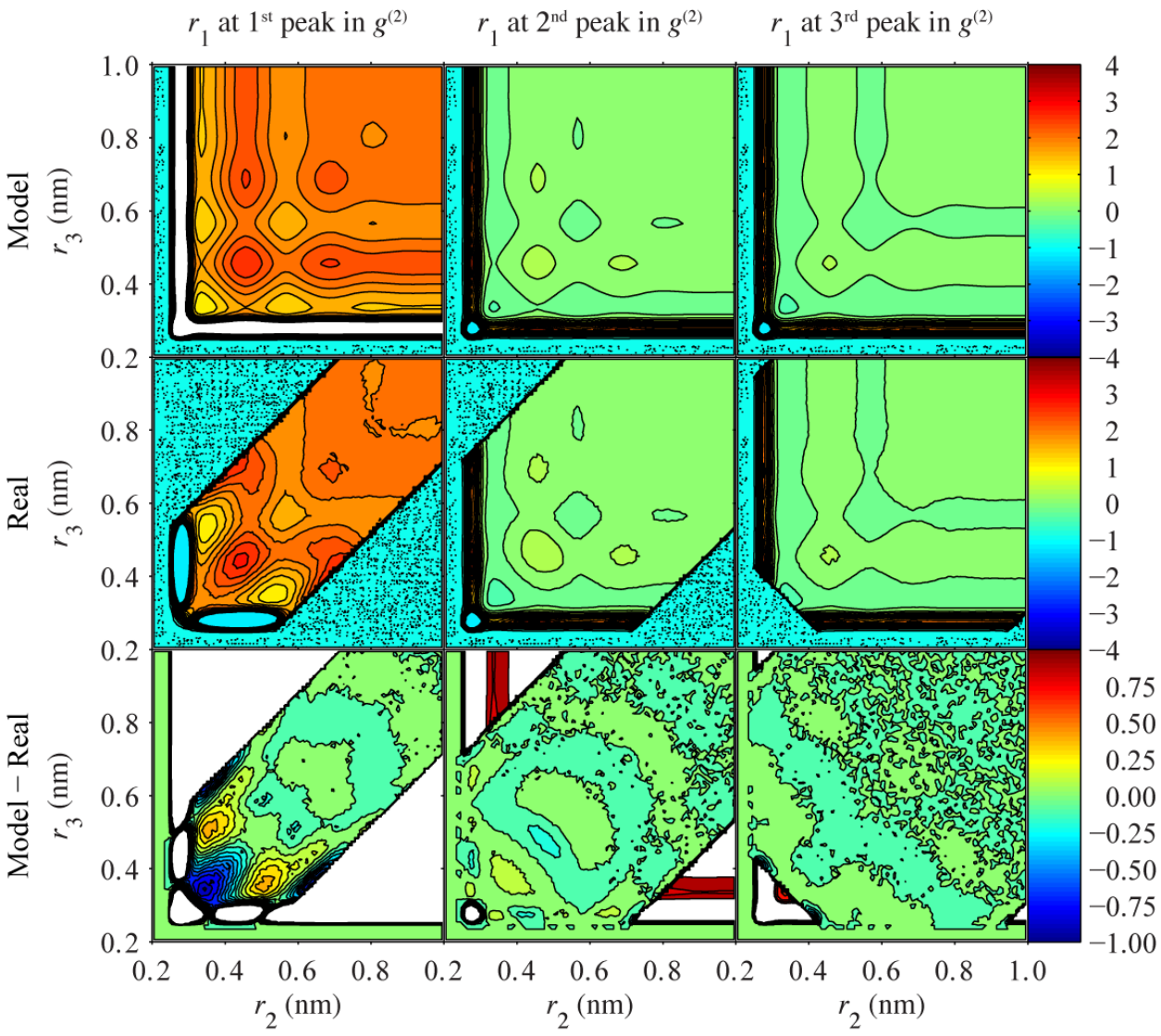

Figure 5.8 Triplet correlation functions for pure water at 300.15K and 1 bar. The top three panels display the values of $g^{(3)}(r_1, r_2, r_3)$-1 corresponding to KSA model. The middle three panels display the real $g^{(3)}(r_1, r_2, r_3)$-1. The bottom three panels display the difference between model and real values. The distances of first, second and third peaks are 0.275, 0.455 and 0.685 nm, respectively.

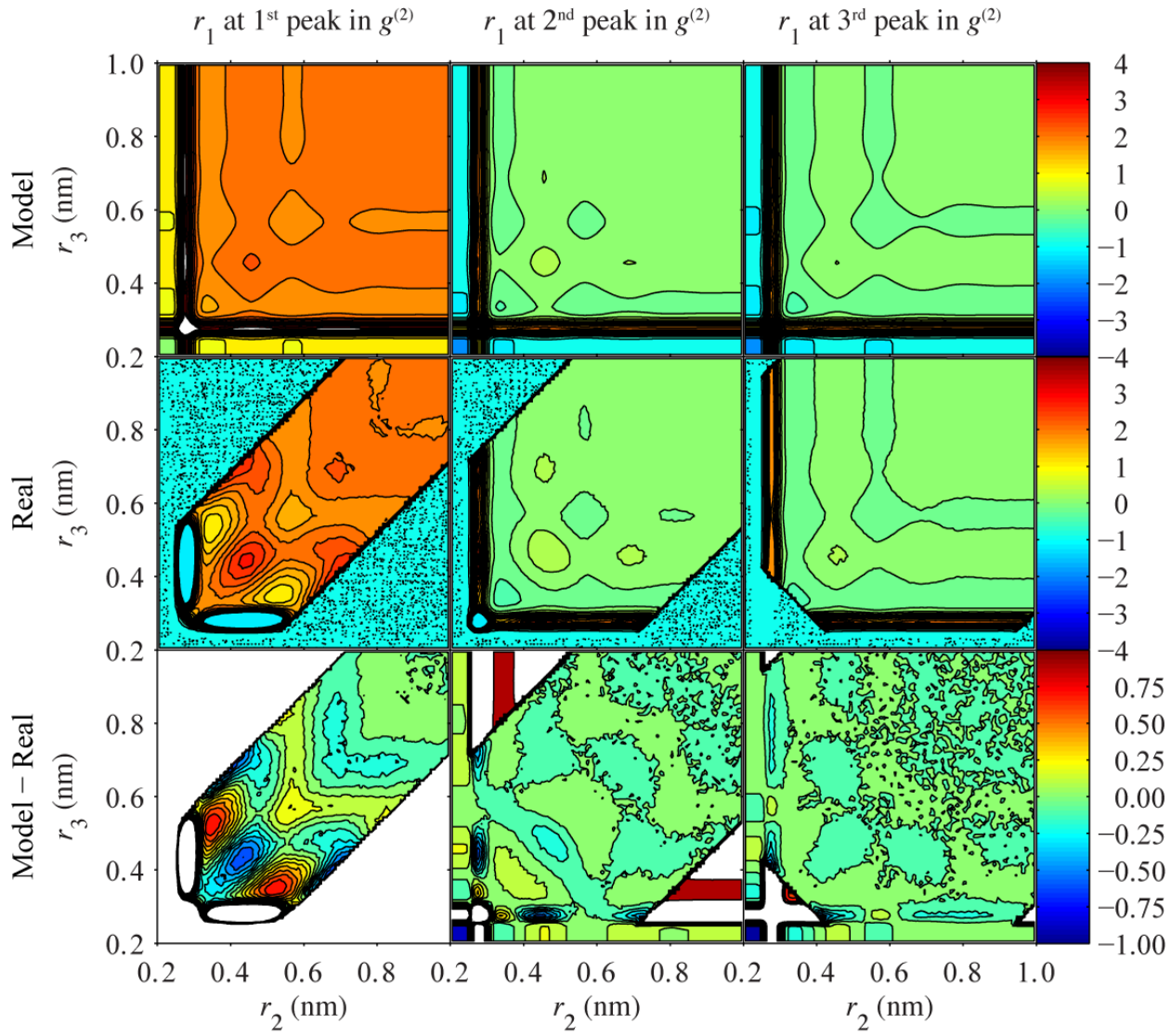

Figure 5.9 Triplet correlation functions for pure water at 300.15K and 1 bar. The top three panels display the values of $g^{(3)}(r_1, r_2, r_3)$-1 corresponding to asymptotic model ($3g^{(2)}$-2). The middle three panels display the real $g^{(3)}(r_1, r_2, r_3)$-1. The bottom three panels display the difference between model and real values. The distances of first, second and third peaks are 0.275, 0.455 and 0.685 nm, respectively.

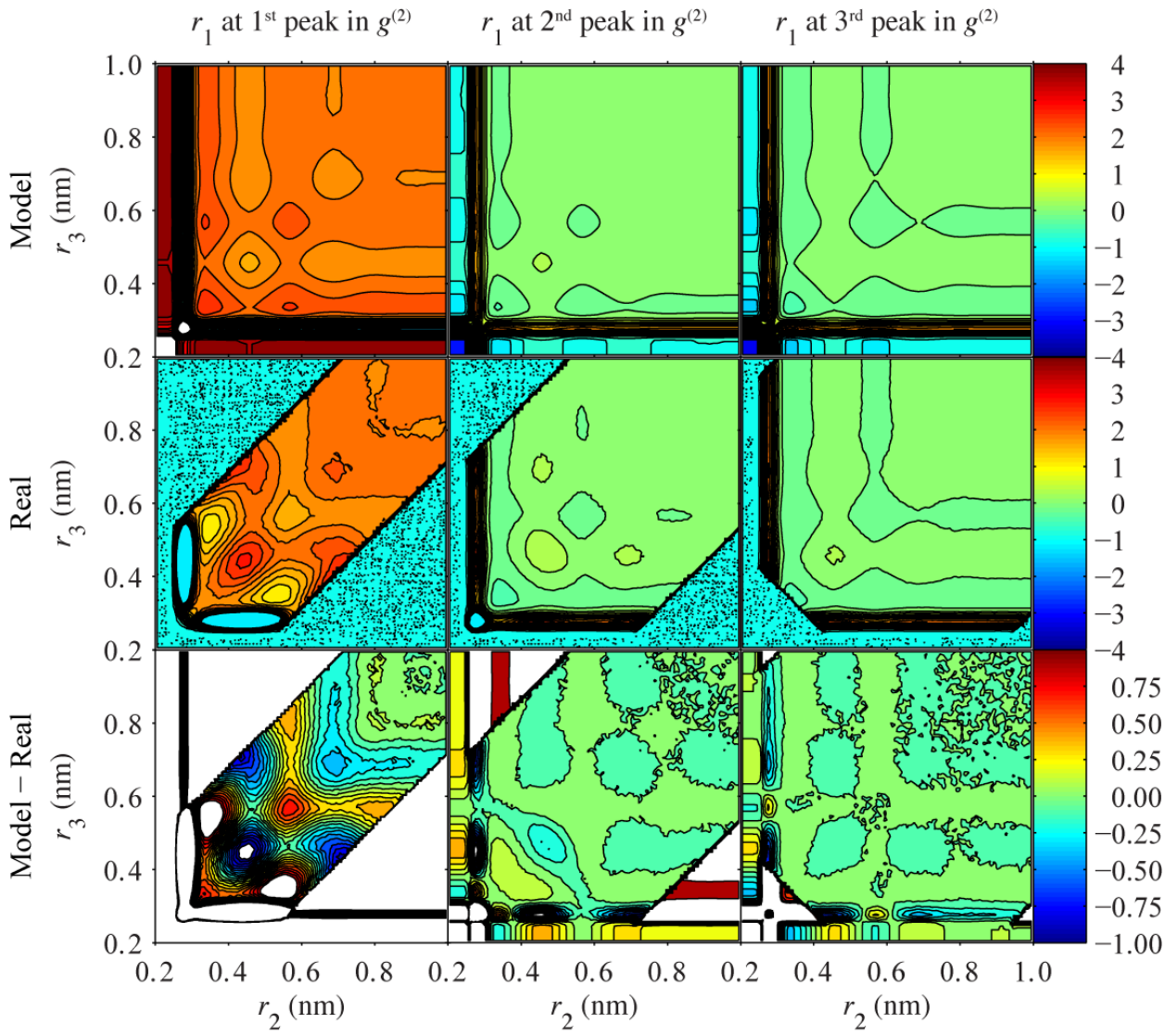

Figure 5.10 Triplet correlation functions for pure water at 300.15K and 1 bar. The top three panels display the values of $g^{(3)}(r_1, r_2, r_3)$-1 corresponding to μ=5.96 (value corresponding to the SPC/E water). The middle three panels display the real $g^{(3)}(r_1, r_2, r_3)$-1. The bottom three panels display the difference between model and real values. The distances of first, second and third peaks are 0.275, 0.455 and 0.685 nm, respectively.

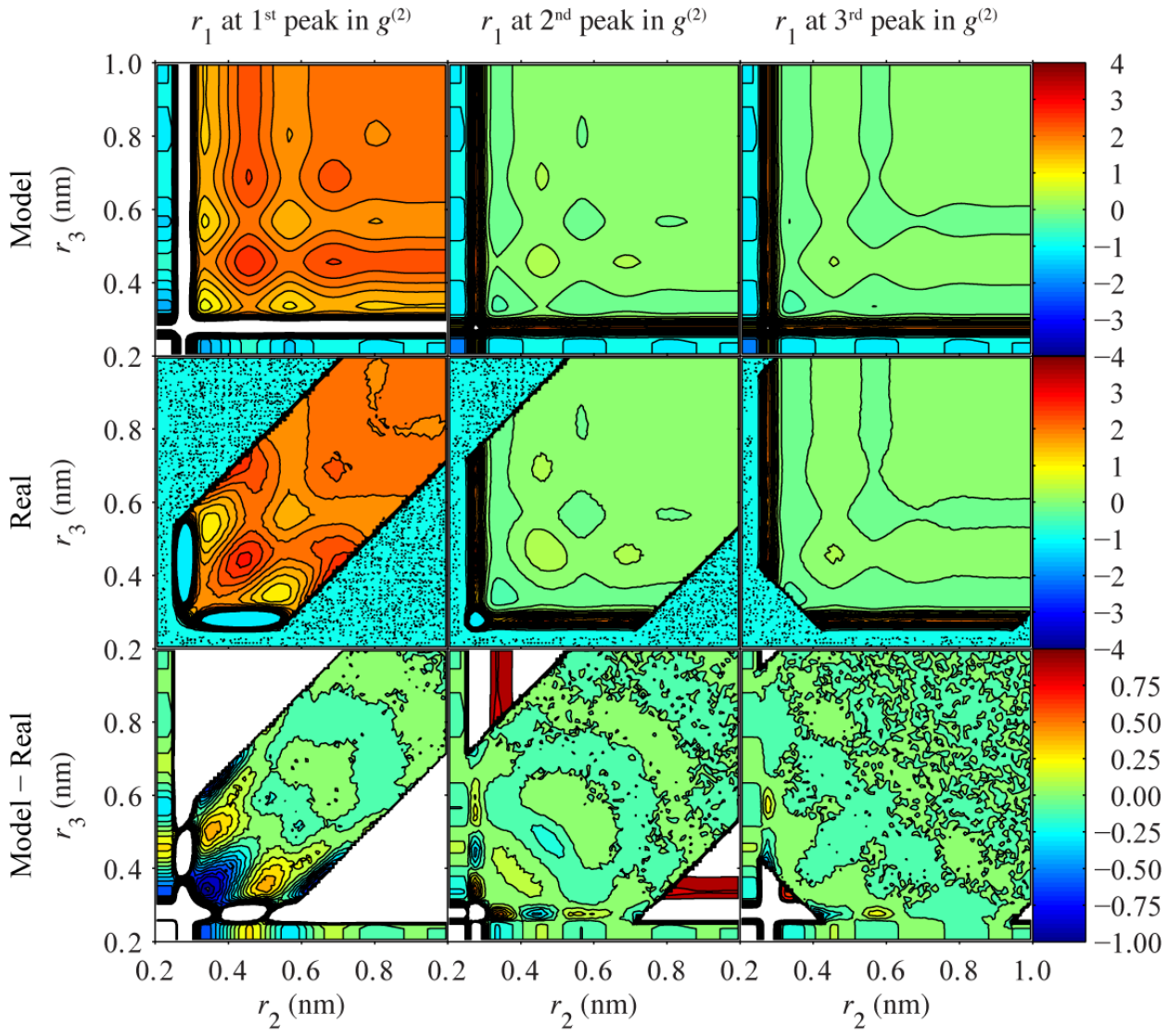

Figure 5.11 Triplet correlation functions for pure water at 300.15K and 1 bar. The top three panels display the values of $g^{(3)}(r_1, r_2, r_3)$-1 corresponding to pressure independent model model (Equation 5.48). The middle three panels display the real $g^{(3)}(r_1, r_2, r_3)$-1. The bottom three panels display the difference between model and real values. The distances of first, second and third peaks are 0.275, 0.455 and 0.685 nm, respectively.

5.6 Conclusions

We have investigated the applicability of the KSA, together with several other approximations relating pair and triplet correlations, for fluid water over a range of temperatures and pressures. The KSA does not perform well with a general underestimation of the three body correlations at low compressibilities (b_{11}), and a general overestimation of the correlations at moderate to high compressibilities. A series of other relationships between the pair and triplet correlations were investigated and all gave good results for values of $b_{11} < 0.2$. An observed power law relationship between the pair and triplet fluctuations reproduced the triplet and quadruplet correlations very accurately up to values of $b_{11} \approx 2.0$, which covers most of the fluid region except for states close to the critical point.

5.7 References

1. Gray, C. G.; Gubbins, K. E. *Theory of Molecular Fluids. Vol. 1:Fundamentals*. Oxford University Press: New York, 1984.

2. Hill, T. L. *Statistical Mechanics: Principle and selected Applications*. New York: McGraw-Hill Book Company Inc., 1956.

3. Hansen J. P, McDonald. I. R. *Theory of Simple Liquids*. Second ed.; Academic Press: London, 1986.

4. Ben-Naim, A. *Molecular theory of solutions*. Oxford university press, 2006.

5. Egelstaff, P. A.; Page, D. I.; Heard, C. R. T. *J. Phys C* **1971,** *4* (12), 1453.

6. Egelstaff, P. A. *Molecular Liquids: New Perspectives in Physics and Chemistry*. Springer, Netherlands, 1992; Vol. 379.

7. Winfield, D. J.; Egelstaff, P. A. *Can. J. Phys.* **1973,** *51* (18), 1965-1970.

8. Cole, G. H. A. *Rep. Prog. Phys.* **1968,** *31* (2), 419.

9. Kirkwood, J. G. *J. Chem. Phys.* **1935,** *3* (5), 300-313.

10. Grouba, V.; Zorin, A.; Sevastianov, L. *Int. J. Mod Phys B* **2004,** *18* (01), 1-44.

11. Sarolea, L.; Mayer, J. E. *Phys. Rev.* **1956,** *101* (6), 1627-1640.

12. Kirkwood, J. G.; Boggs, E. M. *J. Chem. Phys.* **1942,** *10* (6), 394-402.

13. Naim, A. B. *Journal of advances in chemistry* **2014,** *1* (1).

14. Hart, R. W.; Wallis, R.; Pode, L. *J. Chem. Phys.* **1951,** *19* (1), 139-140.

15. Nijboer, B. R. A.; Van Hove, L. *Phys. Rev.* **1952,** *85* (5), 777-783.

16. de Boer, J. *Proc. R. Soc. London, Ser. A* **1952,** *215* (1120), 4-29.

17. Kirkwood, J. G.; Lewinson, V. A.; Alder, B. J. *J. Chem. Phys.* **1952,** *20* (6), 929-938.

18. Salsburg, Z. W.; Zwanzig, R. W.; Kirkwood, J. G. *J. Chem. Phys.* **1953,** *21* (6), 1098-1107.

19. Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1957,** *27* (5), 1208-1209.

20. Wood, W. W.; Jacobson, J. D. *J. Chem. Phys.* **1957,** *27* (5), 1207-1208.

21. Rowlinson, J. S. *Mol. Phys.* **1963,** *6* (5), 517-524.

22. Alder, B. J. *Phys. Rew. Lett* **1964,** *12* (575).

23. Krumhansl, J. A.; Wang, S. s. *J. Chem. Phys.* **1972,** *56* (5), 2034-2041.

24. Krumhansl, J. A.; Wang, S. s. *J. Chem. Phys.* **1972,** *56* (5), 2179-2180.

25. Bildstein, B.; Kahl, G. *J. Chem. Phys.* **1994,** *100* (8), 5882-5893.

26. Ben-Amotz, D.; Stamatopoulou, A.; Yoon, B. J. *J. Chem. Phys.* **1997,** *107* (17), 6831-6838.

27. Dhabal, D.; Singh, M.; Wikfeldt, K. T.; Chakravarty, C. *J. Chem. Phys.* **2014,** *141* (17), 174504.

28. Piasecki, J.; Szymczak, P.; Kozak, J. J. *J. Chem. Phys.* **2013,** *139* (14), 141101.

29. Singer, A. *J. Chem. Phys.* **2004,** *121* (8), 3657-3666.

30. Zahn, K.; Maret, G.; Ruß, C.; von Grünberg, H. H. *Phys. Rev. Lett.* **2003,** *91* (11), 115502.

31. 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.

32. Ploetz, E. A.; Smith, P. E. *J. Phys. Chem. B* **2015,** *119* (25), 7761-7777.

33. Ploetz, E. A.; Karunaweera, S.; Smith, P. E. *J. Chem. Phys.* **2015,** *142* (4), 044502.

34. Ploetz, E. A.; Smith, P. E. *J. Chem. Phys* **2015,** *142*, 094504.

35. Buff, F. P.; Brout, R. *J. Chem. Phys.* **1960,** *33* (5), 1417-1420.

36. Schofield, P. *Proc. Phys. Soc. London* **1966,** *88* (559P), 149-170.

37. Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*. Oxford University Press: 1987.

38. H. Gould, J. T. *Statistical and Thermal Physics*. Princeton University Press: 2010

39. Wagner, W.; Pruss, A. *J. Phys. Chem. Ref. Data* **2002,** *31* (2), 387-535.

40. Harvey, A. H.; Peskin, A. P.; Klein, S. A. *NIST Standard Reference Database 10: NIST/ASME Steam Properties*, Version 2.22; U.S. Department of Commerce: Gaithersburg, 2008.

41. Soper, A. K. *Chem. Phys.* **2000,** *258* (2), 121-137.

42. Soper, A. K. *isrn Phys. Chem.* **2013,** *2013*, 67.

43. Yarnell, J. L.; Katz, M. J.; Wenzel, R. G.; Koenig, S. H. *Phys. Rev. A* **1973,** *7* (6), 2130-2144.

44. Clark, G. N. I.; Cappa, C. D.; Smith, J. D.; Saykally, R. J.; Head-Gordon, T. The structure of ambient water. In *An International Journal at the Interface Between Chemistry and Physics*, Taylor & Francis Group: 2010; Vol. 108, pp 1415-1433.

45. Hura, G.; Sorenson, J. M.; Glaeser, R. M.; Head-Gordon, T. *J. Chem. Phys.* **2000,** *113* (20).

46. Clark, G. N. I.; Hura, G. L.; Teixeira, J.; Soper, A. K.; Head-Gordon, T. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (32), 14003.

47. Head-Gordon, T.; Hura, G. *Chem. Rev.* **2002,** *102* (8), 2651-2670.

48. Beeby, J. L. *J. Phys. C* **1973,** *6* (14), 2262.

49. Stanley, H. E. *Introduction to phase transitions and critical phenomena*. Oxford university press: New York and Oxford, 1971.

50. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987,** *91* (24), 6269-6271.

51. Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. *Bioinformatics* **2013,** *29* (7), 845-54.

52. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

53. Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys* **1993,** *98* (12), 10089-10092.

54. Richards, T. W.; Speyers, C. L. *J. Am. Chem. Soc.* **1914,** *36* (3), 491-494.

55. Gagnon, R. E.; Kiefte, H.; Clouter, M. J.; Edward, W. *J. Chem. Phys.* **1990,** *92*, 1909.

56. Moelwyn-Hughes, E. A. *Physical Chemistry. 2nd Ed*. Pergamon Press: Oxford, 1961.

57. Kharakoz, D. P. *Biophys. J.* **2000,** *79* (1), 511-525.

Chapter 6 - Summary and Future Work

Molecular dynamics simulations have become one of the most important and reliable techniques to study complex biological systems. However, the accuracy of the simulation results depend on the force field parameters. Kirkwood Buff (KB) theory provides an approach to validate and develop force field parameters. Smith and coworkers have been using KB theory with the intention of developing more accurate force fields, known as the Kirkwood Buff derived Force Fields (KBFF), to eventually perform simulations of biological systems. The protein force field is almost compete and the current attempt is to develop force field parameters for phospholipids. As a contribution to that, we have developed force field parameters for esters, with the overall goal of obtaining accurate force field parameters to model phospholipid molecules, and perform molecular dynamics simulations of protein membranes. In addition, a set of small molecular force fields were validated using the KB theory. Chapter 2 and Chapter 3 illustrated the approach of validating and developing force field parameters based on KB theory.

Preferential solvation is an important concept describing solution mixtures that can be evaluated using the KB theory. The usual approach is based on local mole fractions. Here, we have proposed a new approach based on the local volume fraction. Chapter 4 then showed a detailed analysis of different preferential solvation measures using both experimental and simulation data.

An approach based on Fluctuation Solution Theory (FST) has been used to evaluate the Kirkwood Superposition approximation (KSA), together with a series of additional approximate relationships between pair and triplet correlation functions, for fluid water at a series of state points. A comprehensive analysis is described in Chapter 5. Generally, the non KSA approximations provide good agreement with the fluid regions of the phase diagram where the compressibility is small. A simple power low relationship between the pair and triplet fluctuations

is observed for the fluid region. The applicability of the power low relation requires further study for the gas region of the phase diagram of water, as well as for other interesting liquids. This would allow us to obtain a better understanding of the applicability of the power low relationship in attempting to improve current Equation of States methodologies.

Appendix A - Copy of the Permission Letter from the Publisher

3/13/2016 RightsLink® by Copyright Clearance Center

Copyright Clearance Center RightsLink®

Home | Create Account | Help | Live Chat

ACS Publications
Most Trusted. Most Cited. Most Read.

Title: Preferential Solvation in Binary and Ternary Mixtures
Author: Gayani N. Pallewela, Paul E. Smith
Publication: The Journal of Physical Chemistry B
Publisher: American Chemical Society
Date: Dec 1, 2015
Copyright © 2015, American Chemical Society

LOGIN
If you're a copyright.com user, you can login to RightsLink using your copyright.com credentials. Already a RightsLink user or want to learn more?

Quick Price Estimate

Permission for this particular request is granted for print and electronic formats, and translations, at no charge. Figures and tables may be modified. Appropriate credit should be given. Please print this page for your records and provide a copy to your publisher. Requests for up to 4 figures require only this record. Five or more figures will generate a printout of additional terms and conditions. Appropriate credit should read: "Reprinted with permission from {COMPLETE REFERENCE CITATION}. Copyright {YEAR} American Chemical Society." Insert appropriate information in place of the capitalized words.

I would like to...	reuse in a Thesis/Dissertation
Requestor Type	Author (original work)
Portion	Full article
Format	Print and Electronic
Will you be translating?	No
Select your currency	USD - $
Quick Price	Click Quick Price

This service provides permission for reuse only. If you do not have a copy of the article you are using, you may copy and paste the content and reuse according to the terms of your agreement. Please be advised that obtaining the content you license is a separate transaction not involving Rightslink.

QUICK PRICE CONTINUE

To request permission for a type of use not listed, please contact the publisher directly.

https://s100.copyright.com/AppDispatchServlet#formTop 1/1