MOLECULAR DYNAMICS SIMULATIONS OF SOLUTION MIXTURES AND SOLUTION/VAPOR INTERFACES

by

FENG CHEN

B.S., SiChuan University, China, 1999
M.S., Kansas State University, USA, 2003

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

2010

Abstract

In the past several decades, molecular dynamics (MD) simulations have played an important role in providing atomic details for phenomena of interest. The force field used in MD simulations is a critical factor determining the quality of the simulations. Kirkwood-Buff (KB) theory has been applied to study preferential interactions and to develop a new force field. KB theory provides a path from quantities determined from simulation data to the corresponding thermodynamic data. Here we combine KB theory and molecular simulations to study a variety of intermolecular interactions in solution. First, a force field for the computer simulation of aqueous solutions of alcohols is presented. The force field is designed to reproduce the experimentally observed density and KB integrals for a series of alcohols, allowing for an accurate description of alcohols' activity. Other properties such as the translational diffusion constant and heat of mixing are also well reproduced. Second, the newly developed force field is then extended to more complicated systems, such as peptide or mini-proteins, to determine backbone dihedral potentials energetics. The models developed here provide a basis for an accurate force field for peptides and proteins. Third, we have then studied the surface tension of a variety water models. Results showed that different simulation conditions can affect the final values of surface tension. Finally, by using the Kirkwood-Buff theory of solution and surface probability distributions, we attempted to characterize the properties of the Gas/Liquid interface region. The same approach is then used to understand the relationship between changes in surface tension, the degree of surface adsorption or depletion, and the bulk solution properties.

MOLECULAR DYNAMICS SIMULATIONS AND THEORY OF INTERMOLECULAR INTERACTIONS IN SOLUTIONS

by

FENG CHEN

B.S., SiChuan University, China, 1999
M.S., Kansas State University, USA, 2001

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

2010

Approved by

Major Professor
Dr. Paul E. Smith

Copyright

Abstract

In the past several decades, molecular dynamics (MD) simulations have played an important role in providing atomic details for phenomena of interest. The force field used in MD simulations is a critical factor determining the quality of the simulations. Kirkwood-Buff (KB) theory has been applied to study preferential interactions and to develop a new force field. KB theory provides a path from quantities determined from simulation data to the corresponding thermodynamic data. Here we combine KB theory and molecular simulations to study a variety of intermolecular interactions in solution. First, a force field for the computer simulation of aqueous solutions of alcohols is presented. The force field is designed to reproduce the experimentally observed density and KB integrals for a series of alcohols, allowing for an accurate description of alcohols' activity. Other properties such as the translational diffusion constant and heat of mixing are also well reproduced. Second, the newly developed force field is then extended to more complicated systems, such as peptide or mini-proteins, to determine backbone dihedral potentials energetics. The models developed here provide a basis for an accurate force field for peptides and proteins. Third, we have then studied the surface tension of a variety water models. Results showed that different simulation conditions can affect the final values of surface tension. Finally, by using the Kirkwood-Buff theory of solution and surface probability distributions, we attempted to characterize the properties of the Gas/Liquid interface region. The same approach is then used to understand the relationship between changes in surface tension, the degree of surface adsorption or depletion, and the bulk solution properties.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to show unbounded gratitude to my advisor Dr. Paul E. Smith. I am hard pressed to imagine a better role model, in terms of his inimitable enthusiasm, insight, dedication, guidance and instruction. I feel extraordinarily lucky to have had the opportunity to work with and learn from such as advisor who inspires affection.

To members of Dr. Smith's group present and departed, who have supported me through years of graduate study. It makes me feel extraordinarily lucky to have had the opportunity to work with and learn from such group of people who always inspire affection.

I am grateful too, for all of my Ph.D committee members Dr. Christopher Culbertson, Dr. Viktor Chikan and Dr. Om Prakash for their valuable time and efforts.

Finally, I would like to show my appreciation for my family for all their love, support and guidance over the past five and-a-half years. It is them that have kept me going through all the difficulties and not being lost on the way.

Dedication

To my parents

CHAPTER 1 - Introduction

1.1 General Introduction

Despite many years of research, protein folding and unfolding remain as one of the most challenging topics in molecular biology.[1-4] So far, folding studies are based on the properties of native proteins, but detailed information about the folding mechanics may help to light our way to better prediction algorithms. Characterization of the unfolding process is equally important, both from the perspective of fully understanding a fundamental biochemical phenomenon and for providing detailed insights on the folding process. An understanding of protein folding/unfolding would also have an important role for understanding other biological processes, including protein translocation,[5] aging,[6] protein degradation,[7] and variety human diseases.[8] In order to map the folding/unfolding process, one has to characterize every single ensemble state along the whole pathway, from native to denatured state. It is known that the fold/unfolded structure conformation has a strong connection with inter- and intramolecular interactions.[9-11] Such interactions play a crucial role in maintaining protein conformation. Whenever the balance of those interactions is disturbed, proteins themselves may experience misfolding or even denaturation. Many techniques have been designed to study inter- and intramolecular interactions and the ways they affect changes in peptide or protein conformation. One of the early approaches was to determine the effect of solvent on protein folding/unfolding.[12,13] It has been observed that small organic molecules in aqueous solution can have great effects on protein stability, structure, and function. The use of these solvents to stabilize or destabilize proteins is common in today's chemical labs. Among those solvents, water is one of the most studied. Water can affect the manner of protein self-aggregation and equilibrium between the folded and the unfolded states.

Other chemical denaturation agents, such as urea, guanidinium chloride, and lithium perchlorate, have provided ways to investigate protein stability, the effects of mutations, and protein unfolding.

The study of protein folding has been greatly advanced in recent years by the development of fast, time-resolved techniques. Experimental techniques for studying protein folding include : circular dichroism (CD),[14, 15] dual polarisation interferometry,[16] vibrational circular dichroism (VCD),[17,18] fluorescence,[19] infrared,[20] NMR,[21, 22] as well as electron transfer experiments. Despite of all these efforts, our understanding of these biological systems at atomic level is still not sufficient enough to quantitatively describe them due to their internal complexity, as well as the limitations of current experimental methods. The usefulness of molecular dynamics simulations at the present time can hardly be underestimated.[23] It can provide detailed information about the relationships between the bulk properties of solution and the underlying interactions among the particles in the liquid, solid or gaseous state.[23, 24] Therefore, it can provide us with valuable insights about these effects at the atomic level. The increasing power of computers makes it possible to calculate even more accurate data for larger systems. Applications of molecular dynamics simulations can be found in all branches of chemistry as well as physics, chemistry, biochemistry, materials science and pharmaceutical industry.

1.2 Molecular Simulation

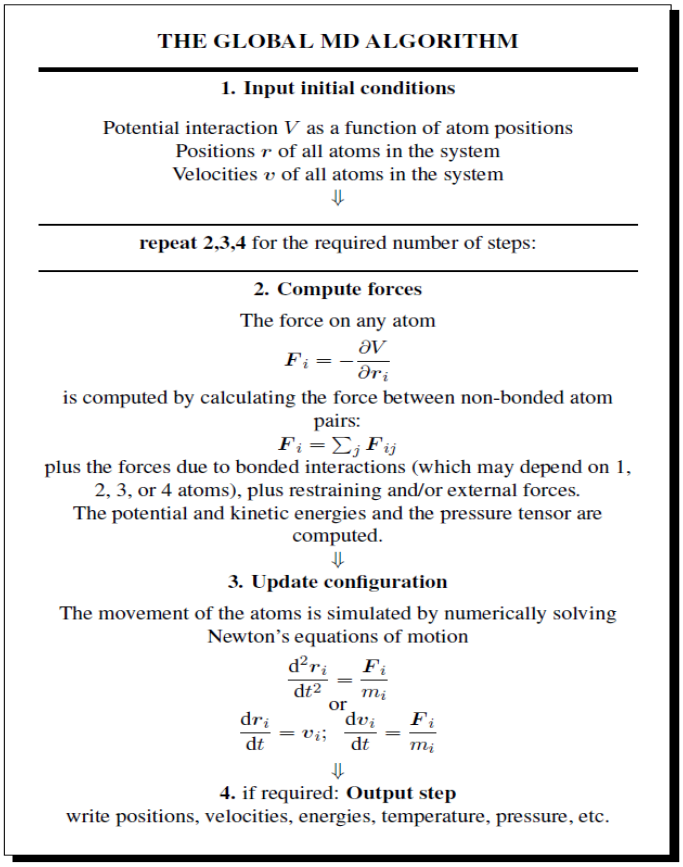
The prologue to the modern age of Molecular Dynamics (MD) simulation can be traced back to half century ago. In 1957, Alder et al.,[25] carried out the first MD study on the properties of hard spheres. Later, the first real system to be simulated involved liquid argon and was performed by Rahman.[26] Afterwards, similar simulation approaches were applied to study

properties of liquid water.[27] Karplus and his colleagues's[28] pioneered work on protein MD simulations and have shown strong indications of a promising future for employing theoretical approaches to investigate biological properties of proteins.

In the world of simulations, several different computational techniques have been developed. Among them, Monte-Carlo (MC) simulations[29,30] and molecular dynamics (MD) [31-34] simulations are the two major approaches. In the case of the Monte-Carlo approach, results depend on calculations using random sampling. Only certain configuration obeying the Boltzmann distribution is accepted, [35] which makes it suitable for most physical and mathematical systems with a large number of degrees of freedom. Molecular dynamics (MD) is a computer simulation approach in which atoms and molecules are allowed to interact for a period of time based upon the laws of physics, providing a view of the motion of the particles. Proteins and biomolecules, as well as molecules in materials science, can be frequently studied by this kind of simulation. In the MD simulation method, the forces applied to the atoms are obtained by evaluating Newton's laws for a short time interval. Within these specific time intervals Newton's equations are solved repeatedly to determine the dynamic properties of the system.[36] Unlike MC studies, MD approach can provide detailed information about particle motions along the whole trajectory of the simulation. The ability to study the dynamical properties of interesting systems makes molecular dynamics simulations a better tool to investigate biological systems.

A global flow scheme for regular molecular dynamics simulation is shown in Figure 1.1.

Figure 1.1 The global MD algorithm[37]



As shown in Figure 1.1, MD simulations are based on empirical energy functions. Newton's equations are solved to determine the motion and coordinates of particles along the folding and unfolding pathway. The forces can be obtained by taking the negative derivative of the potential function with respect to the atomic positions. All equations are solved simultaneously for every single step along the simulation. Solving Newton's equation for particles with a simple force field involves using a classical approach to describe atom motions.

It is worth pointing out that there are several approximations that have to be made when performing molecular dynamics simulations, such as assuming all particles are classical. Even though this is not entirely accurate, classical based molecular dynamics simulations still provide

a powerful tool to study biological system at molecular level.[38,39] They can be used to investigate the properties of a model system, normally more easily than experimental approaches for the same system. The increase in number of simulation packages available makes it easier to run simulation studies on the properties of biological macromolecules, such as peptides and proteins. Hence, the number of publications using molecular dynamics has increased dramatically in the past several years.

1.3 Force Fields for Molecular Dynamics Simulation

A force field[40-42] is made up from two distinct components:1) The set of equations (called the potential functions) used to generate the potential energies and their derivatives, the forces, and 2) the parameters used in this set of equations. The total potential energy is given by:

$$
\begin{aligned}
U = & \sum_{bonds} \frac{1}{2} k_r (r - r_0)^2 && \text{Bond stretches} \\
& + \sum_{angles} \frac{1}{2} k_\theta (\theta - \theta_0)^2 && \text{Angle bending} \\
& + \sum_{torsions} k_\phi [1 + \cos(n\phi - \phi_o)] && \text{Torsional rotation} \\
& + \sum_{improper} \frac{1}{2} k_\xi (\xi_{ijkl} - \xi_0)^2 && \text{Improper torsion} \\
& + \sum_{elec} \frac{q_i q_j}{r_{ij}} && \text{Electrostatic interaction} \\
& + \sum_{LJ} 4\varepsilon_{ij} \left(\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right) && \text{Lennard-Jones interaction}
\end{aligned}
\qquad (1.1)
$$

$$U_{\text{nonbonded}} = U_{\text{vanderWaals}} + U_{\text{electrostatic}} \quad (1.2)$$

$$U_{\text{bonded}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedaral}} \quad (1.3)$$

Values for the force field parameter sets are generally derived from experiments and quantum calculations. Experimentally determined geometries, such as bond angles and bond lengths, are then used to optimize those developed bonded parameters. It is necessary to take into account additional experimental data in the final stage of parameterization. Bonded parameters are usually optimized from experimental data such as gas-phase geometries and vibrational spectra, and torsional energy surfaces supplemented with *ab initio* results. For the optimization of nonbonded parameters, various sources of data can be used, including molecular volumes, experimental enthalpy of mixing, compressibility, density, dipole moments and *ab initio*/QM calculated values. Comparisons between simulations and experimental values are made in the final phase of parameter development, to assure the molecular models reproduce certain physical properties in the correct manner.

The current generation of force fields provides a reasonably good compromise between accuracy and computational efficiency. They are often calibrated to experimental results and quantum mechanical calculations of small model compounds. Their ability to reproduce physical properties measurable by experiment is also tested; these properties include structural data obtained from x-ray crystallography and NMR, dynamic data obtained from spectroscopy, inelastic neutron scattering, and thermodynamic data. The development of a parameter set is a very laborious task, requiring extensive optimization. This is an area of continuing research and many groups have been working over the past two decades to derive functional forms and

parameters for potential energy functions of general applicability to biological molecules. Among the most commonly used potential energy functions are the AMBER,[43] CHARMM,[44] GROMOS[45] and OPLS[46] force fields. The continuing development of force fields remains an intense area of research with implications for both fundamental researches as well as for applied study in areas such as pharmaceutical industry.

There are many ways to improve the quality of force fields. One approach to improve force fields is to introduce explicit polarization into force field. In polarizable models the dynamics of the electronic degrees of freedom are introduced by including nonpairwise additive forces between particles. Unlike traditional force fields, polarizable methods include the effects of changes in environment and should be far more transferable than simple pair potentials.[47,48] Generally speaking, the polarizable many-body forces are represented by induced dipoles, accounting for the changes in the electronic structure of ions and molecules. The advantages over traditional nonpolarizable molecular dynamics simulations are that in the course of the simulations spontaneous polarization can occur because the electronic structure problem is solved more accurately. However, the disadvantages of such polarizable calculations are the short simulation times due to high computational costs compared to nonpolarizable fixed charge methods. Complicated biological systems typically involve a large number of molecules which will require significant computation time. Even rapidly folding small proteins often require tens of microseconds.[49] Such time scales are beyond the limit of typical simulations. To improve the efficiency of simulation studies several methods have been developed, such as continuum solvation model and coarse grained force fields.[50-52] Although these are useful for simulations of biomolecules, these are approximate methods with certain limitations and problems related to parameterization and treatment of ionization effects.

Another way to achieve the same goal is to continue working on the existing version of force fields in terms of developing a better representation of the local environment and pair-wise interactions. For existing classical nonpolarizable force fields there is also still plenty of room for improvement. There is a constant need for improved force fields which better reproduce the available experimental data for a wide range of systems. In particular, issues still arise from unreasonable conformational preferences for protein and inadequate descriptions of solvation effects.[55]

In the past decade, the Smith group has continued working on the development and improvement of simple nonpolarizable united-atom force fields. The goals of their studies are to maintain the delicate balance between solute-solute interactions and solute-solvent (solvation) interactions in solution mixtures by using Kirkwood-Buff theory as a guide.[53-60]

1.4 Kirkwood-Buff Theory

In 1951 Kirkwood and Buff derived a new theory relating the thermodynamic properties of a solution mixture to the molecular distribution functions between the molecules in solution mixture systems.[61] Kirkwood-Buff (KB) theory is one of the most important theories of solutions ever to be developed. In reality, the theory itself uses integrals over radial distribution functions to certain finite distance R. The theory is totally general and valid for any solution mixture over the full range of compositions (Figure 1.1).There is no approximation or limitations involved in KB theory which makes it more suitable for solution mixtures than other theories. Twenty years after the first appearance of KB theory, Ben-Naim developed the KB inversion procedure.[62] For the first time, he outlined how to obtain information on the affinity between a pair of species in the solution mixture from existing experiment data. Since then, many chemists and physicists including Smith, Marcus, Ruckenstein, Shimizu, Hall, Zielkiewicz, Lepori, and others have

continued the trend by using KB theory to the study of a variety of solution mixtures.[63-73] The radial distribution function(rdf) plays a central role in KB theory. A radial distribution function(rdf),[74,75] measures the relative probability of finding an atom at a distance *r* away from a central atom. Consider a system of *N* particles in a volume *V* and at a temperature *T*.[74,75] The probability of finding molecule 1 in dr_1, molecule 2 in dr_2, etc., is given by

$$P^{(n)}(r_1, \dots, r_n)dr_1 \dots dr_N = \frac{e^{-\beta U_N} dr_1 \dots dr_N}{Z_N} \tag{1.3}$$

where $\beta = 1/kT$, U_N is the potential energy and Z_N is the configurational integral. To obtain the probability of finding molecule 1 in dr_1 and molecule *n* in dr_n, irrespective of the remaining *N*-*n* molecules, one has to integrate over the coordinates of molecule *n* + 1 through *N*:

$$P^{(n)}(r_1, \dots, r_n) = \frac{\int \dots \int e^{-\beta U_N} dr_{n+1} \dots dr_N}{Z_N} \tag{1.4}$$

The probability of finding a random atom at distance dr_1 and a random atom in dr_n is:

$$\rho^{(n)}(r_1,...,r_n) = \frac{N!}{(N-n)!} \cdot P^{(n)}(r_1,...r_n) \tag{1.5}$$

In a homogeneous system, the probability of finding a particle can be defined:

$$\frac{1}{V}\int \rho^{(1)}(r_1)dr_1 = \rho^{(1)} = \frac{N}{V} = \rho \tag{1.6}$$

Then, $g^{(n)}$ can be introduced as a correlation function:

$$\rho^{(n)}(r_1,...,r_n) = \rho^n g^{(n)}(r_1,...,r_n) \tag{1.7}$$

From equation 1.4, 1.5 and 1.7, it can be shown that

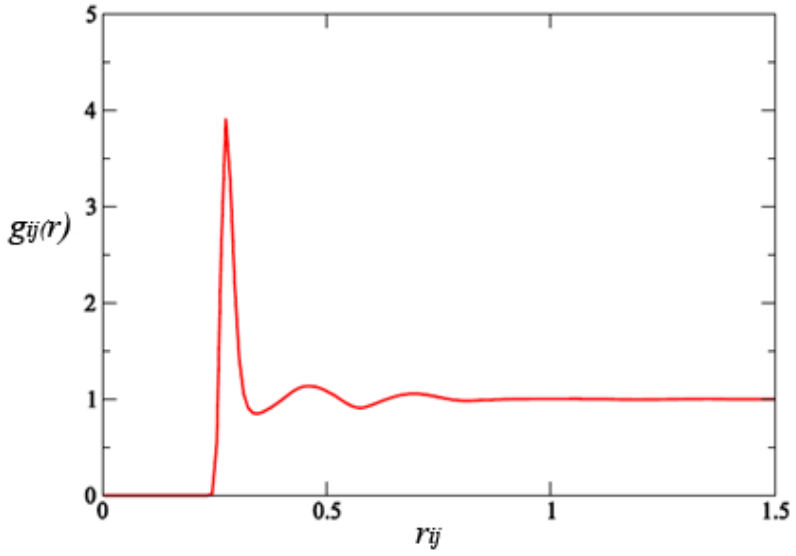
$$g^{(n)}(r_1,...r_n) = \frac{V^n N!}{N^n (N-n)!} \cdot \frac{\int ... \int e^{-\beta U_N} dr_{n+1} ... dr_N}{Z_N} \tag{1.8}$$

Now $\rho g(r) dr$ can be defined as the probability of finding an atom at **r** given that there is an atom at the origin of **r**.

$$\int_0^\infty \rho g(r) 4\pi r^2 dr = N - 1 \approx N \tag{1.9}$$

In Figure 1.2, a sample plot of radial distribution function for Lennard-Jones fluid is provided.

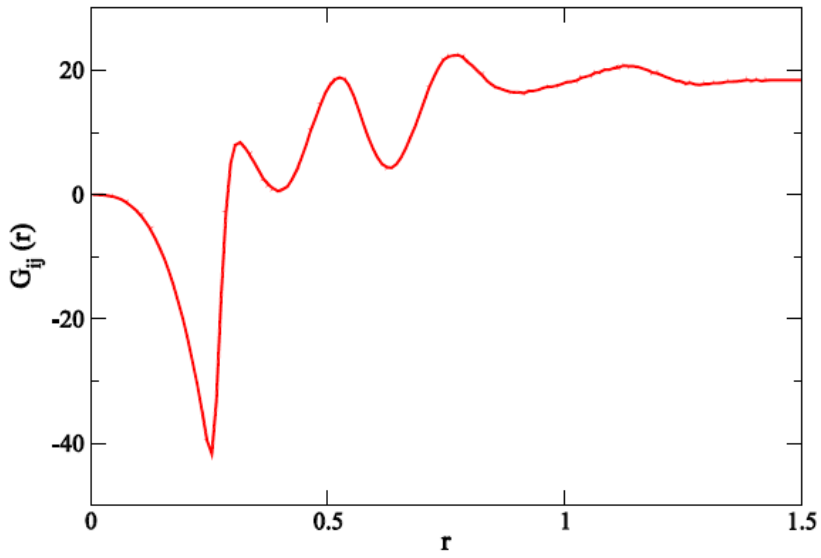
Figure 1.2 Radial distribution function (rdf). The rdf displays the local solution structures for species *i* and species *j* as a function of distance r_{ij}.[78]



Due to a strong repulsive force, *g(r)* has zero value at short distances (less than atomic diameter). In figure 1.2 the first (and large) peak occurs at 0.25 Å. This means that it is four times more likely that two molecules would be found at this separation than expected due to a random distribution. The radial distribution function then falls and passes through a minimum

value around 0.8 Å. The chances of finding two atoms with this separation are less than random. At large distances (beyond 0.8 Å), *g(r)* approaches unity, which indicates closing to bulk solution random distribution. Radial distribution function can also be measured experimentally using X-ray diffraction. The regular arrangement of the atoms in solution gives the characteristic X-ray diffraction pattern with bright, sharp spots. The X-ray diffraction pattern can then be analyzed to estimate an experimental distribution function, which makes it possible to compare with simulated solution *g(r)* values for small molecules.

Figure 1.3 An example of KB integral G_{ij} as a function of integration distance *r* (nm) between species *i* and *j*. This KB integral corresponds to the rdf displayed in Figure 1.2.[78]



The thermodynamic properties of a solution mixture can be expressed using the KB integrals between the different solution components as described as

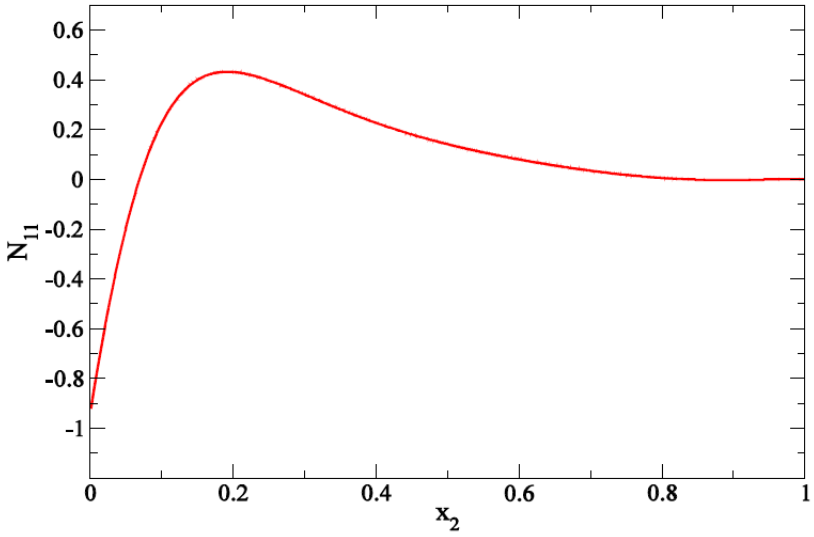
$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}^{\mu VT}(r) - 1 \right] r^2 dr \tag{1.10}$$

where, G_{ij} is the KB integral between species *i* and *j*, $g_{ij}^{\mu VT}(r)$ is the corresponding radial distribution function (rdf) in the μVT ensemble, and r is the center of mass to center of mass distance. KB integrals are determined from typical simulation data (NpT ensemble) by assuming that,[62,79-81]

$$G_{ij}(R) = 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1 \right] r^2 dr \tag{1.11}$$

In equation 1.11, R represents a correlation region within which the solution composition differs from the bulk composition. All rdfs are assumed to be unity beyond a distance R from the central atom. There is no approximation at any level involved during the derivation of the above equations.[71] Previous studies have indicated that a combination of KB theory and NpT simulations can provide quantitative information concerning the thermodynamics of solutions. Excess coordination numbers are defined as $N_{ij} = \rho_j G_{ij}$ and have a simple physical meaning. A value of N_{ij} greater than zero indicates an excess of species j in the vicinity of species i (over a random distribution), while a negative value corresponds to a depletion of species j surrounding i. A sample plot of excess coordination numbers N_{ij} is displayed in Figure 1.4.

Figure 1.4 An example of excess coordination number N_{ij} vs solution conposition.

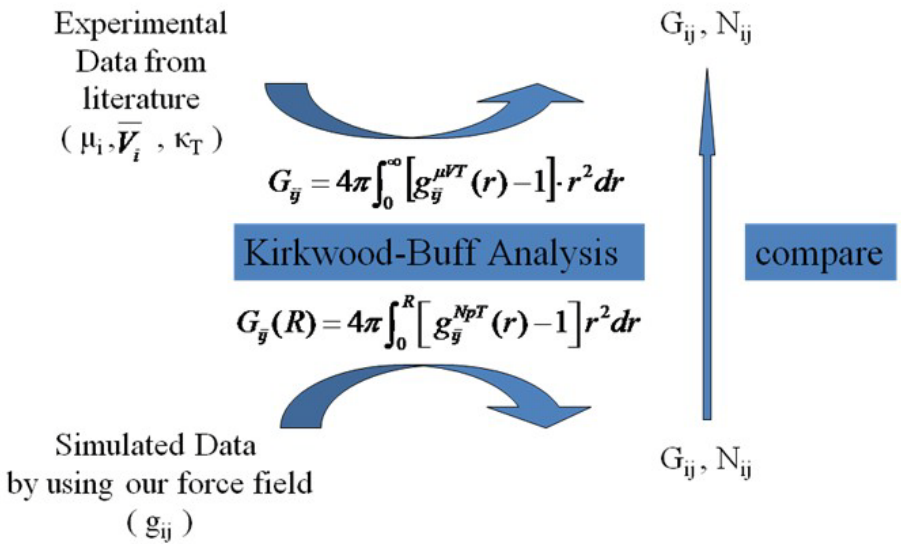


Theoretically speaking, KB theory provides a direct relationship between species self-aggregation (N_{ij}) and activity derivatives and should provide a good test of a particular force field. As noted elsewhere, in practice, a slight larger simulation volume than usual is required in order to ensure that the rdfs approach unity at large distances.[69,70] Simulation studies performed by the Smith group in the past have indicated that a combination of KB theory and NpT simulations can provide quantitative information concerning the thermodynamics of solutions.[54-60]

1.5 Kirkwood-Buff Derived Force Fields

The development of accurate force field for proteins is a central aspect of bimolecular simulation. KB theory has been used to investigate a variety of experimental and theoretical solvation interactions. It is an exact theory of solution mixtures. Most importantly KB theory does not involve any approximations of limitations about the size or character of the molecules to which it can be applied. Therefore, the KB technique can be used to quantify cosolvent effects on peptides and proteins which make it a suitable tool for developing accurate force fields. As shown in Figure 1.5, the KB approach provides a link between simulation data and experimentally measured values.

Figure 1.5 A simple overview of the Kirkwood-Buff approach



Simulation data can then be evaluated by comparing the experimentally measured KB integrals along with thermodynamic properties obtained from them to the corresponding experimental data. Furthermore, the KB integral is one of the most appropriate measures for the molecular affinity which probes the interactions among component pairs. Results have shown, unfortunately, that many existing force fields perform poorly in their ability to reproduce these integrals.[82] Therefore, many of the common force fields currently in use do not necessarily reproduce the correct solution activities, [82] and this can lead to inaccurate simulation data.

Tremendous effort has been focused on the development next generation force fields, including work from the groups of Brooks,[84] Jorgensen,[85] Berenson,[86] Levitt,[87] van Gunsteren[88] and others.[89-91] Despite relatively rapid progress, current available force fields sets are still not perfect. The prediction of properties of interest to biological system in condensed phases is not currently reliable. Most existing force fields also have not been extensively tested under a variety conditions. Therefore, unfortunately, one cannot expect good performance from these force

fields under conditions that have not been considered in their development. This, in turn, severely limits the reliability and predictive ability of molecular modeling methods. Continuous improvement in the quality of force fields has become a long-term goal for theoretical chemists/physicists. A transferable and system independent force field is still in high demand. Over the past several years, the Smith group has been developing a next generation united atom nonpolarizable force field. Kirkwood-Buff (KB) theory has served as a central aspect of their work to help quantify solute-solute and solute-solvent interactions in solution mixtures. As noted elsewhere,[70-73,92,93] KB theory is a powerful theoretical tool to evaluate the ability of a force field to represent the correct relative distribution of molecules in solution.

The development of a KirkWood-Buff parameter set can be traced back to the year 2003. Parameterization and testing of KB force fields for organic molecules and peptides with fix bond lengths and bond angles were described with our previous studies. The van der Waals parameters for hydrocarbons were taken from elsewhere.[94] A list of publications of force fields developed with the aid of Kirkwood-Buff theory are shown as follows:

Urea	**Weerasinghe and Smith, JCP, v118, 5910, 2003**
	Weerasinghe and Smith, JPCB, v107, 3891, 2003
Acetone	**Weerasinghe and Smith, JCP, v118, 10663, 2003**
NaCl	**Weerasinghe and Smith, JCP, v119, 11342, 2003**
GdmCl	**Weerasinghe and Smith, JCP, v121,2180,2004**
CH3OH	**Weerasinghe and Smith, JPCB ,v109,15080,2005**
NMA	**Kang and Smith, JCC, v27, 1477, 2006**
Sulfur	**Bentenitis, Cox, and Smith, JPCB,v113,12306,2009**

The general form of the KB force field involves a Lennard–Jones (LJ) 6-12 plus Coulomb potential, together with the SPC/E water model. The molecular geometries are normally obtained

from the available crystal structures, with bonded parameters mostly taken from the GROMOS96 force field.[45] The force field dependence on the molecular charge distributions of particles is then explored thoroughly by Smith group during the course of KB force field development. Results show the KB integrals are relatively sensitive to the atomic charges.

Currently, the central topic of computational studies of proteins and other biological macromolecules are to solve the protein folding problem. The results from molecular level simulation approaches, however, show significant dependence on the quality of the force fields employed.[98,100] To achieve representative conformational distributions in theoretical studies, appropriate treatment of peptide/protein backbone ϕ/ψ potentials is important. Based on previous studies existing parameter sets such as CHARMM,[95] OPLS,[96] AMBER[97] and GROMOS[45] have shown inherent limitations in reproducing correct structure distributions. High propensities towards π helices in polypeptides/protein solution simulations have been observed.[98] As indicated elsewhere,[99,100] such phenomena are often related to incorrect force field biases. For peptide systems limitations in reproducing high level QM calculated energy surfaces for the Glycine dipeptide and Alanine dipeptides still exist. Overall, the need for an accurate treatment of protein backbone ϕ/ψ potentials is evident.

1.6 A Comparison between the KB derived Force Field and other force fields

There are several advantages of using KB theory in the process of parameterization: 1) more data for testing, 2) ideal probe of thermodynamic properties of solution mixtures and 3) sensitivity to atomic charge distributions. As reported in the literature,[69,71,72,101] several existing force fields do not reproduce activities accurately in solution mixtures. Therefore, they do not provide a correct picture concerning the solvation effects in solution. A significant advantage of KB derived force fields is its ability to find a reasonable representation for the interaction of

cosolvent and solutes. Several comparisons with existing force fields have been made in our previous papers. Here, our KBFF model for N-methylacetamide (NMA) will be compared to other empirical force fields including AMBER, CHARMM, GROMOS, and OPLS. In proteins, the carboxyl and amino groups of neighboring amino acids combine to constitute peptide bond. NMA can serve as a simple model for a peptide group. Hence it is crucial to have an accurate force field for NMA. Glycine is also one of the fundamental building blocks for peptides/proteins, which makes it a good a model for interactions between charged side chains. In Figure 1.6 – 1.10, properties such as the excess coordination numbers, density, partial molar volume, and enthalpy of mixing are compared for a series of force fields.

Figure 1.6 Excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of composition for NMA (2) in water (1) solutions.[67]

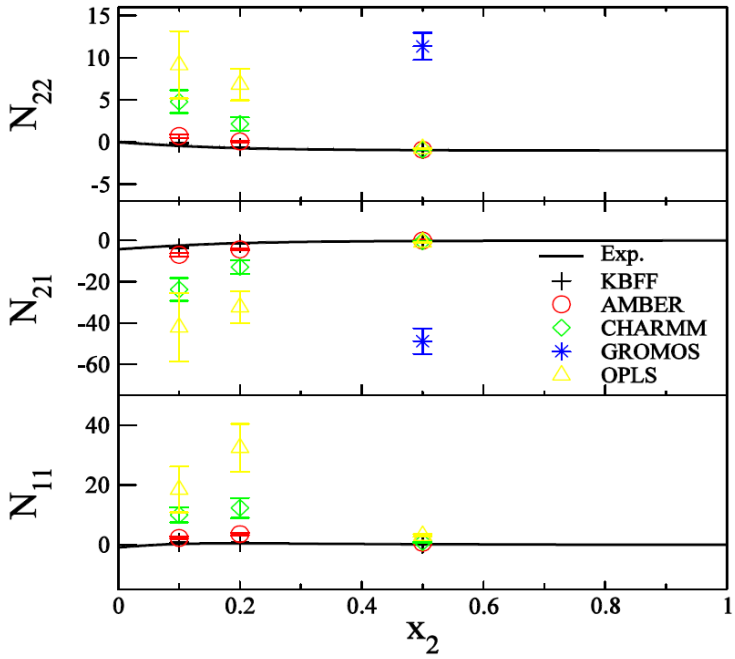


Figure 1.7 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for NMA(2) and water(1) solutions.[67]

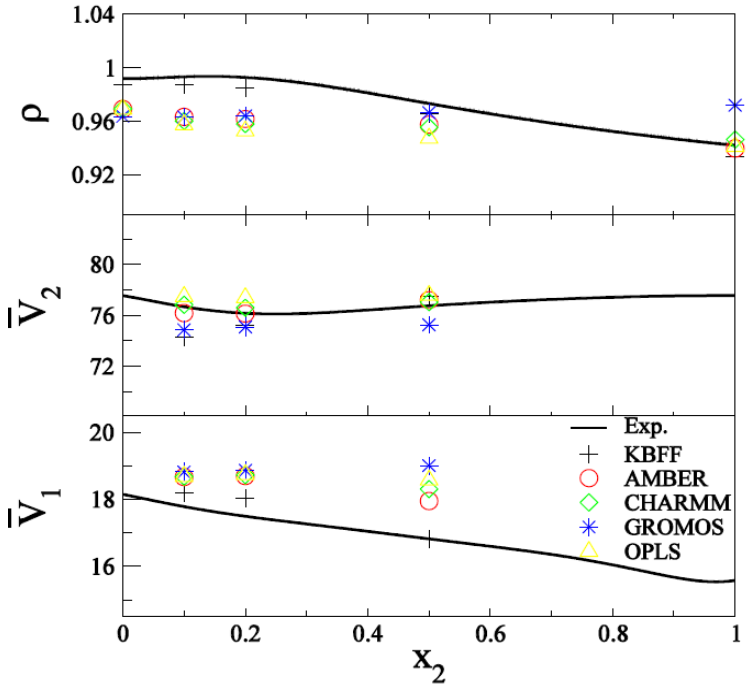
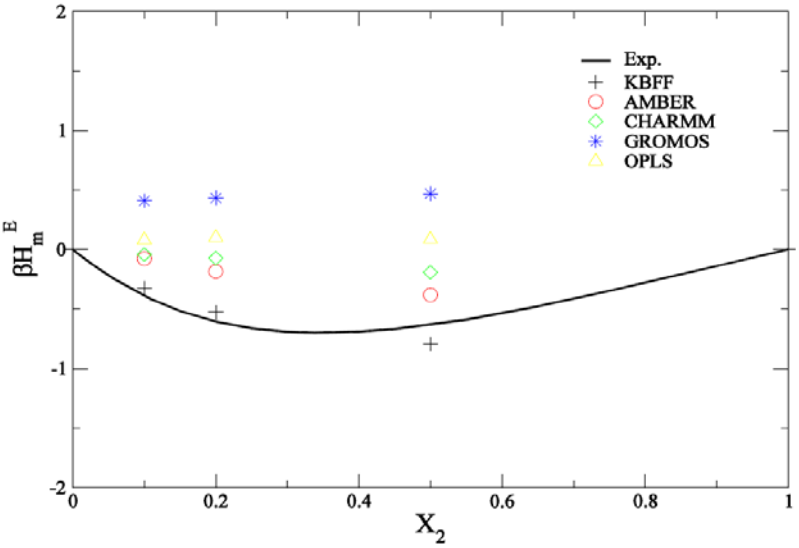


Figure 1.8 Enthalpy of mixing (H_m^E) for NMA solutions. The experimental data is for 308K and the simulations were performed at 313K.[67]



Another comparison was also performed for KBFF glycine solutions. Simulation results for glycine and water mixtures are shown in Figures1.9 and 1.10.

Figure 1.9 Excess coordination numbers (N_{ij}) as a function of composition for glycine (2) in water (1) solutions.[102]

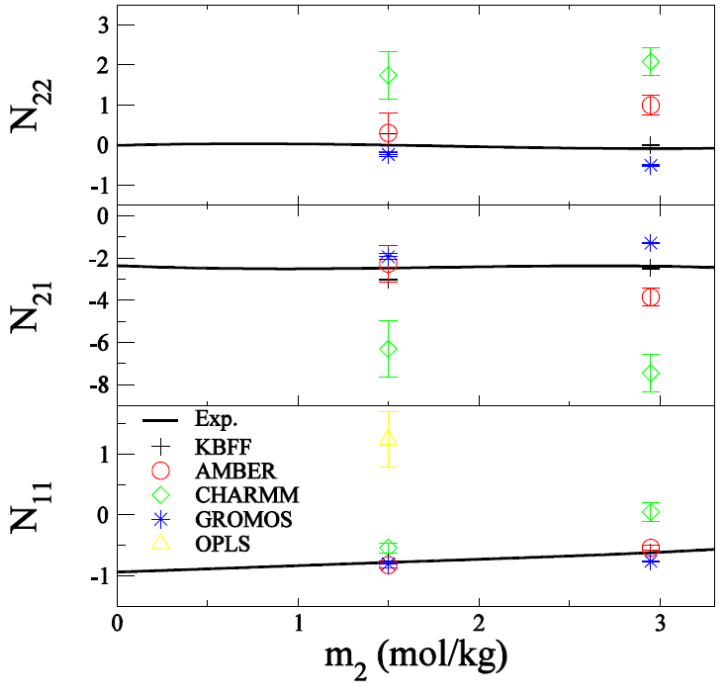
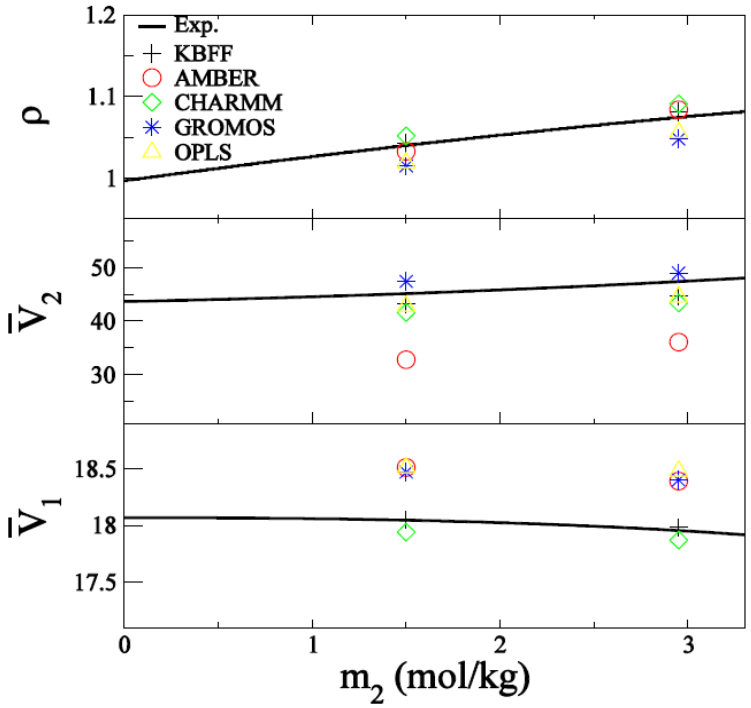


Figure 1.10 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for glycine (2) and water (1) solutions.[102]



It is remarkable how well the experimental values are reproduced with our KB derived force fields. It appears other force fields tested here either overestimate or underestimate certain properties. The KBFF models show the best agreement with experiment values. One should not be so surprised, since our KBFF model was originally designed to reproduce KB integrals while others were not. The reason behind this kind of phenomenon is that it is crucial to obtain correct representations of solvent-solute interactions and molecular distributions in force field development. The results shown here indicate the future possibilities of using the KBFF approach for developing a more accurate and complete force field. In the meantime, it also provided us enough confidence to continue our work on toward a complete KB derived force field for proteins.

1.7 Surface tension of common water models

The unique characteristics on interfacial regions make them important for chemical, physical and biological processes. Thermodynamics and dynamics properties of interfacial region require an understanding of atomic level of information. Therefore, a number of theoretical and experimental methods have been used to help understand surface adsorption and exclusion.[103-107] Even today's modern experimental techniques have a limited ability to explore many of the detailed insights of the interface structure. Classical molecular dynamics (MD) simulations can play an important role in helping our understanding of interfacial structure and to rationalize experimental observations. The majority of empirical force fields, including KBFF, were all developed to study bulk solution properties. As discussed previously,[108] none these models have been fully tested under conditions other than those for which they were initially designed. Surface tension is a crucial property results from the behavior of water at different interfaces. The ability of computer simulations to reproduce the interfacial water behavior

depends on the quality of the water model and force field used. Evidence has shown that several of the previous theoretical studies of the surface tension of SPC/E water have overestimated, for a variety of reasons, the true value by 15%. Here we try to address several numerical issues which may cause conflicting values of SPC/E water from different MD simulations. We then explain the differences between the several previous studies and show that a variety of simulation conditions can affect the final surface tension values. The major issues arising from such studies are that surface structure largely relies on the quality of force field used, and the length scale of simulations to precisely measure surface tension values and surface probability distributions. Another issue involves the polarization effects at the surface region. It has been pointed out that the polarizability of a particle is the driving force for surface solvation.[109-111] The propensity of charged particles towards the aqueous surface is proportional to its polarizability. However, our KB derived model with simple fixed charge does not have polarization effects included. This may limit KB model's ability for predicting surface solvation, since the effects of polarizability and particle size are neglected.

1.8 A Combined Approach for Solution KB theory and Thermodynamics of Surface

The central idea of Kirkwood-Buff Theory is to relate simulation results to solution activity derivatives. Our previous studies have shown remarkable agreement with experimental data for bulk properties of solution mixtures. However, force fields available today, including our KB derived model, were targeted on reproducing bulk solution properties. None of them have been fully tested under conditions other than bulk region. It is still unknown, how well force fields themselves behave when conditions of the simulations are different than what they were initially designed for. Therefore, it is worth investigating the results when one extends the

scope of study into regions other than bulk area. Recently, a wealth of information has been provided by simulations of solutes at the water vapor (or vacuum) interface.[112-117] Interfacial phenomena due to their wide spread in nature have attracted the attention of researchers for a long time. Understanding the detailed insights of interfacial structure and behavior is crucial to many research areas including coating or adsorption, industrial separation, and micellar and membrane systems. Consequently, a variety of experimental and theoretical approaches have been used to help understand surface adsorption or exclusion. It is well established that an increase in the surface tension of a solution due to the addition of a solute indicates exclusion of that solute from the interfacial region and vice versa.[118] Due to technical difficulties, currently available experimental approaches for studying aqueous solution/air interface is lacking at atomic level. In principle, molecular simulations can provide such detail. In the past decades, computer simulations have become an important tool to obtain information on the bulk properties of homogeneous mixtures as well as on their interfacial structure. Nevertheless, theoretical studies which combined bulk solution properties and surface characteristics to achieve a unified picture of solution mixtures hasn't been achieved yet. Here, we have developed a new approach using surface probability distributions to characterize the interface regions, coupled with radial distribution functions and the Kirkwood-Buff theory of solutions to characterize the bulk solution properties. Such an approach is then used to help understand the relationship between changes in the surface tension and the degree of surface adsorption or depletion of two aqueous solution systems.

Paru, Tobias and coworkers have suggested that the uses of polarizable force fields in theoretical studies are essential to catch the correct picture of particles and anions at the surface.[119,120] Hence our use of nonpolarizable Kirkwood-Buff derived models in the current study on surface tension and surface particle probability distribution could be problematic. However, as

results will show, the current version of the KB force fields was capable of reproducing both bulk solution and interfacial thermodynamic data for a system where changes in polarization would be expected to be significant. When a well parametrized force field is applied this suggests that nonpolarizable force fields can be used with enough confidence to study macroscopic and microscopic properties at surface as well.

1.9 Summary

Molecular dynamics (MD) simulations have become powerful tools to investigate the motion of molecules at the atomic level by using classical mechanics. For the first time, people can determine motions and coordinates of particles along the folding and unfolding pathway with atomic detail. The central aspect of our work is to apply Kirkwood-Buff theory to quantify bulk solution and interfacial properties. The ability of the Kirkwood-Buff approach to relate simulation results with experimental data will provide an increased understanding of the thermodynamics and other properties of interested systems. Our long term goal is to achieve a full set of force field parameters to study protein folding and unfolding pathways based on existing KB theory.

In chapter 2, KB theory is applied for the parameterization of a new united atom nonpolarizable force field of a series of primary, secondary and tertiary alcohols. Our goal is to reproduce the experimental density and the Kirkwood-Buff integrals as a function of alcohol mole fractions, which allowing for an accurate description of alcohol cosolvent and water solvent activities. Results have shown significant improvement in simulating properties such as enthalpy of mixing and translational diffusion constants.

In order to complete full Kirkwood-Buff derived force field for proteins, we have continuously worked on developing parameters for protein backbone tensional terms. Based on

our previous work on models for small molecular such as acetone, Urea, NMA etc, φ/ψ backbone dihedral potentials have been determined for several dipeptides. Therefore, eventually one can use our newly developed KB derived force field to look at properties of large system, such as proteins and polypeptides.

In seeking a reason behind the low water surface tension values in MD simulations we have explained the differences between the several previous studies and show that a variety of simulation conditions can affect the final surface tension values. Evidence shows several of the previous theoretical studies of the surface tension of SPC/E water have been overestimated for a variety of reasons. In chapter 4 we try to address several numerical issues which may cause conflicting values of SPC/E water over MD simulations.

The research field of Molecular Dynamics Simulation has been extended from bulk solution to surface. In chapter 5, for the first time, we have combined the surface probability, and the Kirkwood-Buff theory of solutions to quantify the relationship between thermodynamics of surfaces and bulk solution distributions. The approach is then used to understand the relationship between changes in the surface tension, the degree of surface adsorption or depletion, and the bulk solution properties of different aqueous solute systems. The simulated results support the theoretical relationships described here and provide a consistent picture of the thermodynamics of solution interfaces involving any number of components which can be applied to a wide variety of systems.

In chapter 6, a to-do list is provided. The descriptions of future directions of our current projects and a brief summary of present work will also be presented.

Reference List

1. Fersht, A. R.; Daggett, V. Protein folding and unfolding at atomic resolution. *Cell* **2002,** *108* (4), 573-582.

2. Daggett, V.; Fersht, A. The present view of the mechanism of protein folding. *Nature Reviews Molecular Cell Biology* **2003,** *4* (6), 497-502.

3. Daggett, V. Molecular dynamics simulations of the protein unfolding/folding reaction. *Accounts of Chemical Research* **2002,** *35* (6), 422-429.

4. Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-folding dynamics: Overview of molecular simulation techniques. *Annual Review of Physical Chemistry* **2007,** *58*, 57-83.

5. Steiner, J. M.; Loffelhardt, W. Protein translocation into and within cyanelles (Review). *Molecular Membrane Biology* **2005,** *22* (1-2), 123-132.

6. Krankenpfl J, On the track of the enigma of aging. The personal "un-folding".1992, 30(6), 248-256.

7. Ciechanover, A. Intracellular protein degradation: From a vague idea thru the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting. *Experimental Biology and Medicine* **2006,** *231* (7), 1197-1211.

8. Brorsson AC,Kumita JR,Macleod I,Bolognesi B,Speretta E,Luheshi LM,Knowles TP,Dobson CM,Crowther DC. Methods and models in neurodegenerative and systemic protein aggregation diseases. *Front Biosc* **2010**, 15, 373-396.

9. Sapra, K. T.; Besir, S.; Oesterhelt, D.; Muller, D. J. Characterizing molecular interactions in different bacteriorhodopsin assemblies by single-molecule force spectroscopy. *Journal of Molecular Biology* **2006,** *355* (4), 640-650.

10. Samori, B.; Zuccheri, G.; Baschieri, P. Protein unfolding and refolding under force: Methodologies for nanomechanics. *Chemphyschem* **2005,** *6* (1), 29-34.

11. Ma, B. Y.; Tsai, C. J.; Nussinov, R. Binding and folding: in search of intramolecular chaperone-like building block fragments. *Protein Engineering* **2000,** *13* (9), 617-627.

12. Garcia, A. E.; Hillson, N.; Onuchic, J. N. Solvent effects on protein folding/unfolding. *Progress of Theoretical Physics Supplement* **2000,** 138, 282-291.

13. S.M.Vaiana, M.Manno,A.Emanuele,M.B.Palma-Vittorelli and M.U.Palma. The Role of Solvent in Protein Folding and in Aggregation. *Journal of Biological Physics* **2009**, 27(2-3), 133-145.

14. Whitmore, L.; Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: Methods and reference databases. *Biopolymers* **2008,** *89* (5), 392-400.

15. Greenfield, N. J. Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols* **2006,** *1* (6), 2876-2890.

16. Boudjemline, A.; Clarke, D. T.; Freeman, N. J.; Nicholson, J. M.; Jones, G. R. Early stages of protein crystallization as revealed by emerging optical waveguide technology. *Journal of Applied Crystallography* **2008,** *41*, 523-530.

17. Malon, P.; Kobrinskaya, R.; Keiderling, T. A. Vibrational Cd of Polypeptides .12. Reevaluation of the Fourier-Transform Vibrational Cd of Poly(Gamma-Benzyl-L-Glutamate). *Biopolymers* **1988,** *27* (5), 733-746.

18. Yasui, S. C.; Keiderling, T. A. Vibrational Circular-Dichroism of Polypeptides and Proteins. *Mikrochimica Acta* **1988,** *2* (1-6), 325-327.

19. Cremer, C.; Cremer, T. Considerations on A Laser-Scanning-Microscope with High-Resolution and Depth of Field. *Microscopica Acta* **1978,** *81* (1), 31-44.

20. Hamm, P.; Lim, M. H.; Hochstrasser, R. M. Structure of the amide I band of peptides measured by femtosecond nonlinear-infrared spectroscopy. *Journal of Physical Chemistry B* **1998,** *102* (31), 6123-6138.

21. Wuthrich, K. Protein-Structure Determination in Solution by Nmr-Spectroscopy. *Journal of Biological Chemistry* **1990,** *265* (36), 22059-22062.

22. Frydman, L.; Harwood, J. S. Isotropic Spectra of Half-Integer Quadrupolar Spins from Bidimensional Magic-Angle-Spinning Nmr. *Journal of the American Chemical Society* **1995,** *117* (19), 5367-5368.

23. Scheraga, H. A.; Khalili, M.; Liwo, A. Protein-folding dynamics: Overview of molecular simulation techniques. *Annual Review of Physical Chemistry* **2007,** *58*, 57-83.

24. Valerie Daggett ; Protein Folding-Simulation *Chem. Rev.* **2006**, 106(5),1898-1916.

25. Alder, B. J.; Wainwright, T. E. Phase Transition for A Hard Sphere System. *Journal of Chemical Physics* **1957,** *27* (5), 1208-1209.

26. Rahman, A. Correlations in Motion of Atoms in Liquid Argon. *Physical Review A-General Physics* **1964,** *136* (2A), A405-11.

27. Rahman, A.; Stillinger, FH. Molecular Dynamics Study of Liquid Water. *Journal of Chemical Physics* **1971,** *55* (7), 3336-59.

28. Mccammon, J. A.; Gelin, B. R.; Karplus, M. Dynamics of Folded Proteins. *Nature* **1977,** *267* (5612), 585-590.

29. Metropolis, N.; Ulam, S. The Monte Carlo Method. *Journal of the American Statistical Association* **1949,** *44* (247), 335-341.

30. Mosegaard, K.; Tarantola, A. Monte-Carlo Sampling of Solutions to Inverse Problems. *Journal of Geophysical Research-Solid Earth* **1995,** *100* (B7), 12431-12447.

31. Schlick, T. Mathematical Applications to Biomolecular Structure and Dynamics, IMA Volumes in Mathematics and Its Applications. *New York: Springer-Verlag* **1996**, 218-247.

32. Alder, B. J.; Wainwright, T. E. Studies in Molecular Dynamics .1. General Method. *Journal of Chemical Physics* **1959,** *31* (2), 459-466.

33. M.P.Allen; Tildesley,D.J. Computer simulation of liquids. Oxford University 1989.

34. J.A.McCammon, S. C. H. Dynamics of Proteins and Nucleic Acids. *Cambridge University Press*, **1987**.

35. Atkins, P. W.; DePaula. *Physical Chemistry, W.H. Freeman and Company: New York*. **2003**.

36. Adcock, S. A.; McCammon, J. A. Molecular dynamics: Survey of methods for simulating the activity of proteins. *Chemical Reviews* **2006,** *106* (5), 1589-1615.

37. Gromacs User Manual Version 4.0. http://www.gromacs.org/Documentation/Manual: **2009**.

38. Shea, J. E.; Brooks, C. L. From folding theories to folding proteins: A review and assessment of simulation studies of protein folding and unfolding. *Annual Review of Physical Chemistry* **2001,** *52*, 499-535.

39. Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology* **2002,** *9* (9), 646-652.

40. Ponder, J. W.; Case, D. A. Force fields for protein simulations. *Protein Simulations* **2003,** *66*, 27.

41. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochimica et Biophysica Acta-Proteins and Proteomics* **2006,** *1764* (11), 1647-1676.

42. Leckband, D. Force as a probe of membrane protein structure and function. *Current Opinion in Structural Biology* **2001,** *11* (4), 433-439.

43. 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. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.

44. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** *4* (2), 187-217.

45. Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **1998,** *19* (5), 535-547.

46. Jorgensen, W. L.; Tiradorives, J. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* **1988,** *110* (6), 1657-1666.

47. Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *Journal of Chemical Theory and Computation* **2007,** *3* (6), 1960-1986.

48. Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *Journal of Molecular Biology* **1976,** *103* (2), 227-249.

49. Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. The alpha-helix folds on the millisecond time scale. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (13), 7232-7237.

50. Roux, B.; Simonson, T. Implicit solvent models for biomolecular simulation - Preface. *Biophysical Chemistry* **1999,** *78* (1-2), 217.

51. Roux, B.; Simonson, T. Implicit solvent models. *Biophysical Chemistry* **1999,** *78* (1-2), 1-20.

52. Korkut, A.; Hendrickson, W. A. A force field for virtual atom molecular mechanics of proteins. *Proceedings of the National Academy of Sciences of the United States of America* **2009,** *106* (37), 15667-15672.

53. Bentenitis, N.; Cox, N. R.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Thiols, Sulfides, and Disulfides. *Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

54. Smith, P. E. On the Kirkwood-Buff inversion procedure. *Journal of Chemical Physics* **2008,** *129* (12), 124509(1-5).

55. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

56. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

57. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics* **2004,** *121* (5), 2180-2186.

58. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003,** *119* (21), 11342-11349.

59. Weerasinghe, S.; Smith, P. E. Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics* **2003,** *118* (23), 10663-10670.

60. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

61. Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions .1. *Journal of Chemical Physics* **1951,** *19* (6), 774-777.

62. A.Ben-Naim Statistical Thermodynamics for Chemists and Biochemists. Plenum, New York: 1992.

63. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. Recent applications of Kirkwood-Buff theory to biological systems. *Cell Biochemistry and Biophysics* **2008,** *50* (1), 1-22.

64. Smith, P. E. Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophysical Journal* **2006,** *91* (3), 849-856.

65. Guha, A.; Mukherjee, D. Kirkwood-Buff parameter for the solubility of gases in water at high pressure. *Journal of the Indian Chemical Society* **1997,** *74* (3), 195-198.

66. Guha, A.; Ghosh, N. K. Kirkwood-Buff parameters for the binary mixtures and determination of partial structure factor in the long wavelength limit. *Indian*

Journal of Chemistry Section A-Inorganic Bio-Inorganic Physical Theoretical & Analytical Chemistry **1998,** *37* (2), 97-101.

67. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

68. Smith, P. E. On the Kirkwood-Buff inversion procedure. *Journal of Chemical Physics* **2008,** *129* (12), 124509.

69. Weerasinghe, S.; Smith, P. E. Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics* **2003,** *118* (23), 10663-10670.

70. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

71. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003,** *119* (21), 11342-11349.

72. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics* **2004,** *121* (5), 2180-2186.

73. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

74. B.Widom Statistical Mechanics: *A Concise Introduction for Chemists. Cambridge:* **2002**.

75. D.A.McQuarrie. *Statistical Mechanics. Harper Collins Publishers:* **1976**.

76. Wynveen, A.; Bresme, F. Interactions of polarizable media in water: A molecular dynamics approach. *Journal of Chemical Physics* **2006,** *124* (10), 104502.

77. Chandler, David *Introduction to Modern Statistical Mechanics*. Oxford University Press 1987.

78. MYUNGSHIM KANG. *PHD thesis Kansas State University, KS 66506*. **2009**.

79. Chitra, R.; Smith, P. E. Preferential interactions of cosolvents with hydrophobic solutes. *Journal of Physical Chemistry B* **2001,** *105* (46), 11513-11522.

80. Chitra, R.; Smith, P. E. Properties of 2,2,2-trifluoroethallol and water mixtures. *Journal of Chemical Physics* **2001,** *114* (1), 426-435.

81. Weerasinghe, S.; Pettitt, B. M. Ideal Chemical-Potential Contribution in Molecular-Dynamics Simulations of the Grand-Canonical Ensemble. *Molecular Physics* **1994,** *82* (5), 897-912.

82. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

83. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

84. Brooks, C. L. Characterization of Native Apomyoglobin by Molecular-Dynamics Simulation. *Journal of Molecular Biology* **1992,** *227* (2), 375-380.

85. Tiradorives, J.; Jorgensen, W. L. Molecular-Dynamics Simulations of the Unfolding of Apomyoglobin in Water. *Biochemistry* **1993,** *32* (16), 4175-4184.

86. Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. C. A molecular dynamics study of the 41-56 beta-hairpin from B1 domain of protein G. *Protein Science* **1999,** *8* (10), 2130-2143.

87. Daggett, V.; Levitt, M. Protein Unfolding Pathways Explored Through Molecular-Dynamics Simulations. *Journal of Molecular Biology* **1993,** *232* (2), 600-619.

88. Mark, A. E.; Vangunsteren, W. F. Simulation of the Thermal-Denaturation of Hen Egg-White Lysozyme - Trapping the Molten Globule State. *Biochemistry* **1992,** *31* (34), 7745-7748.

89. Pande, V. S.; Rokhsar, D. S. Molecular dynamics simulations of unfolding and refolding of a beta-hairpin fragment of protein G. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (16), 9062-9067.

90. Duan, Y.; Kollman, P. A. Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution. *Science* **1998,** *282* (5389), 740-744.

91. Tsai, J.; Levitt, M.; Baker, D. Hierarchy of structure loss in MD simulations of src SH3 domain unfolding. *Journal of Molecular Biology* **1999,** *291* (1), 215-225.

92. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

93. Bentenitis, N.; Cox, N. R.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Thiols, Sulfides, and Disulfides. *Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

94. Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **1998,** *19* (5), 535-547.

95. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. Charmm - A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *Journal of Computational Chemistry* **1983,** *4* (2), 187-217.

96. Jorgensen, W. L.; Tiradorives, J. The Opls Potential Functions for Proteins - Energy Minimizations for Crystals of Cyclic-Peptides and Crambin. *Journal of the American Chemical Society* **1988,** *110* (6), 1657-1666.

97. 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. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.

98. Best, R. B.; Buchete, N. V.; Hummer, G. Are Current Molecular Dynamics Force Fields too Helical. *Biophysical Journal* **2008,** *95* (9), 4494.

99. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

100. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006,** *65* (3), 712-725.

101. Perera, A.; Sokolic, F. Modeling nonionic aqueous solutions: The acetone-water mixture. *Journal of Chemical Physics* **2004,** *121* (22), 11272-11282.

102. Gee, M. B; Smith, P. E. The figures for glycine are taken from the unpublished data developed by Smith, P. E. and Gee, M. B. **2009**.

103. Pappenheimer, J. R.; Lepie, M. P.; Wyman, J. The Surface Tension of Aqueous Solutions of Dipolar Ions. *Journal of the American Chemical Society* **1936,***58***,** 1851-1855.

104. Onsager, L.; Samaras, N. The surface tension of Debye-H.ovrddot.uckel electrolytes. *Journay of Chemical Physics* **1934**, *2*, 528-536.

105. Richmond, G. L.; Robinson, J. M.; Shannon, V. L. 2Nd Harmonic-Generation Studies of Interfacial Structure and Dynamics. *Progress in Surface Science* **1988,** *28* (1), 1-70.

106. Raymond, E. A.; Richmond, G. L. Probing the molecular structure and bonding of the surface of aqueous salt solutions. *Journal of Physical Chemistry B* **2004,** *108* (16), 5051-5059.

107. Cacace, M. G.; Landau, E. M.; Ramsden, J. J. The Hofmeister series: salt and solvent effects on interfacial phenomena. *Quarterly Reviews of Biophysics* **1997,** *30* (3), 241-277.

108. Zorn, D.; Lin, V. S. Y.; Pruski, M.; Gordon, M. S. An interface between the universal force field and the effective fragment potential method. *Journal of Physical Chemistry B* **2008,** *112* (40), 12753-12760.

109. Perera, L.; Berkowitz, M. L. Many-Body Effects in Molecular-Dynamics Simulations of Na+(H2O)N and Cl-(H2O)N Clusters. *Journal of Chemical Physics* **1991,** *95* (3), 1954-1963.

110. Dang, L. X.; Smith, D. E. Molecular-Dynamics Simulations of Aqueous Ionic Clusters Using Polarizable Water. *Journal of Chemical Physics* **1993,** *99* (9), 6950-6956.

111. Dang, L. X. Characterization of water octamer, nanomer, decamer, and iodide-water interactions using molecular dynamics techniques. *Journal of Chemical Physics* **1999,** *110* (3), 1526-1532.

112. Zorn, D.; Lin, V. S. Y.; Pruski, M.; Gordon, M. S. An interface between the universal force field and the effective fragment potential method. *Journal of Physical Chemistry B* **2008,** *112* (40), 12753-12760.

113. Wilson, M. A.; Pohorille, A. Interaction of Monovalent Ions with the Water Liquid Vapor Interface - A Molecular-Dynamics Study. *Journal of Chemical Physics* **1991,** *95* (8), 6005-6013.

114. Jungwirth, P.; Tobias, D. J. Ions at the air/water interface. *Journal of Physical Chemistry B* **2002,** *106* (25), 6361-6373.

115. Bhatt, D.; Newman, J.; Radke, C. J. Molecular dynamics simulations of surface tensions of aqueous electrolytic solutions. *Journal of Physical Chemistry B* **2004,** *108* (26), 9077-9084.

116. Kuo, I. F. W.; Mundy, C. J.; Eggimann, B. L.; McGrath, M. J.; Siepmann, J. I.; Chen, B.; Vieceli, J.; Tobias, D. J. Structure and dynamics of the aqueous liquid-vapor interface: A comprehensive particle-based simulation study. *Journal of Physical Chemistry B* **2006,** *110* (8), 3738-3746.

117. Chen, B.; Siepmann, J. I.; Klein, M. L. Vapor-liquid interfacial properties of mutually saturated water/1-butanol solutions. *Journal of the American Chemical Society* **2002,** *124* (41), 12232-12237.

118. Moore, W. J. *Physical Chemistry*; Prentice-Hall: Englewood Cliffs, NJ. 1972.

119. Jungwirth, P.; Tobias, D. J. Ions at the air/water interface. *Journal of Physical Chemistry B* **2002,** *106* (25), 6361-6373.

120. Jungwirth, P.; Tobias, D. J. Specific ion effects at the air/water interface. *Chemical Reviews* **2006,** *106* (4), 1259-1281.

CHAPTER 2 - A Kirkwood-Buff Derived Force Field for Alcohols

Feng Chen and Yuanfang Jiao

Abstract

A united atom nonpolarizable force field for the computer simulation of aqueous solutions of a series of alcohols such as ethanol, n-propanol, n-butanol, iso-propanol, tert-butanol, and n-octanol is presented. The force field is designed to reproduce the experimental density and the Kirkwood-Buff integrals as a function of the alcohol mole fraction thereby allowing for an accurate description of alcohol cosolvent and water solvent activities. In addition, the models perform well for other known properties of alcohols including the enthalpy of mixing, and translational diffusion constants.

2.1 Introduction

It is known that the accuracy of the results obtained from simulations studies depend largely on the quality of the force field describing the intermolecular and intermolecular interactions. Therefore, improved force fields which provide better agreement with the available experimental data for a large range of systems are highly desired. Recently, we have been developing a force field (KBFF) for molecular dynamics (MD) simulations which is specifically designed to reproduce the Kirkwood-Buff (KB) integrals obtained from the experimental data for solution mixtures. The KB integrals have been shown to be a sensitive probe of the molecular charge distributions observed for different solutions.[1-4] The KBFF method is primarily aimed at providing accurate force fields for the simulation of solution mixtures. In this study, we have extended our previously developed KBFF methanol model[5] to investigate a series of alcohols, such as ethanol, n-propanol, n-butanol, iso-propanol, tert-butanol, and n-octanol. Computer

simulations are carried out to study the properties of alcohols and water mixtures covering the entire composition range in an attempt to validate the force fields. The present version of the force fields employ a simple nonpolarizable classical approach which can be easily used for other simulation packages. The proposed force field is specifically designed for use with the SPC/E[6] water model.

2.2 Methods

2.2.1 Kirkwood-Buff Theory.

The development of KB theory is described in detail elsewhere.[7-9] The thermodynamic properties of a solution mixture can be expressed in terms of the KB integrals between the different solution components as defined by

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

where, G_{ij} is the KB integral between species *i* and *j*, $g_{ij}^{\mu VT}(r)$ is the corresponding radial distribution function (rdf) in the μVT ensemble, and r is the center of mass to center of mass distance. KB integrals were determined from the present simulation data (NpT ensemble) by assuming that, [10-12]

$$G_{ij}(R) = 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1 \right] r^2 dr \tag{2.2}$$

where R represents a correlation region within which the solution composition differs from the bulk composition. All rdfs are assumed to be unity beyond R.

For a binary solution consisting of water (w) and a cosolvent (c), a variety of thermodynamic quantities can be defined in terms of the KB integrals G_{ww}, G_{cc}, G_{cw}, and the number densities (or molar concentrations) ρ_w and ρ_c . The partial molar volumes of the

components ($\overline{V_i}$); derivatives of the chemical potential (μ); the derivative of the cosolvent activity ($\alpha_c = y_c\rho_c$) and derivatives of the cosolvent mole fraction scale activity coefficients (f_c), at a pressure (p) and a temperature (T) are given by

$$\overline{V_w} = \frac{1+\rho_c\left(G_{cc}-G_{cw}\right)}{\eta}; \qquad \overline{V_c} = \frac{1+\rho_w\left(G_{ww}-G_{cw}\right)}{\eta}; \tag{2.3}$$

$$a_{cc} = \left(\frac{\partial \ln a_c}{\partial \ln \rho_c}\right)_{p,T} = 1+\left(\frac{\partial \ln y_c}{\partial \ln \rho_c}\right)_{p,T} = \frac{1}{1+\rho_c\left(G_{cc}-G_{cw}\right)}; \tag{2.4}$$

$$f_{cc} = \left(\frac{\partial \ln f_c}{\partial \ln x_c}\right)_{p,T} = -\frac{\rho_w x_c\left(G_{ww}+G_{cc}-2G_{cw}\right)}{1+\rho_w x_c\left(G_{ww}+G_{cc}-2G_{cw}\right)} \tag{2.5}$$

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

No approximations are made during the derivation of the above equations.[1] Previous studies have indicated that a combination of KB theory and NpT simulations can provide quantitative information concerning the thermodynamics of solutions.[13-16]

Excess coordination numbers are defined as N. A value of N_{ij} greater than zero indicates an excess of species j in the vicinity of species i (over a random distribution), while a negative value corresponds to a depletion of species j surrounding i. Hence, KB theory provides a direct relationship between alcohol self-aggregation (N_{cc}) and alcohol activity derivatives through eq 2.4 and should provide a good test of a particular force field.

2.2.2 Kirkwood-Buff Analysis of the Experimental Data.

A Kirkwood-Buff analysis of the experimental data for alcohol cosolvent (c) and water (w) mixture at 298 K and 1atm was performed using the available activity coefficients[17] and density data.[18,19] The KB integrals can be obtained from experimental data on the chemical

potential (μ_i), partial molar volumes ($\overline{V_i}$), and isothermal compressibilities (κ_T) of the binary mixtures at constant pressure (p) and temperature (T) according to,[20, 21]

$$G_{ij} = RT\kappa_T - \frac{\overline{V_i}\overline{V_j}}{(1+f_{cc})V_m} \tag{2.6}$$

and

$$G_{ii} = G_{ij} + \frac{1}{x_i}\left(\frac{\overline{V_j}}{1+f_{cc}} - V_m\right) \tag{2.7}$$

where R is the gas constant, x_i is the mole fraction of i, $V_m = V/(N_c+N_w)$ is the molar volume, and

$$\beta\left(\frac{\partial\mu_c}{\partial\ln x_c}\right)_{p,T} = 1+\left(\frac{\partial\ln f_c}{\partial\ln x_c}\right)_{p,T} = 1+f_{cc} \tag{2.8}$$

with $(\beta = 1/RT)$ and f_c equal to the cosolvent activity coefficient on the mole fraction scale with the pure cosolvent solution as the standard state.

Partial molar volumes were determined from the experimental density data by calculating the excess molar volume,

$$X_m^E = X_m - x_c X_{m,c}^0 - x_w X_{m,w}^0 \tag{2.9}$$

where X is the volume (V) of the solution and $V_{m,i}^0$ is the molar volume of the pure i. The raw data were then fitted to a Redlich-Kister equation,[22]

$$X_m^E = x_c x_w \sum_{i=0}^{n} a_i (x_w - x_c)^i \tag{2.10}$$

where a_i are fitting constants. The partial molar volumes at any composition are then given by,

$$Y_i = X_m^E - x_j\left(\frac{\partial X_m^E}{\partial x_j}\right)_{p,T} \tag{2.11}$$

with $Y = \overline{V_m^E}$ and $X = V$. In general, the KB integrals are not sensitive to the exact values of the isothermal compressibility and therefore the following approximate expression was used,

$$\kappa_T = \phi_c \kappa_{T,c}^0 + \phi_w \kappa_{T,w}^0 \qquad (2.12)$$

where $\varphi_i = \rho_i \overline{V_i}$ is the volume fraction of *i* in the solution. Isothermal compressibilities for the pure solutions ($\kappa_{T,i}^0$) were taken from the literature.

The excess molar Gibbs energy (G_m^E) was obtained by assuming the form given in eq 2.10 with $X = \beta G$, and then fitting the excess chemical potential ($\beta\mu_i^E = \ln f_i$) of both alcohol and water to the experimental data using eq 2.11 with $Y_i = lnf_i$ and $X = \beta G$. The resulting data are in agreement with previous determinations of the excess and partial molar volumes of alcohol and water,[23-25] and a previous determination of the KB integrals for regions where the KB integrals are statistically reliable.[26]

2.3 Molecular Dynamics Simulations

All molecular dynamics simulations were performed using our newly developed KBFF force fields together with the SPC/E water model as implemented in the GROMACS 4.0.5 package.[27-30] All simulations were performed in the isothermal isobaric ensemble at 300K and 1 atm. The weak coupling technique[31] was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively. A time-step of 2 fs was used and the bond length were constrained using Lincs,[32] while the water geometry was constrained using SETTLE.[33] The particle mesh Ewald technique[34] was used to evaluate electrostatic interactions. A real space convergence parameter of 3.5 nm^{-1} was used in combination with twin range cutoffs of 0.8 and 1.5 nm, and a nonbonded update frequency of 10 steps. Random initial configurations of

molecules in a cubic box were used. Initial configurations of the different solutions were generated from a cubic box (L≅6.0 nm) of equilibrated water molecules by randomly r eplacing waters with alcohol until the required concentration was attained. The steepest descent method was then used to perform 100 steps of minimization. This was followed by extensive equilibration, which was continued until all intermolecular potential energy contributions and rdfs displayed no drift with time (typically 15 ns). Total simulation times were in the 20-35 ns range, and the final 15-30 ns were used for calculating ensemble averages. Configurations were saved every 0.1 ps for the calculation of various properties.

Translational self-diffusion constants (D_i) were determined using the mean square fluctuation approach,[35] and excess enthalpies of mixing (ΔH_m^E) from the average potential energies (ΔE_{pot}). Errors ($\pm 1\sigma$) in the simulation data were estimated by using five or six block averages.

2.4 Parameter Development

The force field used in this study corresponded to the Lennard-Jones (LJ) 6-12 plus Coulomb potential, which is the most commonly used potential for biomolecular simulation. In this scheme each pair of atoms *i* and *j* interact with an interaction energy given by

$$V_{ij} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \qquad (2.13)$$

Here, we will use the same approach based on our previous study to obtain LJ parameters.[5] United atom carbon group parameters were taken from the literature.[36] The molecular geometry was taken from the OPLS force field.[41] The charges on the atoms were then adjusted to best reproduce the density and KB integrals for KB integrals for solution mixtures with x_c = 0.125, 0.250, 0.375, 0.50, 0.625, 0.75, and 0.875. The final parameters are presented in the Table 2.1.

TABLE 2.1: Nonbonded Force Field Parameters Used in the Simulations

Model	atom	ε,kJ/mol	σ, nm	q, |e|
ETOH	O	0.6506	0.3192	-0.82
	H	0.0880	0.1580	0.52
	CH_3	0.8672	0.3748	0.30
	CH_2	0.4105	0.4070	0
n-PROH	O	0.6506	0.3192	-0.82
	H	0.0880	0.1580	0.52
	CH_2	0.8672	0.3748	0.30
	CH_2	0.4105	0.4070	0
	CH_3	0.4105	0.4070	0
n-BUOH	O	0.6506	0.3192	-0.82
	H	0.0880	0.1580	0.52
	CH_2	0.8672	0.3748	0.30
	CH_2	0.4105	0.4070	0
	CH_2	0.4105	0.4070	0
	CH_3	0.4105	0.4070	0
i-PROH	CH_1	0.0949	0.5019	0.33
	O	0.6506	0.3192	-0.902
	H	0.0880	0.1580	0.572
	CH_3	0.4105	0.4070	0.00
	CH_3	0.4105	0.4070	0.00
t-BUOH	C	0.4170	0.3770	0.36
	O	0.6506	0.3192	-0.984
	H	0.0880	0.1580	0.624
	CH_3	0.4105	0.4070	0.00
	CH_3	0.4105	0.4070	0.00
	CH_3	0.4105	0.4070	0.00
Water				
SPC/E	O	0.6506	0.3166	-0.8476
	H	0	0	0.4238

[a] SPC/E parameters were taken from ref [37]. Combination rules used are: $\sigma_{ij}=\sqrt{\sigma_{ij}\sigma_{ij}}$ and $\epsilon_{ij}=\sqrt{\epsilon_{ij}\epsilon_{ij}}$

TABLE 2.2: Bonded Force Field Parameters Used in the Simulations

bonds	r_0		r_0
CH_2-O	0.1430	CH_2-CH_2	0.1530
CH_2-CH_3	0.1530	CH_2-CH_3	0.1530
O-H	0.0945	CH_1-CH_3	0.1530
CH_1-O	0.1430	C-O	0.1430
C-CH_3	0.1530		
Angles		k_θ	θ_0
ETOH	O-CH_2-CH_3	520.0	109.5
	H-O-CH_2	450.0	109.5
n-PROH	O-CH_2-CH_3	520.0	109.5
	H-O-CH_2	450.0	109.5
	CH_2-CH_2-CH_3	530.0	111.0
n-BUOH	O-CH_2-CH_3	520.0	109.5
	H-O-CH_2	450.0	109.5
	CH_2-CH_2-CH_2	530.0	111.0
	CH_2-CH_2-CH_3	530.0	111.0
i-PROH	O-CH_1-CH_3	530.0	111.0
	H-O-CH_1	450.0	109.5
	CH_3-CH_1-CH_3	530.0	111.0
t-BUOH	O-C-CH_3	610.0	108.0
	H-O-C	450.0	109.5
	CH_3-C-CH_3	530.0	112.0

Dihedrals	k_ψ	δ	*n*
CH_3-CH_2-O-H	0.85	0.0	1
	0.40	0.0	2
	3.00	0.0	3
CH_3-CH_2-CH_2-O	2.55	0.0	1

	1.20	0.0	2
	9.00	0.0	3
CH_2-CH_2-CH_2-O	2.55	0.0	1
	1.20	0.0	2
	9.00	0.0	3
CH_3-CH_2-CH_2-CH_2	5.00	0.0	1
	2.00	0.0	2
	7.00	0.0	3
CH_3-CH_1-O-H	0.85	0.0	1
	0.40	0.0	2
	3.00	0.0	3
CH_3-C-O-H	0.85	0.0	1
	0.40	0.0	2
	3.00	0.0	3

Impropers	k_ω	ω_o
CH_1-O-CH_3-CH_3	334.8	35.26

The force fields for alcohols including charge distributions were optimized in order to achieve the best reproduced values of KB integrals for alcohol solutions across the whole range of concentration. The last column of Table 2.1 shows the final charge distributions obtained for each alcohol in present study. Our KB force field is focusing on reproducing the thermodynamic properties of solution mixtures such as enthalpy of mixing and excess Gibbs free energy of the solution.[38] Therefore, effort has been made on pursuing the effective charge distributions for alcohols in polar solvent. The objective of the charge tuning process is to ensure the correct balance of alcohol-alcohol and alcohol-solvent interactions. The difference between the charges on central carbon of primary and secondary alcohol is 0.03*e*. Moreover, the central carbon charge difference between primary and tertiary alcohol is 0.06*e*. The higher central carbon

charges on secondary/tertiary alcohols suggests that hydrogen bonding is less likely between primary alcohol molecules compared to molecules of secondary/tertiary alcohols. In the whole process of parameter development, approximately 15 charge distributions were tested.

2.5 Kirkwood-Buff Analysis of the Simulated Data

The KB integrals obtained from an analysis of the experimental data correspond to integrals over rdfs in the μVT ensemble. The simulated KB integrals were obtained by assuming that,

$$G_{ij} = 4\pi\int_0^{\infty}\left[g_{ij}^{\mu VT}(r)-1\right]r^2dr \approx 4\pi\int_0^{R}\left[g_{ij}^{NpT}(r)-1\right]r^2dr \qquad (2.14)$$

where R is a cutoff distance at which the rdfs are essentially unity, i.e. the bulk solution values. In practice, this condition is difficult to achieve precisely unless one uses very large systems. However, a reasonable approximation is to determine $G_{ij}(R)$ and average the values over a short distance range, typically one molecular diameter. In this work the final KB integral values have been obtained by averaging between 1.50 and 2.00 nm. Future justification for this approximation can be found in the result section. Using the simulated KB integrals one can determine the corresponding partial molar volumes,

$$\overline{V_i} = \frac{1+\rho_j\left(G_{jj}+G_{ij}\right)}{\rho_i+\rho_j+\rho_i\rho_j\Delta G} \qquad (2.15)$$

and activity derivative,

$$f_{cc} = \left(\frac{\partial \ln f_c}{\partial \ln x_c}\right)_{p,T} = -\frac{x_c\rho_w\Delta G}{1+x_c\rho_w\Delta G} \qquad (2.16)$$

where $\Delta G = G_{cc} + G_{ww} - 2G_{cw}$. The parameters used to describe the variation of the simulation G_m^E with composition (eq 10) were then obtained from the simulated values of f_{cc} and the thermodynamic relationship,

$$f_{cc} = x_c x_w \beta \left(\frac{\partial^2 G_m^E}{\partial x_c^{\,2}} \right)_{p,T} \tag{2.17}$$

The isothermal compressibility was not determined from the KB integrals as it is typically statistically unreliable.

2.6 Results and Discussion

A summary of the simulations performed is presented in Table 2.3. They cover the entire composition range and all mixtures were simulated up to 30 ns to ensure reasonable precision in the data.

TABLE 2.3: Summary of the Alcohol and Water Simulations

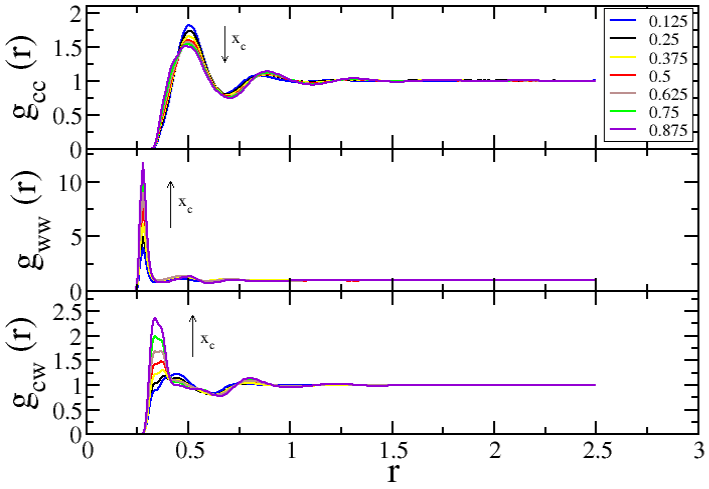
x_c	N_c	N_w	V, nm^3	ρ_c, M	ρ, g/cm^3	E_{pot}, kJ/mol	T_{sim}, ns
Ethanol							
0.0	0	2000	60.16	0.0	0.990	-46.45	30
0.125	702	917	211.834	5.50	0.947	-46.29	30
0.25	1152	3456	212.379	9.00	0.901	-45.60	30
0.375	1465	2441	214.165	11.36	0.864	-44.88	30
0.50	1695	1695	216.093	13.06	0.834	-44.20	30
0.625	1871	1122	217.894	14.26	0.810	-43.54	30
0.75	2010	670	219.886	15.18	0.790	-42.92	30
0.875	2123	303	222.022	15.87	0.772	-42.26	30
1.0	2216	0	224.542	16.38	0.754	-41.46	30
n-Propanol							
0.0	0	2000	60.16	0.0	0.995	-46.45	30

0.125	646	4522	214.884	4.99	0.929	-45.77	30
0.25	1006	3019	216.577	7.71	0.880	-44.85	30
0.375	1236	2060	217.981	9.41	0.848	-43.96	30
0.50	1395	1395	219.099	10.57	0.825	-43.12	30
0.625	1512	907	220.151	11.40	0.808	-43.35	30
0.75	1602	534	221.279	12.02	0.794	-41.65	30
0.875	1672	239	222.302	12.48	0.782	-41.01	30
1.0	1730	0	223.816	12.83	0.771	-40.28	30
n-Butanol							
0.0	0	2000	60.16	0.0	0.995	-46.45	30
0.50	1188	1188	221.111	8.92	0.822	-41.82	30
0.625	1271	763	221.697	9.51	0.808	-40.77	30
0.75	1334	445	222.374	9.96	0.798	-39.80	30
0.875	1383	197	223.095	10.29	0.789	-38.92	30
1.0	1422	0	223.9738	10.54	0.781	-38.05	30
i-Propanol							
0.0	0	2000	60.16	0.0	0.995	-46.45	30
0.25	996	2989	216.536	7.64	0.872	-47.04	30
0.50	1376	1376	221.611	10.31	0.805	-47.07	30
0.75	1576	525	225.115	11.63	0.768	-47.37	30
1.0	1700	0	228.505	12.36	0.742	-48.19	30
t-Butanol							
0.0	0	2000	60.16	0.0	0.995	-46.45	30
0.25	873	2619	212.347	6.83	0.875	-48.41	30
0.50	1151	1151	214.856	8.90	0.820	-49.78	30
0.75	1288	429	216.988	9.86	0.790	-51.20	30
1.0	1370	0	216.664	10.50	0.778	-53.48	30

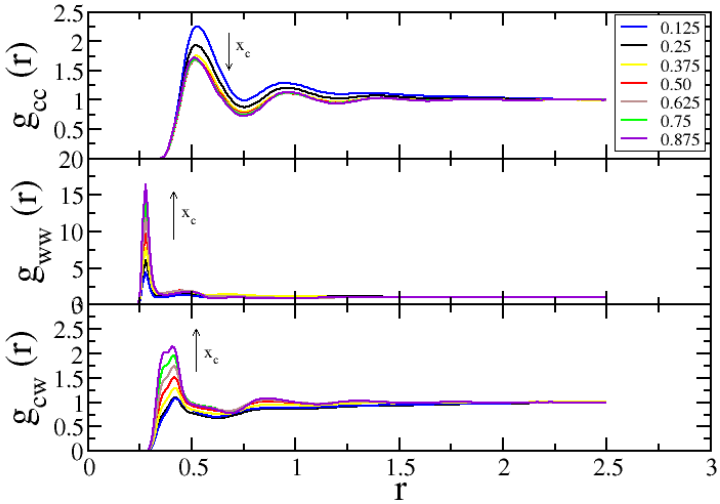
The center of mass rdfs are displayed in Figure 2.1 as a function of composition. As it would be expected as the carbon chain length increases, the first salvation shell (the first peak) of

g_{cc} shifted from 0.5 to 0.55 due to the radius change of alcohol molecular. For all three alcohols, the first maximum increased and the first minimum decreased with alcohol mole fraction. All rdfs were essentially unity beyond 1.5 nm. The most prominent feature was an increase in the first shell hydration for water to water suggesting an increasing degree of water self-association with increasing alcohol mole fraction in agreement with neutron diffraction data.[42] The corresponding first shell coordination numbers are displayed in Figure 2.1. All three first shell coordination numbers displayed a linear dependence on alcohol (or water) mole fraction as observed previously.[43,44] The values in pure water were 5.1 compared to 2.0 in pure alcohol. The experimental value for pure alcohol is 1.9 at 0.34 nm.[39, 40]

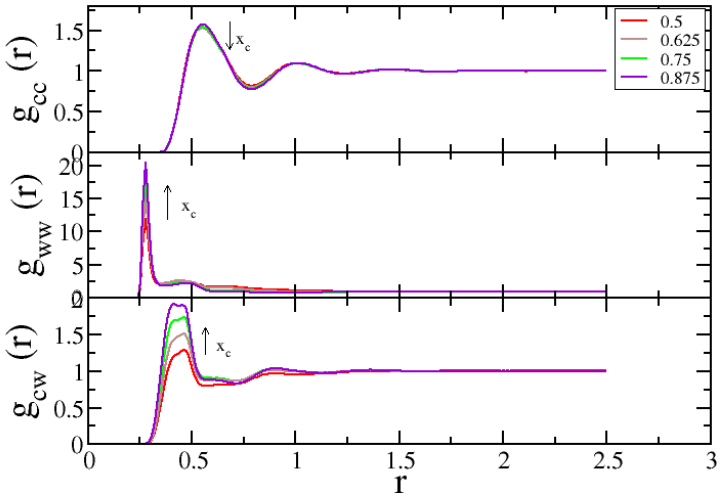
Figure2.1 (a), radial distribution functions as a function of distance (nm) and ethanol mole fraction. (b), radial distribution functions as a function of distance (nm) and n-propanol mole fraction. (c), radial distribution functions as a function of distance (nm) and n-butanol mole fraction. (d), radial distribution functions as a function of distance (nm) and i-propanol mole fraction. (e), radial distribution functions as a function of distance (nm) and t-butanol mole fraction. Mole fractions of 0.125, 0.250, 0.375, 0.50, 0.625, 0.75, and 0.875 for ethanol and n-propanol are displayed. n-Butanol's concentration is from 0.5 to 1, due to the low solubility in water at lower concentrations. Mole fractions of 0.25, 0.50, and 0.75 are displayed for i-propanol and t-butanol.



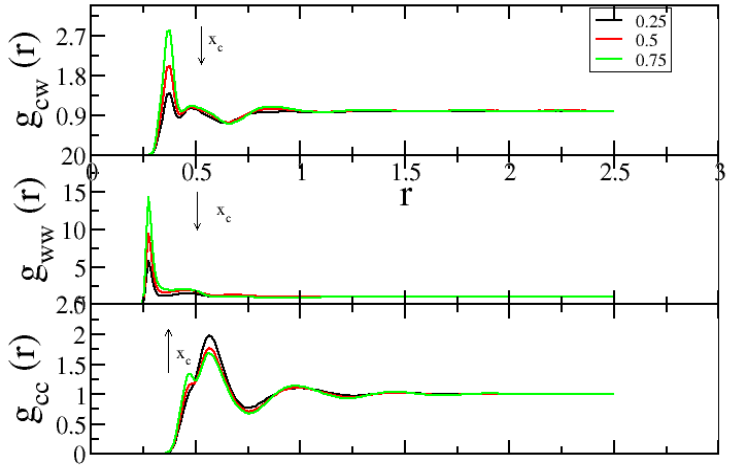
a)



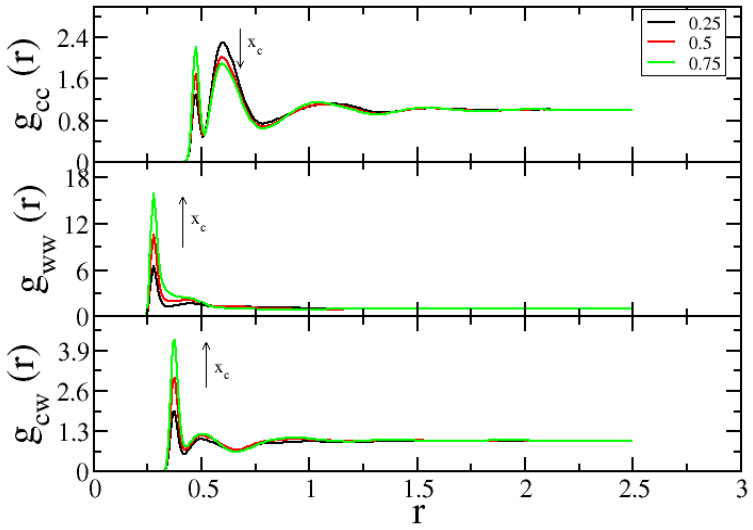
b)


0.5
0.625
0.75
0.875
g_{cc} (r)
g_{ww} (r)
g_{cw} (r)
x_c
r

c)

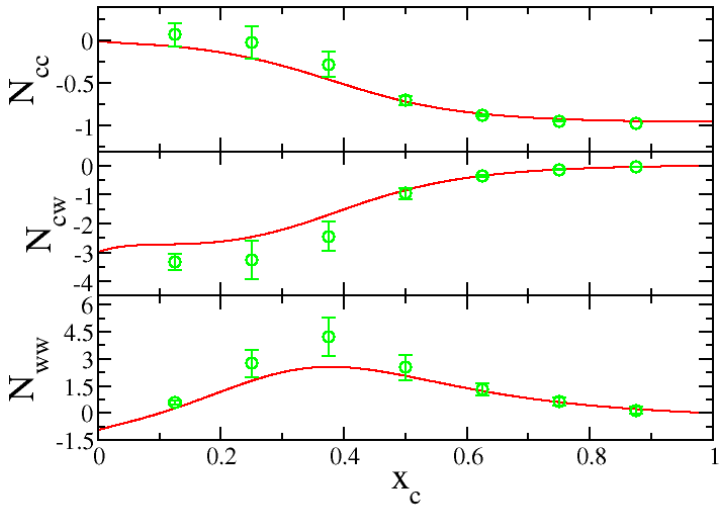

0.25
0.5
0.75
g_{cw} (r)
g_{ww} (r)
g_{cc} (r)
x_c
r

d)

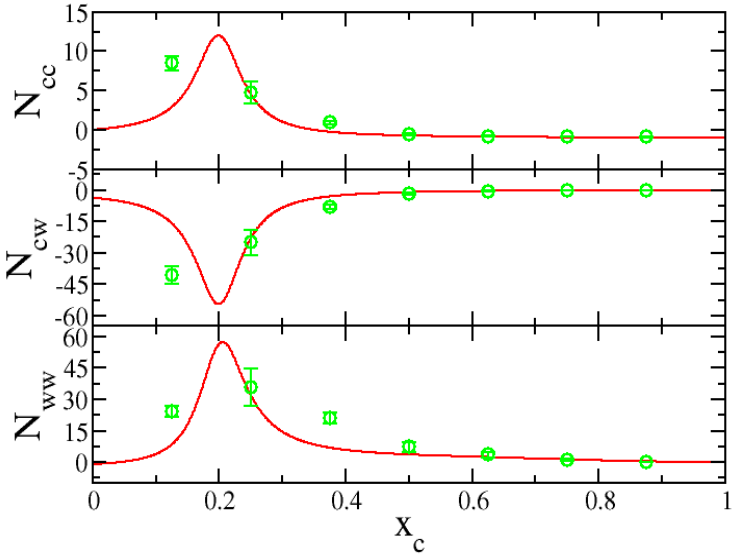


e)

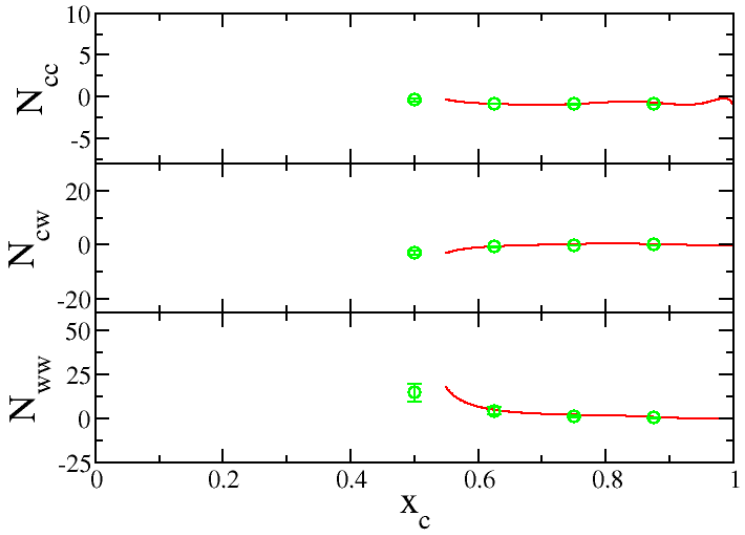
Figure 2.2 a). excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of ethanol mole fraction. b). excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of n-propanol mole fraction. c). excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of n-butanol mole fraction. d). excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of i-propanol mole fraction. e). excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) as a function of t-butanol mole fraction. The solid lines correspond to the experimental data, the circles to the raw simulation data.



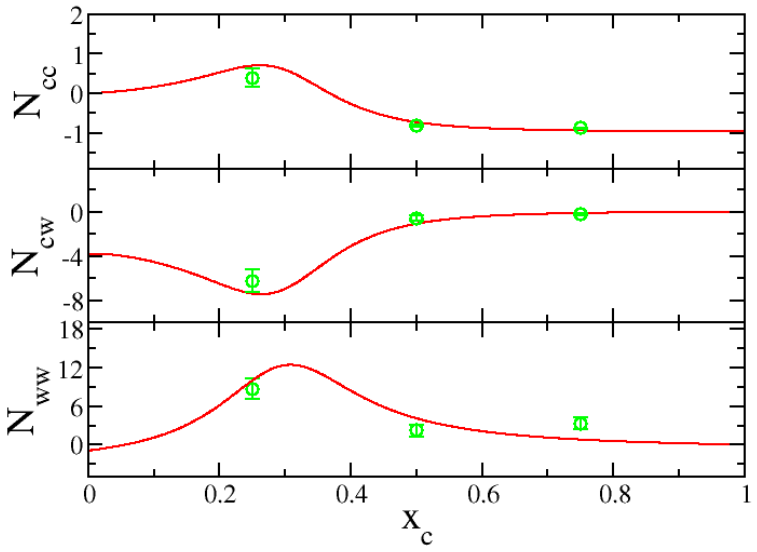
a)



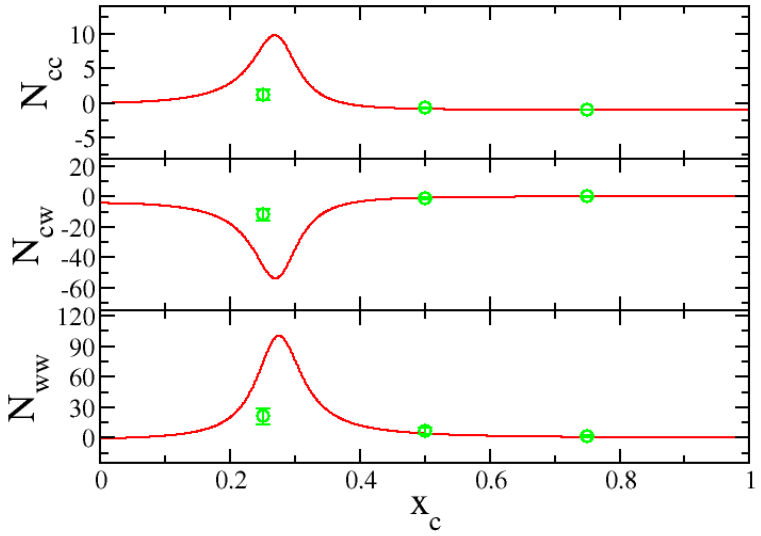
b)



c)



d)



e)

The experimental and simulated KB integrals are compared in Figure 2.2 as excess coordination numbers ($N_{ij} = \rho_j G_{ij}$). The use of excess coordination numbers helps to suppress the inherent uncertainties in both the experimental and simulated G_{ij} integrals at low j concentrations. The trends in the experimental data were well reproduced. There was essentially quantitative agreement for N_{cc}, N_{cw} and N_{ww} over the composition range from 0.25 to 1, whereas N_{ij} of n-propanol was slightly off when x_c under 0.25.

The experimental and simulated densities are compared in Figure 2.3. The density of the pure alcohol solution was slightly underestimated and hence one observes a gradually increasing deviation from the experimental density with increasing x_c.

Figure 2.3 a). The density (g/cm^3) as a function of ethanol mole fraction. b). The density (g/cm^3) as a function of n-propanol mole fraction. c). The density (g/cm^3) as a function of n-butanol mole fraction.

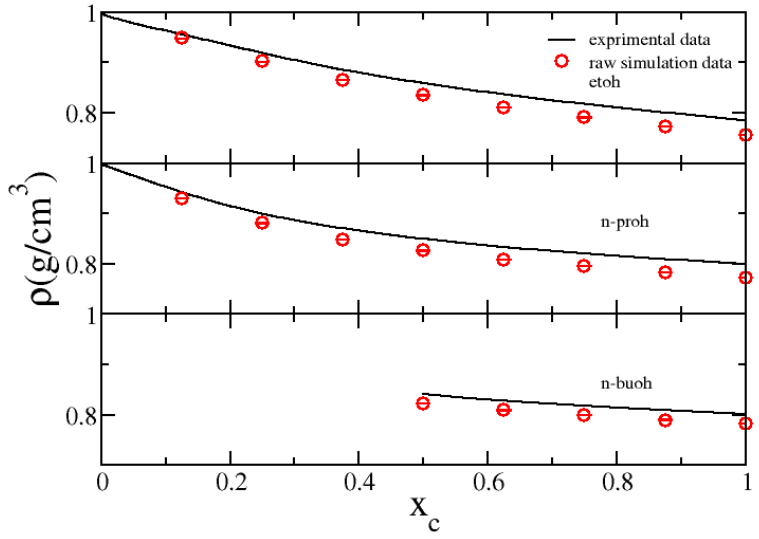
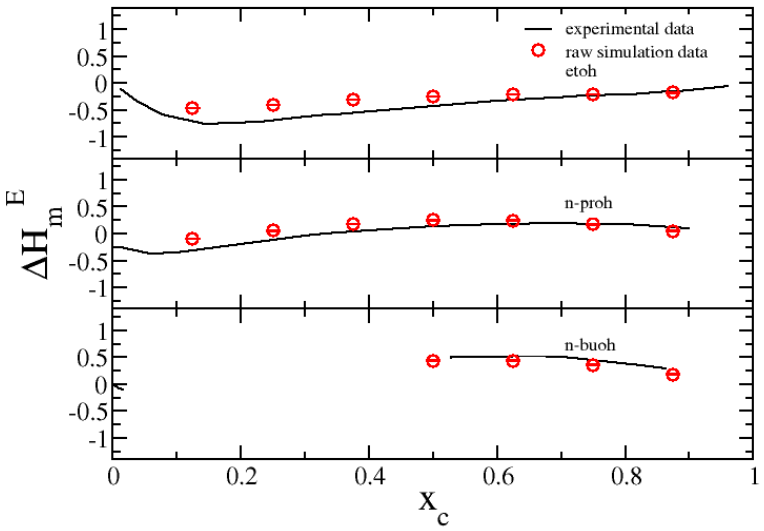
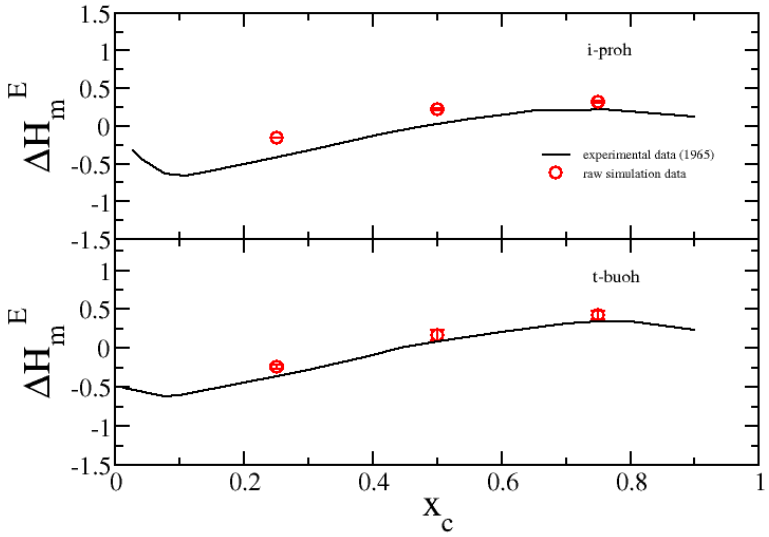


Figure 2.4 The excess molar enthalpy of mixing (H_m^E in kJ/mole) as a function of methanol mole fraction for ethanol, n-propanol, n-butanol, i-propanol and t-butanol. The solid lines correspond to the experimental data, and the circles to the raw simulation data.

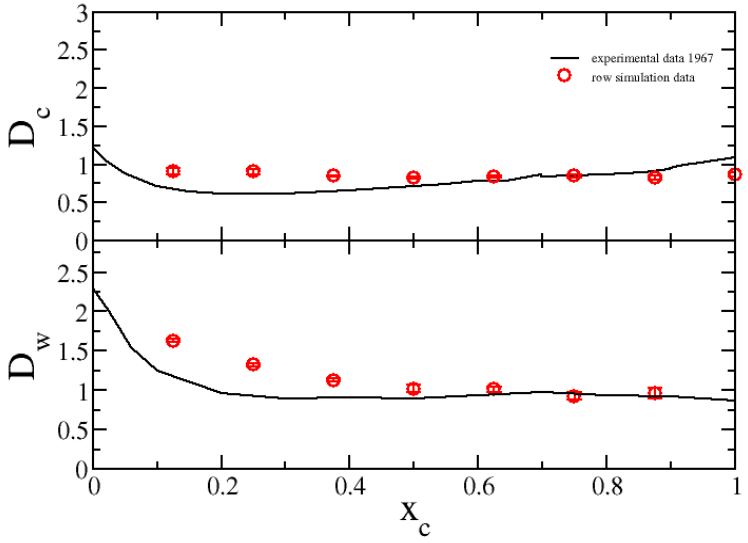




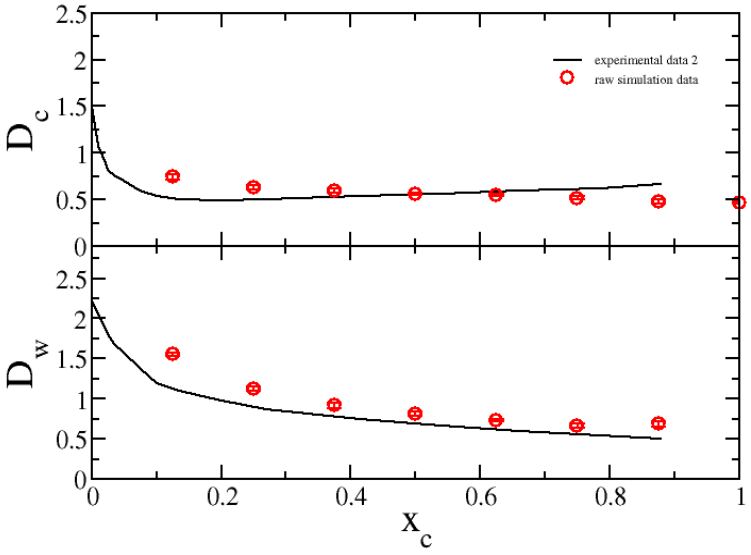
The (excess) molar enthalpy of mixing is displayed in Figure 2.4. The simulated enthalpy was less favorable at lower concentrations. However, the differences are within those typically observed for other force fields. As we pointed out previously,[5] it is possible to obtain a good enthalpy of mixing from a particular model and yet the contributions from each component may be incorrect. We believe there are two major possible sources for this discrepancy. Either composition-dependent polarization effects are significant, or the enthalpy contributions from composition-dependent vibrational and rotational frequency shifts have to be included for a more accurate comparison. Both of these effects are absent from the current force field. However, the results are still reasonable.

The self diffusion coefficient of both water and alcohol are displayed in Figure 2.5. The experimental trends were well reproduced, although the exact compositions corresponding to the low composition ones were overestimated. This is to be expected as the diffusion constant for pure SPC/E water is higher than experiment while the simulated value for alcohol was lower than experiment.

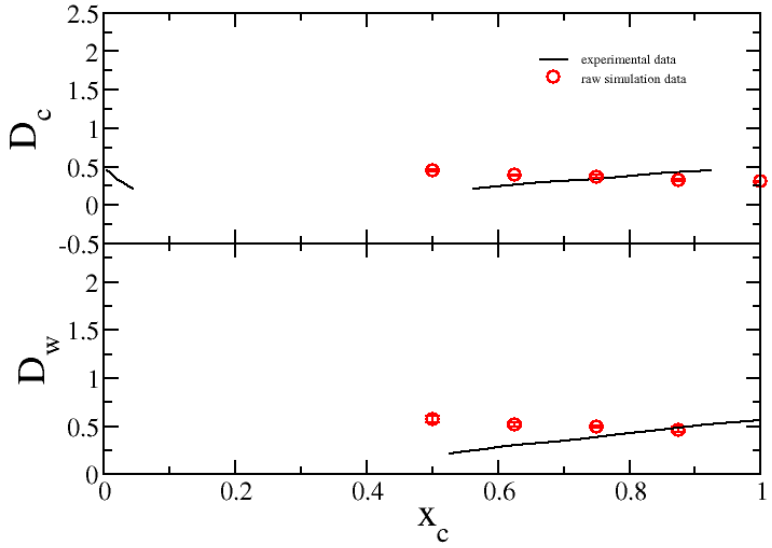
Figure 2.5 Alcohol (D_c) and water (D_w) translational self-diffusion constants ($\times 10^{-9}$ $m^2 s^{-1}$) as a function of alcohol mole fraction. a) ethanol b) n-propanol c) n-butanol d) i-propanol e) t-butanol. The solid lines correspond to the experimental data, and the circles to the simulation data. The experimental data have been scaled (<6%) using the pure solution values to correct for isotopic substitution effects.



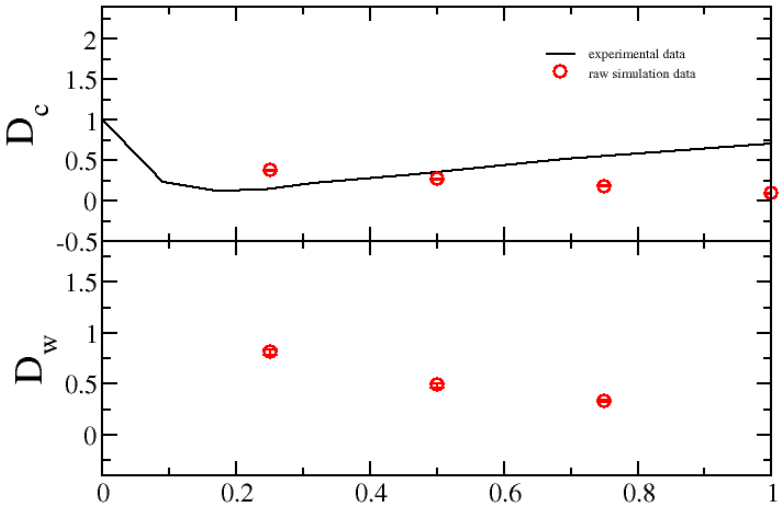
a)



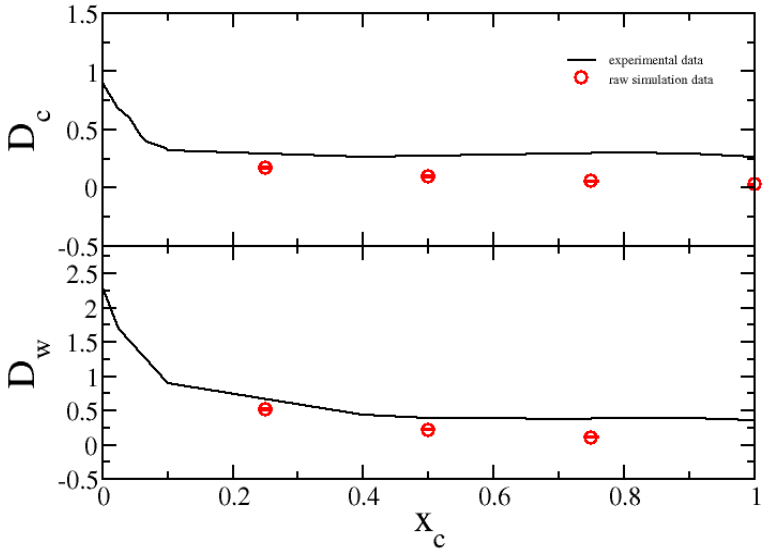
b)


experimental data
raw simulation data
D_c
D_w
x_c
2.5
2
1.5
1
0.5
0
-0.5
0.2
0.4
0.6
0.8

c)


experimental data
raw simulation data
D_c
D_w
2
1.5
1
0.5
0
-0.5
0.2
0.4
0.6
0.8

d)



e)

Due to the low solubility of n-butanol and n-octanol in water, these two alcohols barely dissolve in water; therefore will lead to a phase separation. Knowledge about the structure of phase separation between two immiscible liquids turns out to be a good test case for the current force fields. It will be interesting to see if the newly developed force fields examine such phenomenon or not. To that purpose, we have carried out two simulations by employing our newly developed parameters. The initial system was set up by placing all molecules randomly in an elongated box of 60 x 60 x 94 Å for n-butanol (x= 0.25) and 60 x 60 x 120 Å for n-Octanol (x =0.1010). In the case of n-butanol/water 4239 water molecules were added to the simulation box with 1422 n-butanol molecules, and n-Octanol/water box include 810 alcohols and 7209 water molecules. For each system, initial configurations were equilibrated using NPT ensemble (pressure coupling were applied for only x and y directions) at a constant temperature of 298K. Total simulation times were 45-55 ns, and the final 50 ns were used for analysis. An interface starts forming while in equilibration and eventually leads to a phase separation for both cases. Even though, there are still some water molecules staying within butanol and octanol phase. A

Snapshots of n-butanol/water and n-octanol/water system after equilibration are shown in Figure 2.6.

Figure 2.6 a) Snapshot of representative configuration from n-butanol. b) Snapshot of representative configuration from n-octanol

a) b)

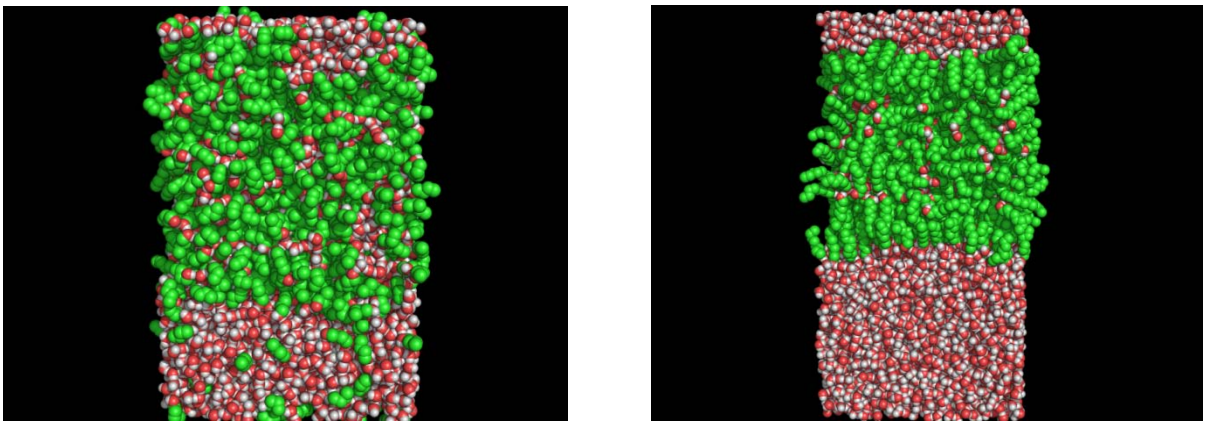
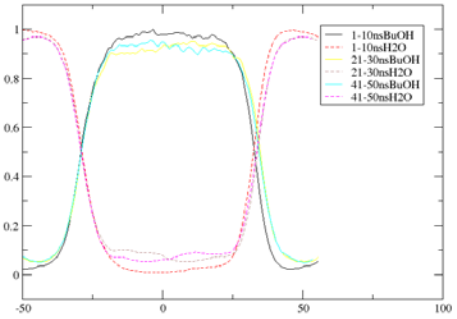
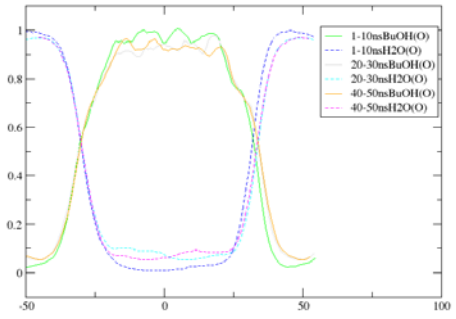


Figure 2.7 a) Center of mass density profile of n-butanol/water, as indicated: solid line : n-butanol: dotted line: water b) Oxygen density profile of n-butanol/water, as indicated: solid line : n-butanol: dotted line: water

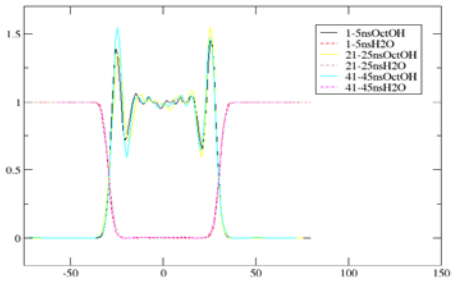


a)

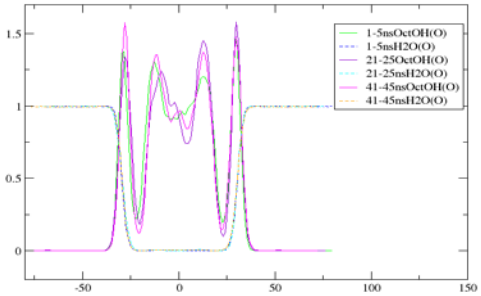


b)

Figure 2.8 a) Center of mass density profile of n-octanol/water, as indicated: solid line : n-octanol: dotted line: water b) Oxygen density profile of n-octanol/water, as indicated: solid line : n-octanol: dotted line: water



a)



b)

Along with snapshots, in figure 2.7 and 2.8 we plot density profile as a function of z direction based on center of mass and hydroxyl oxygen. From these profiles, n-octanol has formed more uniform and less "wet" alcohol phase than butanol does. This is due to the fact that n-octanol is less soluble than n-butanol, which is in agreement with experimental observations.

2.7 Conclusions

Simple nonpolarizable models for alcohols have been determined by attempting to reproduce the experimental KB integrals for a series of aqueous alcohol solutions. The experimental KB integrals are well reproduced. The bulk properties of a series of alcohol solutions have been studied by using radial distribution functions coupled with KB theory. The model quantitatively reproduces the thermodynamic properties of mixtures of water and alcohol for variety compositions. At higher mole fractions small deviations are observed which appear to relate to deviations in the enthalpy of mixing and partial molar enthalpies of alcohol and water. It is possible that these small deviations could be due to environmentally dependent polarization effects although this is difficult to determine. It is satisfying that our alcohol and water oxygen parameters are very similar as the electro negativities of carbon and hydrogen are very similar,

and that this results in the correct balance between solvation of alcohol by other alcohol molecules and solvation by water molecules as displayed by the KB integrals. One of the key criterion involved in a successful set of force field is transferability. Even a force field set can accurately reproduce the molecular behavior of a given system, transferability test on the more complicated systems and properties of interest is still very important. A force field with good transferability can dramatically reduce the computational cost and, therefore, improve the simplicity of parameter set. Here, we proposed a simple fix charge KB derived united atom model. The transferability has been tested for primary alcohols from methanol to n-octanol. Results show the same set of parameters reproduced a range of primary alcohols' properties well. This confirms the transferable feature of proposed model. However, our goal here is to find the parameter set which can provide the best representation of the thermodynamic properties of solution mixtures, such as enthalpy of mixing and excess Gibbs free energy of the solution.[38] As shown in Table 2.1, force field parameters have been modified for secondary and tertiary alcohols accordingly to accommodate for such goal.

`

Reference List

1. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

2. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003,** *119* (21), 11342-11349.

3. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics* **2004,** *121* (5), 2180-2186.

4. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

5. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

6. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.

7. Kirkwood, J. G.; Buff, F. P. The Statistical Mechanical Theory of Solutions . *Journal of Chemical Physics* **1951,** *19* (6), 774-777.

8. Ben-Naim, A. Inversion of Kirkwood-Buff Theory of Solutions - Application to Water-Ethanol System. *Journal of Chemical Physics* **1977,** *67* (11), 4884-4890.

9. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York. 1992.

10. Smith, P. E. Local chemical potential equalization model for cosolvent effects on biomolecular equilibria. *Journal of Physical Chemistry B* **2004,** *108* (41), 16271-16278.

11. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003,** *107* (16), 3891-3898.

12. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for the simulation of aqueous guanidinium chloride solutions. *Journal of Chemical Physics* **2004,** *121* (5), 2180-2186.

13. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

14. Kang, M.; Smith, P. E. Kirkwood-Buff theory of four and higher component mixtures. *Journal of Chemical Physics* **2008,** *128* (24), 244511.

15. Weerasinghe, S.; Smith, P. E. Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics* **2003,** *118* (23), 10663-10670.

16. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003,** *119* (21), 11342-11349.

17. 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. Chemical potential and concentration fluctuation in some aqueous alkane-mono-ols at 25 degrees. *Canadian Journal of Chemistry-Revue Canadienne de Chimie* **2003,** *81*, 141-149.

18. Hoiland, H. Partial Molal Volumes, Expansibilities, and Compressibilities for Aqueous Alcohol-Solutions Between 5-Degrees-C and 40-Degrees. *Journal of Solution Chemistry* **1980,** *9* (11), 857-866.

19. Nakanishi, K. Partial Molal Volumes of Butyl Alcohols and of Related Compounds in Aqueous Solution. *Bulletin of the Chemical Society of Japan* **1960,** *33*, 793-797.

20. Jain, A. K.; Tewari, R. K.; Mishra, R. K. Nonequilibrium Thermodynamics of Electroosmosis of Liquid-Mixtures Studies on Aqueous N-Propanol. *Journal of Non-Equilibrium Thermodynamics* **1980,** *5* (5), 275-284.

21. International Citical Tables. *Knovel Corporation* **2009,** vol III. 120-121.

22. Redlich, O.; Kister, A. T. Algebraic representation of thermodynamic properties and the classification of solutions. *Ind. Eng. Chem* **1948**, 40, 345-348.

23. Easteal, A. J.; Woolf, L. A. P,V,T and derived thermodynamic data for toluene, trichloromethane, dichloromethane, acetonitrile, aniline, and n-dodecane. *International Journal of Thermophysics* 1985, *6*(4) , 49-62.

24. Benson, G. C.; Kiyohara, O. Thermodynamics of aqueous mixtures of nonelectrolytes. I. Excess volumes of water-n-alcohol mixtures at several temperatures. *Journal of Solution Chemistry* 1980, *9*(10), 791-804.

25. Safarov, J.; Heydarov, S.; Shahverdiyev, A.; Hassel, E. Excess molar volumes VmE, isothermal compressibilities k, and thermal expansivities of {(1 - x)H2O + xCH3OH} at T = (298.15 to 523.15) K and pressures up to 60 MPa. *Journal of Chemical Thermodynamics* 2004, 36(7), 541-547.

26. Matteoli, E.; Lepori, L. Solute–solute interactions in water. II. An analysis through the Kirkwood–Buff integrals for 14 organic solutes. *Journal of Chemical Physics* **1984**, 80, 2856-2863.

27. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

28. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (3), 435-447.

29. Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26* (16), 1701-1718.

30. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001,** *7* (8), 306-317.

31. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular-Dynamics with Coupling to An External Bath. *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

32. Hess, B. P-LINCS: A parallel linear constraint solver for molecular simulation. *Journal of Chemical Theory and Computation* **2008,** *4* (1), 116-122.

33. Miyamoto, S. and Kollman, P. A. *Journal of Computational Chemistry* 1992, 13(8), 952-62.

34. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

35. Chitra, R.; Smith, P. E. Molecular dynamics simulations of the properties of cosolvent solutions. *Journal of Physical Chemistry B* **2000,** *104* (24), 5854-5864.

36. Daura, X.; Mark, A. E.; van Gunsteren, W. F. Parametrization of aliphatic CHn united atoms of GROMOS96 force field. *Journal of Computational Chemistry* **1998,** *19* (5), 535-547.

37. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269-6271.

38. Bentenitis, N.; Cox, N. R.; Smith, P. E. A Kirkwood-Buff Derived Force Field for Thiols, Sulfides, and Disulfides. *Journal of Physical Chemistry B* **2009,** *113* (36), 12306-12315.

39. Yamaguchi, T.; Hidaka, K.; Soper, A. K. The structure of liquid methanol revisited: a neutron diffraction experiment at -80 °C and +25 °C. *Molecular Physics* **1999,** 96, 1159-1168.

40. Yamaguchi, T.; Hidaka, K.; Soper, A. K. *Mol. Phys.* **1999**, 96, 603-605.

41. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487

42. Dixit, S.; Crain, J.; Poon, W. C. K.; Finney, J. L.; Soper, A. K. *Nature* **2002**, *416*, 829-832.

43. Laaksonen, A.; Kusalik, P. G.; Svishchev, I. M. *J. Phys. Chem. A* **1997**, *101*, 5910-5918.

44. Ferrario, M.; Haughney, M.; Mcdonald, I. R.; Klein, M. L. Molecular-dynamics simulation of aqueous mixtures: Methanol, acetone, and ammonia. *Journal of Chemical Physics* **1990**, *93*, 5156-5166.

CHAPTER 3 - Development and testing of protein backbone torsional potentials for the Kirkwood Buff derived force field of peptides and proteins

Abstract:

Recently, we have been developing a force field for biomolecular simulations of peptides and proteins (KBFF) designed to reproduce the experimental Kirkwood-Buff (KB) integrals observed in solution mixtures. This ensures a reasonable balance between solute-solute interactions and solute solvation – usually by water. Here, we describe the development and testing of the backbone torsion potentials, required for accurate modeling of the conformational preferences of amino acids, which are consistent with the corresponding KBFF nonbonded parameters. Molecular dynamics (MD) simulations were performed for dipeptides of glycine, alanine in solutions and several peptides. Comparisons with crystallographic data and quantum mechanics derived gas phase energy surfaces are also made.

3. 1 Introduction

The relationship between structure and protein function remains as a significant challenge in today's theoretical study of biological macromolecules systems. In the past few decades, empirical force fields represent a powerful tool to obtain protein structure function relationship at the atomic level. Studies applying those tools, however, rely on the quality of the version of force field being used. The implementation of classical force field-based simulations has made significant progress toward better representations of models of more complicated biological systems. The appropriate treatment of protein/peptide backbone turns out to be an important

factor to achieve the correct conformational distributions from molecular dynamics simulations. Several efforts have been made to improve the treatment of protein/peptide backbone.[1-6] Feig *et al.*[7-9] improved the backbone angles for CHARMM22[10] with grid correction terms. Perez and coworkers,[11] have corrected the α/β backbone ratio by using a refined version of AMBER99 parameter set. A variety of scaling factors for protein backbone torsional energetics was introduced by Sorin *et al*[1] in order to improve quality of different version of AMBER force fields. An improved two-dimensional adaptive umbrella sampling approach was introduced by Cao et al, [12] to determine the free energy surface of simple models. More recently, the Arnautova group[13] has used different charge models (single- and multiple-) to reproduce QM calculated electrostatic potentials. Despite several improvements, the relationship between protein backbone φ/ψ dihedral angles and the energy related to conformational changes of protein backbone is still not yet fully understood. Moreover, limitations in the treatment of different energy regions of protein backbone can cause misinterpretation of pathways of protein folding/unfolding. A typical approach in the development of protein φ/ψ parameters for the studies mentioned above was to fit the condensed phase dihedral parameters against high-level quantum mechanical methods (such as LMP2/cc-pVxZ/MP2/6-31G*), for small dipeptides in gas phase. The sources of the limitations of the existing protein backbone φ/ψ parameterization processes include using the solution phase particle charges to match up with gas phase charge distributions. More specifically, the question is that even if one can reproduce a gas phase φ /ψ map, to what degree are the φ /ψ preferences observed in the gas phase present under condensed phased protein conditions. Such limitations in the treatment of protein conformational backbone energy differences still remain. They will lead to systematic deviations in backbone φ/ψ angles in MD simulations of peptide models and in proteins.[14] Moreover, limitations in the treatment of high-energy regions of the protein backbone maps may cause misinterpretation of the pathway of

protein folding. To overcome these limitations, in our present work, the torsional parameters, together with those of side chains, were fitted against gas phase QM calculated energies. However, the charge set of the newly developed Kirkwood-Buff derived force field was scaled down to match up with gas phase values. Such a fitting method can overcome a possible source of overcompensation in existing force fields. Following gas phase parameter fitting, the Kirkwood-Buff derived force field with additional torsional terms has been assessed both by energetic comparison against QM (LMP2/cc-pVxZ/MP2/6-31G*) data and by classical MD simulation and replica exchange molecular dynamics simulations of short glycine, alanine, and proline dipeptides and several peptides.

3.2 Background and theory

3.2.1 Energy Functions of Kirkwood-Buff Derived Force Field

The backbone potential energy function used in Kirkwood-Buff derived force field is given by Figure 3.1

Figure 3.1 Protein Backbone Potential Function for KB Derived Force Field

$$U = \sum_{torsions} k_{\phi}[1 + \cos(n\phi - \phi_0)]$$

In this work, we focus on finding parameters for torsional term. For the description of torsional energy term, k_θ is the dihedral force constant (amplitude), n is dihedral periodicity, ϕ_0 is a phase of the dihedral angle ϕ (which could be either φ or ψ for backbone dihedral terms). The

fourier series for U_{dihedral} is approximated using small number of terms. In our newly developed KB derived force field, our choice of a four term expansion is consistent with the GROMOS[15,16] philosophy of using dihedral terms which can be physically rationalized and generate reasonable number of terms that need to be parameterized. In this case, there are only three parameters needed to be determined which are $k_{\theta,}$ ϕ_0, and *n*.

3.2.2 Optimization of gas-phase geometry

Kirkwood-Buff derived force field gas-phase geometry optimizations were run for every single φ/ψ combination that ab *initio* calculations had been carried out. The *ab initio* energy surface at LMP2/cc-pVxZ/MP2/6-31G* levels used in this study has a grid resolution of 15° which indicates a total of 624 runs are required. The torsional angles defining the energy surface to be fitted were fixed whereas no constraints were imposed for bonds. During calculations, the main dihedrals (CH_3 –C-N-Cα and Cα-C-N-CH3) were all set to a fixed value of 180°. The energy minimizations were done with GROMACS program (v3.3.1).[17,18] A conjugate gradient algorithm for energy minimization was applied and the convergence factor for the calculations was 10^{-6} kJ/mol/nm. The objective of such process is to explore local and global energy minima thoroughly.

3.2.3 Scaling of 1-4 interactions

An accurate description of interactions in a molecule is a crucial element in today's force field development work. 1-4 interactions affect the nobonded and torsional contributions to the total energy and therefore, should be treated correctly. Sorin and colleagues[19] noted that scale factor for 1-4 interaction can change secondary structure propensity of force fields. Many existing parameter sets have used scale factors to scale down Lennard-Jones(LJ) nonbonded and

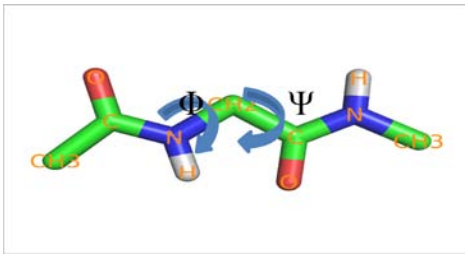
electrostatic interactions between 1-4 atoms. For example, 0.5 is a commonly used scale factor for LJ potentials. Coulomb interactions are also scale down by 2 in OPLS,[20] by 1.2 in AMBER.[21] However, the scale factor must be handled carefully. As demonstrated by Smith and Karplus,[22] the inappropriate usage of scale factor may lead to artifact effects. Our motivation here is to obtain appropriate scale factor values to our potential model which has a positive impact on the accuracy of our force field.

3.2.4 Torsion parameter development for peptides

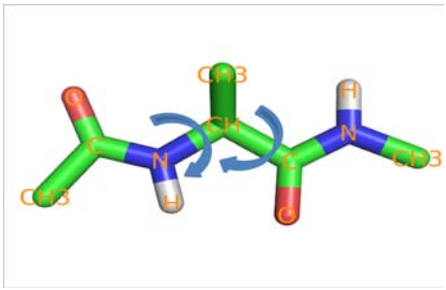
Traditionally, dipeptides are used as models for understanding protein backbone dynamics and to derive force field parameters. Here, blocked glycine and alanine dipeptides (Figure 3.2) are used as prototypes to determine torsional potential parameters for the phi and psi angles.

Figure 3.2 Diagrams of (a) blocked glycine dipeptide (b) blocked alanine dipeptide

a)



b)



QM potential energy maps of dipeptides at the LMP2/cc-pVxZ/MP2/6-31G* levels were taken from literature.[23] Initial attempts to obtain the parameters for four main-chain torsional potentials(C-N-Cα-C, N-Cα-C-N, C-N-Cα-C_β, and N-C-Cα-C_β) was attempted by fitting to a QM calculated energy profile. There are 624 points on this QM energy surface map with a grid resolution of 15° in each direction. We fit the dihedral parameters with the third Fourier terms (1 fold, 2 fold, and 3 fold or total of 3 parameters for each dihedral angle). It is worth mentioning that the goal of fitting is to assure the energetic surface of gly and ala dipeptides are adequately represented. The idea of backbone parameters development is to find a single set of backbone dihedral φ(C-N-Cα-C) /ψ(N-Cα-C-N) for all amino acids. Comparisons between the QM and MD data were made to assure the accuracy of newly developed model. The unique nature of glycine (no β carbon) makes it a good candidate to determine φ/ψ parameters. For those nonglycine amino acids an additional set of φ /ψ parameters will be added to the glycine determined values (here defined as φ_b = C-N-Cα-C, ψ_b = N-C-Cα-Cβ). Therefore, for amino acids other than glycine the total torsional energy will be the summation of energies calculated for φ/ψ and φ_b/ψ_b. Here, glycine dipeptide φ/ψ parameters were developed first, which are based on the best fit of QM gas phase map principle. Once the φ/ψ parameters are determined, they will be used in another round of fitting where φ_b/ψ_b parameters were found. Such procedure is carried out by fitting the parameters to the best reproduce QM energies for alanine dipeptide. The charge set of KB force fields were scaled down by thirty percent against condensed-phase values. Such fitting procedure can be a source of possible complement for existing discrepancy gas-phase fitting techniques.

Table 3.1 Dihedral Parameters: Dihedral Force constants (k_θ), Phase shift of dihedral angle(ϕ_0), and Dihedral Periodicity(n) for Backbone Dihedral Angles

Dihedrals	k_θ	ϕ_0	n
φ	0.772	0.0	1
	2.700	0.0	2
	2.439	0.0	3
ψ	5.359	0.0	1
	-4.481	0.0	2
	-0.551	0.0	3
φ’	-2.515	0.0	1
	-1.104	0.0	2
	-0.796	0.0	3
ψ’	2.518	0.0	1
	0.500	0.0	2
	0.466	0.0	3

Our basic approach to protein backbone force field development is to find a set of parameters which best reproduce QM energies for dipeptides. To achieve this goal we have focused on fine tuning force constants (k_θ), phase shift (ϕ_0), and periodicity(n) term. Other researchers have also explored similar ideas.[12,20,24-26] The major improvement of present gas-phase fitting approach is the usage of gas-phase charge distributions instead of solution values. Briefly, a scale factor of 0.7 for solution charges was used for all fittings. Such an effect can partially mimic the gas-phase electrostatics. Due to the complexity of parameterization, a large number of test cases were carried out. Over 100 sets of k_θ, ϕ_0, and n values were tested. Tables 3.1 shows the final parameters used in our study.

3.3 Simulation details

All molecular dynamics simulations/energy minimizations were performed for dipeptides and test cases, together with the SPC/E water model,[27] as implemented in the GROMACS program (v3.3.1).[17, 28] The water geometry was constrained using SETTLE.[29] A twin range cut-off of 1.5 nm/1.5 nm was employed with a nonbonded pair list update of every 10 steps. Long range electrostatic interactions were evaluated using the PME approach.[30] Dipeptides were blocked with N-methylamide at C terminus and with an acetyl group at N terminus. The startup structure for dipeptide simulations were fully extended backbone dihedrals near (φ = -180, ψ = 180). The dipeptides were solvated by SPC/E water in cubic box with size 35Å x 35Å x 35Å. Before the production run, a 5 ns equilibration run was carried out, followed by a 50 ns NPT MD run to generate trajectories for the purpose of analysis. During the simulation, bond length was constrained by using SHAKE algorithm.[31] The temperature coupling is monitored by velocity rescale method.[32] Scale factors of 0.1 and 0.75, were applied to Lennard-Jones nonbonded and electrostatics interactions, respectively. Non-bonded interactions between atoms that are no further than 3 bonds away are excluded.

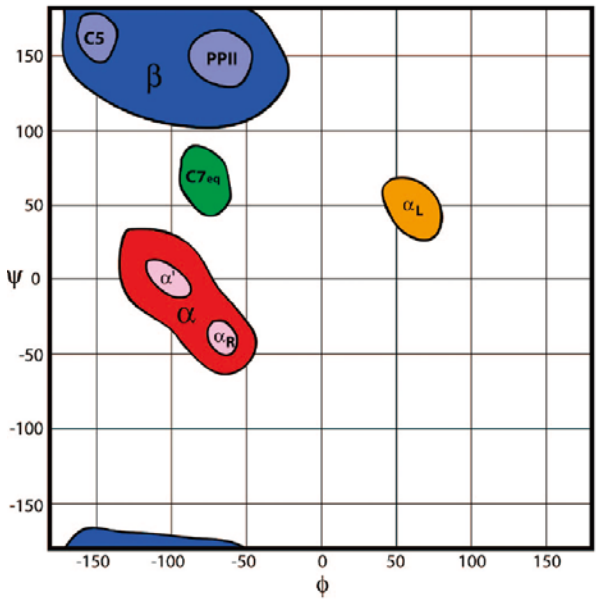
Replica exchange molecular dynamics (REMD),[33,34] involves simulating multiple replicas of the same system at different temperatures and randomly exchanging the complete state of two replicas at regular intervals with the probability. For the case alanine dipeptide solution simulation, we used fourteen replicas from 300K to 382K with exchange at 1ps intervals. Exchange probability for neighboring replicas is 0.3.

3.4 Result and discussion

The flexibility of polypeptide backbone conformations reflects different secondary structure. Alanine and glycine dipeptides can serve as prototype models for protein backbone

parameterization in protein fields. For nonglycine and nonproline residues the well-known Ramachandran plot[35] of ψ versus ϕ identifies two major minima: α_R (φ = -60, ψ = -50) in the R basin and PPII/C5 (φ = -60 to -170, ψ = 120 - 170) in the β basin (see Figure 3.3), which corresponds to R-helical and extended β -strand/-sheet secondary structures when repeated over multiple amino acid residues. Secondary minima with higher relative free energies at α_L (φ = 50, ψ = 50) and $C7_{ax}$ (φ = 50, ψ = -130) are relevant in the formation of turns and loops.[23]

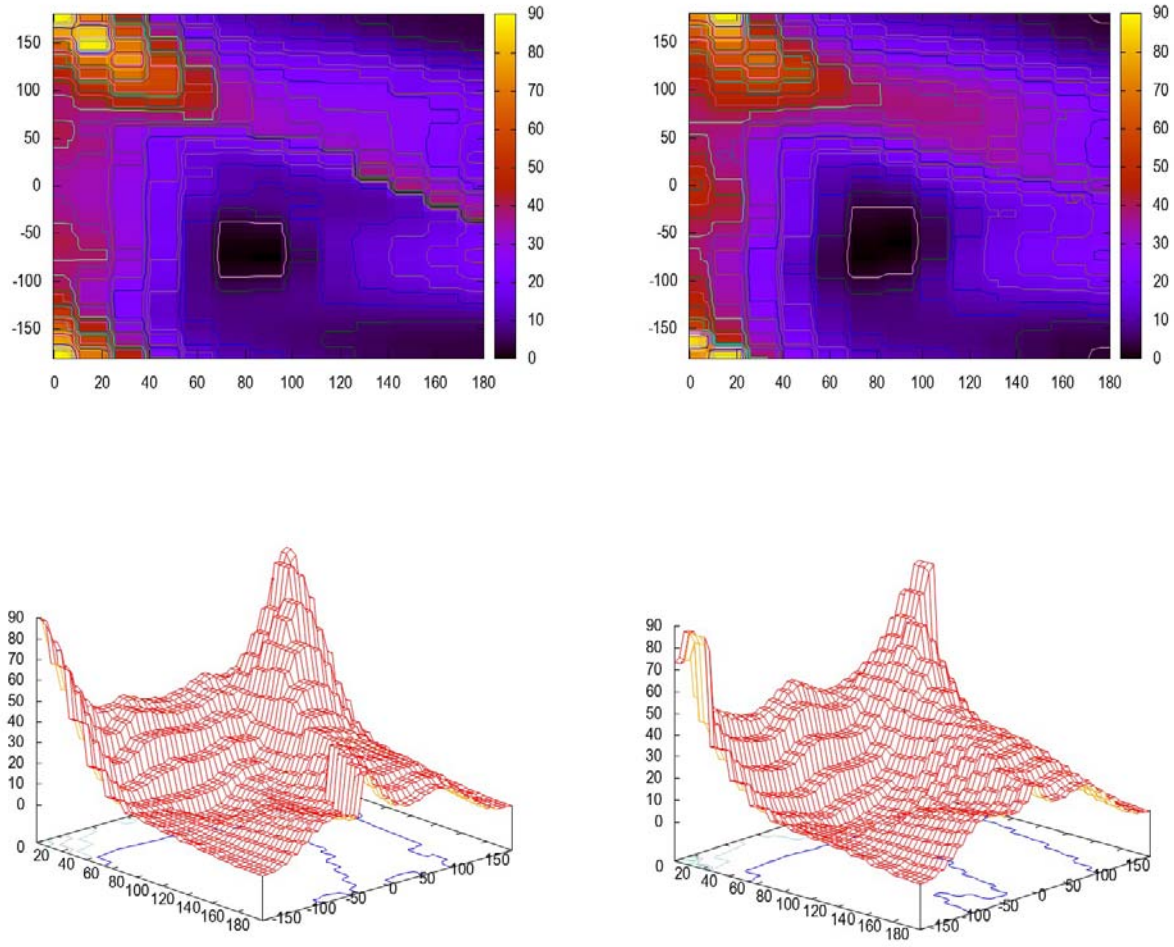
Figure 3.3 Schematics overview of major conformational basins sampled by ϕ/ψ backbone angles in nonglycine, nonproline peptide residues.



3.4.1 Dipeptide gas-Phase Simulations.

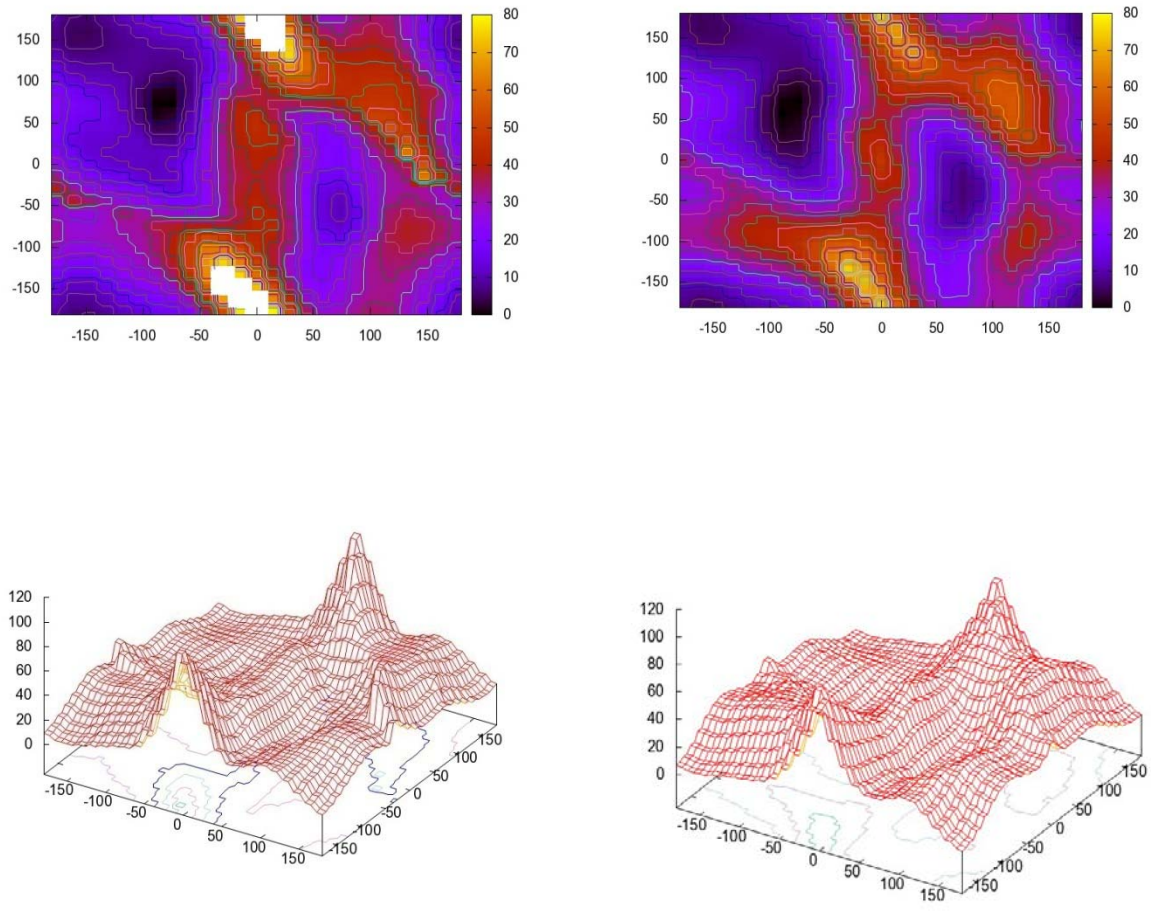
The full Glycine dipeptide map obtained using MD simulation s is presented in Figure 3.4, along with QM calculated gas phase map. As expected, all of the surfaces are similar.

Figure 3.4 QM calculated glycine dipeptide phi, psi energy (kJ/mol) map vs Fitted glycine dipeptide phi, psi energy map



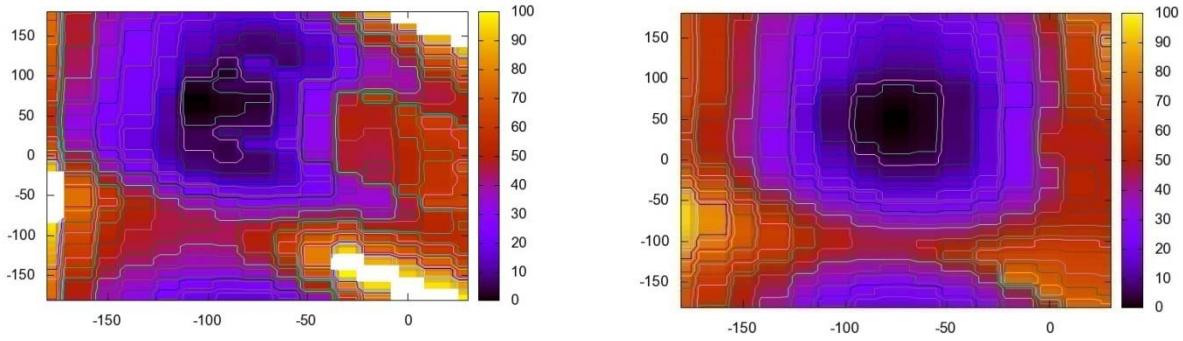
Comparison of the gas phase energy surface of alanine dipeptide is shown in Figure 3.5. The highest MD simulated energy is relative lower than QM values; however, it is subtle and irrelevant. Since protein backbone conformation only related with lower energy region on the map.

Figure 3.5 QM calculated Ala dipeptide gas phase phi, psi energy (kJ/mol) map vs Ala dipeptide phi, psi map with fitted dihedral energy terms



Potential energy surfaces for proline dipeptide are compared in Figure 3.6. Due to the distinctive cyclic structure of proline's side chain, only the range of locks its φ from -180° to 0° were calculated. Only one single minimum is observed on the energy surface map, corresponding to the region (φ = -75°, ψ = 50°). The overall features of QM maps are reproduced well by molecular dynamics gas-phase simulations. It appears that the extended minimum regions towards 0 φ value are slightly unfavorable. Moreover, the global minimum around (φ = -75°, ψ = 50°) is a bit too favorable.

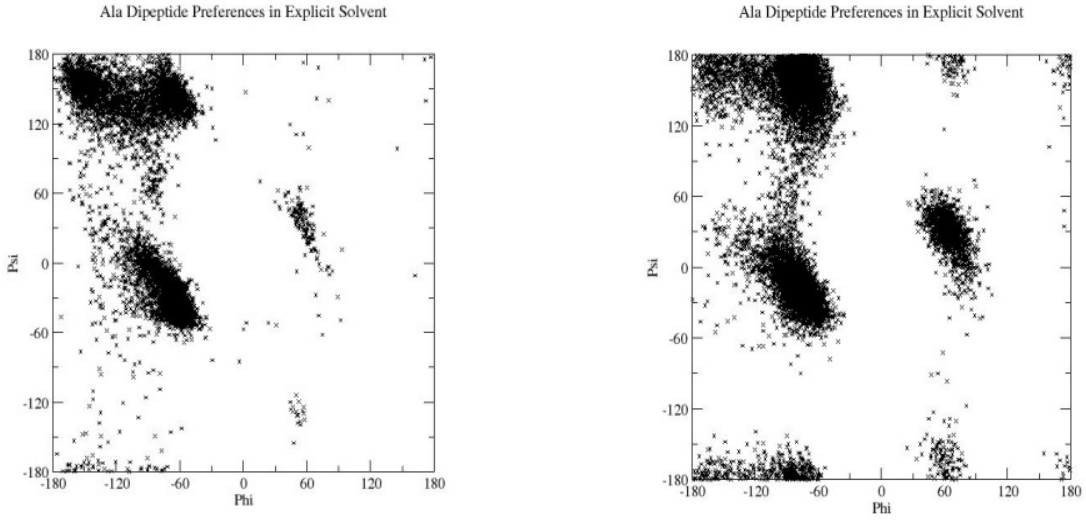
Figure 3.6 QM calculated Proline dipeptide gas phase phi, psi energy (kJ/mol) map vs Proline dipeptide phi, psi map with fitted dihedral energy terms



3.4.2 ϕ/ψ Sampling in Dipeptide versus PDB Structures.

The torsional (ϕ/ψ) backbone distribution was analyzed from MD simulations in explicit water and compared to the distributions from PDB structures. Results for alanine and glycine in ϕ/ψ sampling are shown in Figure 3.6-3.7. The PDB ϕ/ψ distribution has been reproduced very well, especially in those regions corresponding to lower-energy states. Detailed comparison shows the subtle variations are also reproduced well. In alanine, α region is more extended the C7 transition region in between α and β is slightly lower than PDB, the α_R helical basin is relative favorable, and α_L conformation is more pronounced.

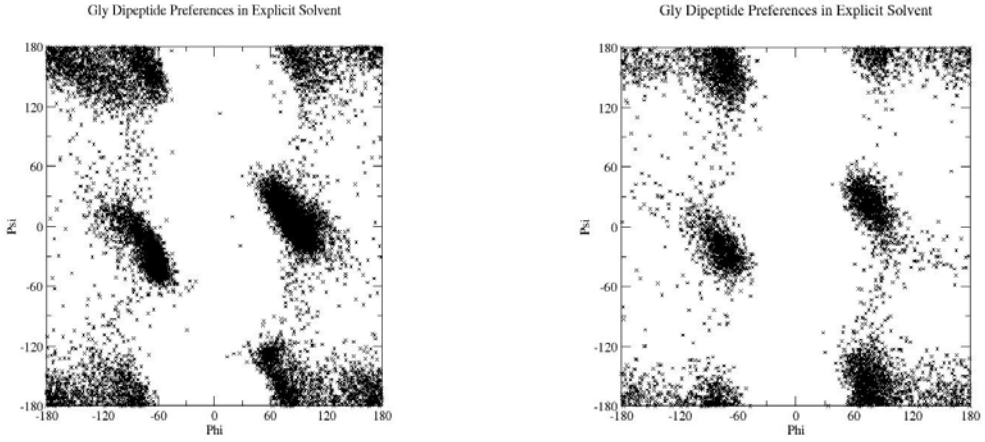
Figure 3.7 a) Alanine dipeptide phi, psi distributions from PDB crystal structures b) alanine dipeptide phi, psi distributions from REMD run by using KB derived force field



a) b)

In glycine, the agreement between from PDB and from the simulations is also good. Every basin regions have been reproduced in our solution simulations, except the preference for (ϕ = -85, ψ = 0) and for (ϕ = 85, ψ = 0). Those two minima regions is less extended. The preference for the minima region around (ϕ = -150, ψ = 180) and (ϕ = -150, ψ = -180) were also less pronounced.

Figure 3.8 Glycine dipeptide phi, psi distributions from a) PDB crystal structures vs b) Glycine dipeptide phi, psi distributions from MD simulations



a) b)

3.4.3 Protein simulations

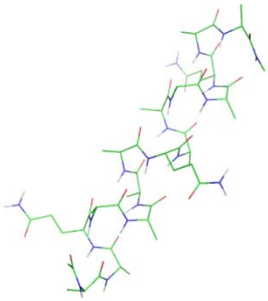
The ultimate goal of current parameterization work is to obtain proper balance of protein secondary structure preferences. Hereby, we have selected a range of peptides and mini-proteins which can be divided into three groups: α helices, β sheets, and loops, respectively. The conformational changes during the protein simulations were compared to experimental structures to trace the degree of realism in the simulations. Averages of final root-mean-square deviation (rmsd) over the entire trajectory during the simulation are reported in Table 3.1. Simulation conditions are set to be the same as under which experimental measurements were took place. Six out of nine rmsd averages from the crystal structure is less than 1 Å and other two are between 1 and 2 Å. For the case of mini protein (2JOF), the deviation is bigger, due to the large fluctuations of residues 8-16 which consist mostly of loop region.

Table 3.2 Root Mean Square Deviations from Experimental Structures in Protein Simulations

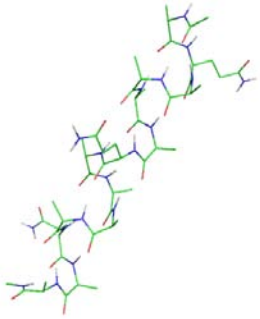
Systems (PDB ID)	Type	Length	Avg. C_α rmsd [Å]	Temperature	MD runs
Trp-cage(1L2Y)	α/coil	20	N/A	282K	60ns
Trpzip2(1LE1)	β(90%)	12	1.10(0.20)	288k	100ns
Mini Protein (2JOF)	α/coil	20	1.65(0.32)	280K	60ns
GB1P(3GB1)	β(30-80%)	16	0.30(0.10)	298K	100ns
AAQAA	α(50%)	15	0.70(0.15)	277K	100ns
AAQAA(incorrect)	α(50%)	15	0.50(0.10)	277K	100ns
AAQAA(PI)	α(50%)	15	0.45(0.10)	277K	10ns
AAQAA(310)	α(50%)	15	0.55(0.10)	277K	10ns
Glu-Lys(14A1)	α(80%)	17	0.90(0.10)	273K	20ns

Experimental measurements predict that $(AAQAA)_3$ has approximately a 50% helical[36] structure. In figure 3.8, we present the result of our simulations using our KBFF with newly developed torsional terms after 100 ns run compared with $(AAQAA)_3$ in 100% helical structure. Several folding and unfolding structures have been observed through the whole simulation time. During the course of simulation our simulated structures have consistently lower helicity values, which is in a reasonable agreement with experimental observations (~50%).

Figure 3.9 A representative simulated structure of $(AAQAA)_3$ compared with ideal helical $(AAQAA)_3$



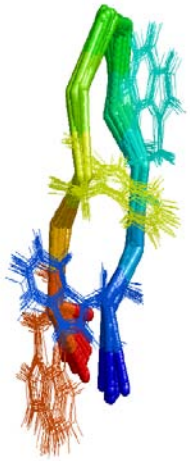
Ideal helix



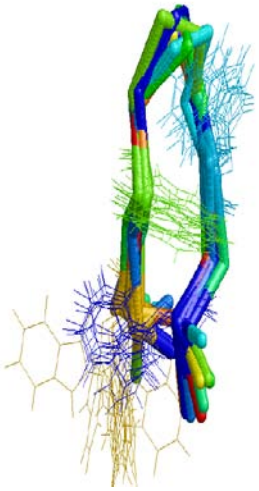
Simulated structure after 100 ns MD run

Beta hairpins are the simplest structure motif involving two beta strands which look like a hairpin. The strong propensity toward sheet structure and unique turn sequence make it a good test case for sampling β form. Here we chose two different hairpins Trpzip2 and GB1P, which have β ratio of 90%[37] and 30-80%,[38] respectively. Figure 3.9 and 3.10 represent the representative folded hairpin structures of Trpzip2 and GB1p. The folded structures were built from NMR structure data of the G B1 domain with a PDB bank ID 3gb1. The computed Cα main chain RMSD value for Trpzip2 and GB1P are 1.10 and 0.30 compared with NMR structures. This demonstrates that both Trpzip2 and GB1P are stable with native secondary structures. Therefore we showed the ability of our new KB derived force field in producing correct folded native conformations.

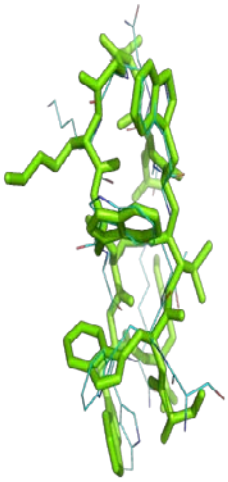
Figure 3.10 A representative simulated structures of Trpzip2(PDB ID: 1LE1)



NMR structure of Trpzip2

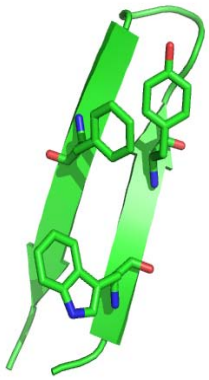


Simulated Trpzip2 structures for every 10^{th} ns over 60 ns MD run

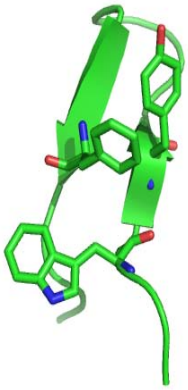


The 1st simulated Tripzip2 structure vs the last simulated Trpzip2 structure

Figure 3.11 A representative simulated structures of GB1P(PDB ID: 3GB1)



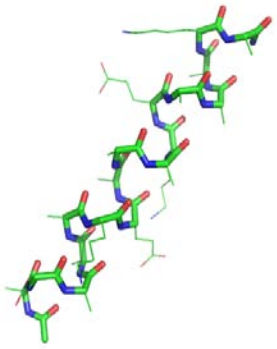
NMR structure of GB1P



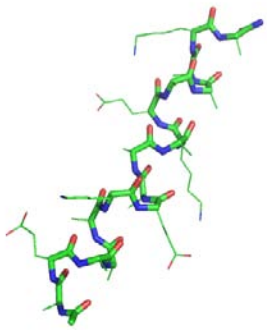
Simulated GB1P structure after 60 ns MD run

To test the ability of predicting preference of helical secondary structure, we selected the peptide designed by Marqusee and Baldwin. This peptide (PDB ID: 14A1) has three glutamic/lysine residue pairs and is 17 amino acids long. Marqusee demonstrated 14A1 is over 80% helical at 273K, due the stabilization by $Glu^{-} \cdots Lys^{+}$ salt bridges. As shown in Figure 3.11 the simulated structure after 20 ns MD run was compared with ideal helical of the same peptide. Over the course of simulation, the peptide keeps low structural deviation from the ideal helix with small RMSD value of 0.90. The strong interactions between Glu^{-} and Lys^{+} were mimicked in the right manner.

Figure 3.12 A representative simulated structures of Glu-Lys peptide



1st simulation structure (ideal helix)



The last simulated Glu-Lys peptide structure

All our test case protein simulation runs were under 100 ns. As reported elsewhere,[39] even simulations up to a microsecond do not completely sample the conformation space accessible during formation of alpha helices and beta sheet. Such limitation of length scale for each simulation may not provide adequate information of sampling of α versus β conformations. However, it is more meaningful to sample relative conformations related to a given basin. Such an approach is expected to be more meaningful since for a given protein system as the conformational changes can be studied thoroughly without any other sources of disruptions.

3.5 Conclusions

Motivated by limitations in common classical molecular force field in reproducing structure properties in both QM and experimental measurements we have continuously worked on developing an accurate nonpolarizable force field with better protein backbone representation. Limitations of other methods largely come from the difference between gas phase particle

charges and condensed phase charges and high propensities towards α helices. As pointed out elsewhere[14] when traditional fitting techniques are performed, accurate treatment of condensed phase ϕ, ψ distributions in both α and β regions is difficult to achieve, therefore require additional adjustments. It is obvious that current potential functions in use can hardly treat the gas and condensed phases with high accuracy simultaneously. Here, we have shown that the use of QM-based energy surface for developing backbone parameters can provide improved agreement with experimental structure distributions.

To overcome such a known limitation protein backbone charges were tuned/scaled against gas phase data prior to parameterization. Torsional preferences of proteins and dipeptides simulations were then compared with experimental measured structures. In the present work it was shown that the use of a high level QM-based energy surface together with scaled condensed phased charges improved the quality of previously developed Kirkwood-Buff force field. The results shown here also in good agreement with φ/ψ angle distributions found in PDB structures.

Reference List

1. Sorin, E. J.; Pande, V. S. Empirical force-field assessment: The interplay between backbone torsions and noncovalent term scaling. *Journal of Computational Chemistry* **2005,** *26* (7), 682-690.

2. Cao, Z. X.; Lin, Z. X.; Wang, J.; Liu, H. Y. Refining the Description of Peptide Backbone Conformations Improves Protein Simulations Using the GROMOS 53A6 Force Field. *Journal of Computational Chemistry* **2009,** *30* (4), 645-660.

3. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins-Structure Function and Bioinformatics* **2006,** *65* (3), 712-725.

4. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

5. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. An improved OPLS-AA force field for carbohydrates. *Journal of Computational Chemistry* **2002,** *23* (15), 1416-1429.

6. Cao, Z. X.; Lin, Z. X.; Wang, J.; Liu, H. Y. Refining the Description of Peptide Backbone Conformations Improves Protein Simulations Using the GROMOS 53A6 Force Field. *Journal of Computational Chemistry* **2009,** *30* (4), 645-660.

7. Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* **2004,** *25* (11), 1400-1415.

8. Mackerell, A. D.; Feig, M.; Brooks, C. L. Improved treatment of the protein backbone in empirical force fields. *Journal of the American Chemical Society* **2004,** *126* (3), 698-699.

9. Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones? *Journal of Chemical Theory and Computation* **2008,** *4* (9), 1555-1564.

10. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** *102* (18), 3586-3616.

11. Rueda, M.; Ferrer-Costa, C.; Meyer, T.; Perez, A.; Camps, J.; Hospital, A.; Gelpi, J. L.; Orozco, M. A consensus view of protein dynamics. *Proceedings of the National Academy of Sciences of the United States of America* **2007,** *104* (3), 796-801.

12. Cao Zanxia; Lin Zhixiong; Wang Jun; Liu Haiyan Refining the Description of Peptide Backbone Conformations Improves Protein Simulations Using the GROMOS 53A6 Force Field. *Journal of computational chemistry* **2008,** *30*(4) 645-660.

13. Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. A new force field (ECEPP-05) for peptides, proteins, and organic molecules. *Journal of Physical Chemistry B* **2006,** *110* (10), 5025-5044.

14. Mackerell, A. D.; Feig, M.; Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* **2004,** *25* (11), 1400-1415.

15. W.F.van Gunsteren and H.J.C.Berendsen Groningen Molecular Simulation (GROMOS) Library Manual. BIOMOS b.v,Groningen: 1987.

16. Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; van Gunsteren, W. F. The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry* **2005,** *26* (16), 1719-1751.

17. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001,** *7* (8), 306-317.

18. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26* (16), 1701-1718.

19. Sorin, E. J.; Pande, V. S. Empirical force-field assessment: The interplay between backbone torsions and noncovalent term scaling. *Journal of Computational Chemistry* **2005,** *26* (7), 682-690.

20. Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B* **2001,** *105* (28), 6474-6487.

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. A 2Nd Generation Force-Field for the Simulation of Proteins, Nucleic-Acids, and Organic-Molecules. *Journal of the American Chemical Society* **1995,** *117* (19), 5179-5197.

22. Smith, J. C.; Karplus, M. Empirical Force-Field Study of Geometries and Conformational Transitions of Some Organic-Molecules. *Journal of the American Chemical Society* **1992,** *114* (3), 801-812.

23. Feig, M. Is alanine dipeptide a good model for representing the torsional preferences of protein backbones. *Journal of Chemical Theory and Computation* **2008,** *4* (9), 1555-1564.

24. Mackerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B* **1998,** *102* (18), 3586-3616.

25. Kony, D.; Damm, W.; Stoll, S.; van Gunsteren, W. F. An improved OPLS-AA force field for carbohydrates. *Journal of Computational Chemistry* **2002,** *23* (15), 1416-1429.

26. Cao, Z. X.; Lin, Z. X.; Wang, J.; Liu, H. Y. Refining the Description of Peptide Backbone Conformations Improves Protein Simulations Using the GROMOS 53A6 Force Field. *Journal of Computational Chemistry* **2009,** *30* (4), 645-660.

27. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The missing term in effective pair potentials. *Journal of Physical Chemistry* **1987,** *91*, 6269-6271.

28. van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* **2005,** *26* (16), 1701-1718.

29. Miyamoto, S.; Kollman, P. A. Settle - An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992,** *13* (8), 952-962.

30. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.

31. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. Numerical-Integration of Cartesian Equations of Motion of A System with Constraints - Molecular-Dynamics of N-Alkanes. *Journal of Computational Physics* **1977,** *23* (3), 327-341.

32. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *Journal of Chemical Physics* **2007,** *126* (1), 014101-014107.

33. Sugita, Y. O. Y. Replica-exchange molecular dynamics method for proteint folding. *Chemical Physics Letters* **1999**, *314*(1,2), 141-151.

34. Okabe, T. K. M. O. Y. Replica-exchange Monte Carlo method for the isobaric-isothermal ensemble. *Chemical Physics Letters* **2001**, *335*(5,6), 435-439.

35. Ho, B. K.; Brasseur, R. Ramachandran plots of glycine and pre-proline. *BMC Structural Biology* **2005**, 5(14), 1472.

36. Shalongo, W.; Dugad, L.; Stellwagen, E. Distribution of Helicity Within the Model Peptide Acetyl(Aaqaa)(3)Amide. *Journal of the American Chemical Society* **1994,** *116* (18), 8288-8293.

37. Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. Tryptophan zippers: Stable, monomeric beta-hairpins. *Proceedings of the National Academy of Sciences of the United States of America* **2001,** *98* (10), 5578-5583.

38. Fesinmeyer, R. M.; Hudson, F. M.; Andersen, N. H. Enhanced hairpin stability through loop design: The case of the protein G B1 domain hairpin. *Journal of the American Chemical Society* **2004,** *126* (23), 7238-7243.

39. Clarke, D. T.; Doig, A. J.; Stapley, B. J.; Jones, G. R. The alpha-helix folds on the millisecond time scale. *Proceedings of the National Academy of Sciences of the United States of America* **1999,** *96* (13), 7232-7237.

CHAPTER 4 - Surface tension of common water model: a simulation study water

The Journal of Chemical Physics **2007**, 126, 221101

Abstract

Initial simulated values of the surface tension for the SPC/E water model have indicated excellent agreement with experiment. More recently, differing values have been obtained which are significantly lower than previous estimates. Here, we attempt to explain the differences between the previous studies and show that a variety of simulation conditions can affect the final surface tension values. Consistent values for the surface tensions of six common fixed charge water models (TIP3P, SPC, SPC/E, TIP4P, TIP5P, and TIP6P) are then determined for four temperatures between 275 and 350 K. The SPC/E and TIP6P models provide the best agreement with experiment.

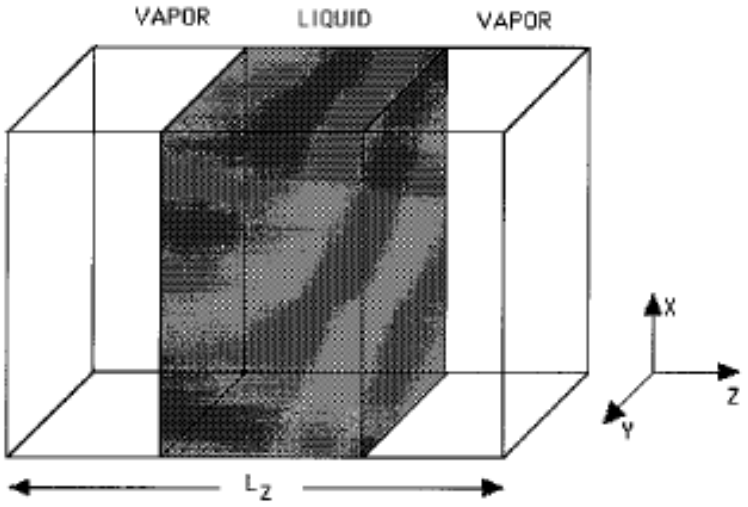
4.1 Introduction

Surface tension is an important property of water which has significant implications for the behavior of water at a variety of interfaces. The ability of computer simulations to reproduce this interfacial water behavior depends on the quality of the water model. Hence, several studies have been performed to determine the surface tension of different water models. Initially, the simulated surface tension values varied quite widely,[1-4] presumably due to the different accuracies of the various water models and the slow convergence properties of the computed surface tension.[5] More recently, consistent values of the surface tension of the SPC/E water model[6] have appeared which indicate excellent agreement with experiment.[7-9] This is somewhat surprising as most water models are developed to reproduce bulk water properties and one would

therefore expect some errors for interfacial systems. Here, we present evidence that several of the previous simulated values of the surface tension of SPC/E water have overestimated, for a variety of reasons, the true value by a significant degree (15%). Three studies of the surface tension of SPC/E water have been presented that appear to be in good agreement with both the experimental data and each other. Alejandre *et al.*, Shi *et al.*, and Lu and Wei determined values of 66 mN/m (328 K),[10] (72 mN/m) (302 K),[11] and 70 mN/m (300 K),[9] which compare well to the experimental values of 67.1, 71.3, and 71.6 mN/m,[12] respectively. Several of the studies also emphasized the need to include long range dispersion interactions in determining the pressure tensor and to include additional *k* vectors in the reciprocal space calculation for rectangular systems.[10,11] More recently, lower values have appearing in the literature, and thereby determine a consistent value for the surface tension of the SPC/E water model. Having obtained consistent values for the surface tension of SPC/E water at 300 K, we then determined the surface tensions of six common fixed charge water models (TIP3P,[13] SPC,[14] SPC/E,[15] TIP4P,[16] TIP5P,[17] and TIP6P[18]) at four different temperatures of 275, 300, 325, and 350 K. The surface tensions of the SPC, TIP5P, and TIP6P models have not been determined previously as a function of temperature, while the results for the SPC/E model are found to be different from current literature values been observed by Wemhoff and Carey,[19] although a reason for the disagreement with earlier values was not provided. In addition, Ismail *et al.* have quoted a significantly lower value of 55.4 mN/m for SPC/E water at 300 K.[20] Our own studies using the SPC/E water model have also consistently underestimated the surface tension in comparison to both the experimental data and previous studies. Hence, our aim here is to explain the reasons for some of the different values appearing in the literature, and thereby determine a consistent value for the surface tension of the SPC/E water model.

4.2 Methods

Figure 4.1 rectangular parallelepiped cell with a liquid slab in the middle and vapor in each side of cell. L_z = 120 A for N = 512. The z axis is perpendicular to the interface.



The initial configuration consists of a rectangular parallelepiped with the molecules located in its center, as shown in Fig. 1. The simulation box has a volume $V=Lx\ Ly\ Lz$. The sides are of length, $Lx= Ly$ =19.7 Å, Lz=100 Å The molecules are arranged in two replicated cubic boxes along the *z* direction. All simulations were performed with the GROMACS program v3.2.1 in single precision.[21,22] The system involved a slab of 512 water molecules in a constant volume box of 1.97 x 1.97 x 10.0 nm^3 coupled to a temperature bath using a Berendsen thermostat.[23] Each system was equilibrated for 1 ns and then simulated for an additional 3–5 ns during which the initial surface tension values were determined from the diagonal elements of the pressure tensor according to the relationship, $\gamma_0= 1/2Lz[Pzz- ½(Pxx+Pyy)]$, where *Lz* is the box length in the *z* direction and $P_{\alpha\alpha}$ is the αα component of the pressure tensor.[24] Electrostatic energies were determined using the particle mesh Ewald (PME) approach[25] with a convergence parameter of 3.1 nm^{-1}, a real space LJ and Coulomb cutoff of 0.98 nm, a grid resolution of 0.12 nm, and tinfoil boundary conditions.[26] The time step was 2 fs and SETTLE was used to constrain

the water geometry.[27] The long range dispersion correction (γ_d) was included in the calculated values,[28] giving a final surface tension of $\gamma = \gamma_o + \gamma_d$. The dispersion correction term varied slightly between water models. The average correction was γ_d =4.4 mN/m and displayed a small decrease with increasing T.

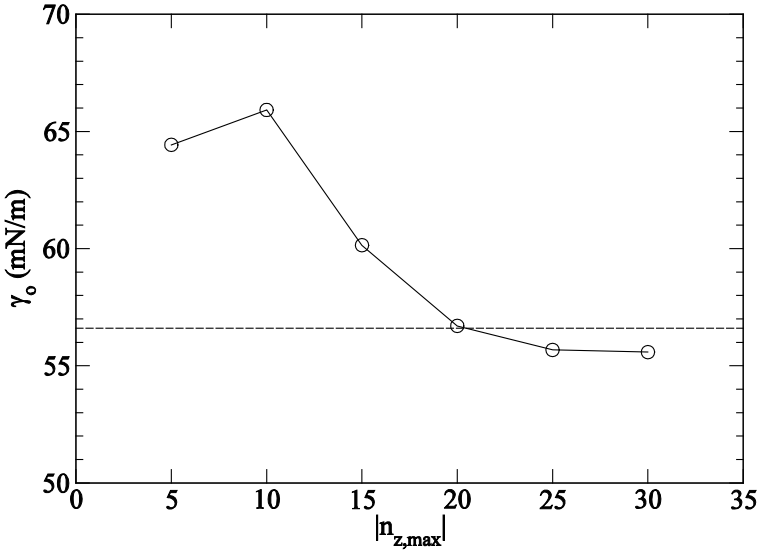
4.3 Results

Before determining the surface tensions of the different models as a function of temperature, it is necessary to investigate the effects of various numerical approximations made during the current and previous simulations. To do so we will focus on the SPC/E model at 300K and quote all surface tension values before the long range dispersion correction (γ_o) and after at least 5 ns of simulation time. The PME result for γ_o is 56.7 mN/m. Estimated standard deviations for the current simulations were 1-2 mN/m, but fluctuations as large as 8 mN/m were observed between 1 ns subaverages. The same results were obtained for the SPC/E model using the double precision version of Gromacs.

Alejandre et al. provided a thorough analysis of surface tension calculations from computer simulations using Ewald sums and emphasized the need to include additional lattice vectors in the reciprocal space sum to account for the increased box dimensions in the extended (formally nonperiodic) z direction.[7] In particular, it is important to maintain a fixed ratio of the maximum number of lattice vectors to box length ($|n_{\alpha,max}|/L_\alpha$) in all three directions, especially at high temperatures. However, the systems simulated here have used the PME approach. This solves the reciprocal space sum using 3D FFT routines and interpolation using a 3D grid. Our calculations were insensitive to the grid resolution (between 0.08 and 0.16 nm) as long as the same grid resolution was maintained in each direction.

In an effort to establish a consistent surface tension for SPC/E water we have reinvestigated the effect of using different numbers of lattice vectors (n_x, n_y, n_z) for the reciprocal space sum by determining the molecular virial, with the electrostatic contribution given by Equation A10 from Reference (7), using 50000 configurations spanning 5 ns obtained from the PME based simulations. In all calculations $|n_{x,max}| = |n_{y,max}| = 5$, while $|n_{z,max}|$ was varied from 5 to 30. We note that the effect of increasing the number of lattice vectors had a negligible effect on the system energy. This is the same approach as presented by Alejandre et al[7], but performed at a different temperature and expanded to include averaging over multiple configurations. The resulting surface tensions are displayed in Figure 4.1. It is clear from Figure 4.2 that one requires values of $|n_{z,max}| \geq 25$ in order to obtain reliable surface tension values even at 300K. Furthermore, the limiting value now agrees with the atomic virial based PME results from the Gromacs programme within the statistical error. The difference in values can be traced to the contribution of P_{zz} which decreased as the number of lattice vectors in the z direction was increased. Unfortunately, although Alejandre et al. recognized this issue and increased the number of lattice vectors in the z direction accordingly, they only included a relatively limited number of additional lattice vectors ($|n_{z,max}| = 10$) in their calculations at low temperatures. The effect of using a limited number of lattice vectors is to overestimate the surface tension. This issue has also been raised by Ismail et al. using an argument based on the Particle-Particle Particle-Mesh (PPPM) mesh size.[5] The recent study of Wemhoff and Carey maintained the appropriate ratio of lattice vectors.[19] They observed lower values of the surface tension for SPC/E water, which is in agreement with the present results.

Figure 4.2. The simulated surface tension (γ_o) for SPC/E water at 300K as a function of the maximum number of lattice vectors in the z direction. The dashed line is the PME result of 56.7 mN/m. The data refer to a system of 512 waters in a box 1.97x1.97x10.0 nm using |nx,max| = |ny,max| = 5 and no long range dispersion correction. The real space contribution (electrostatic plus Lennard-Jones) is 52.3 mN/m



A second approximation which can affect the simulated surface tension values involves the use of the PPPM method for determining the electrostatic interactions during the simulation, followed by the use of the Ewald based virial expression to obtain the electrostatic contribution to the components of the pressure tensor.[8] This approach is often adopted as the calculation of the pressure using PPPM electrostatics is nontrivial and computationally inefficient.[25] We performed an equivalent simulation with the PPPM approach using a 40x40x200 mesh. Analysis of the resulting trajectory using the Ewald virial equation provided a surface tension of 65.3mN/m. This is significantly higher than the PME based result of 56.7 mN/m. Hence, either the PME and PPPM methods produce different results, and/or one has to be consistent when determining the pressure tensor. Evidence for the former comes from the fact that the bulk liquid densities obtained from the two simulations are somewhat different. We find a liquid phase density of 0.987 g/cm^3 for the PME approach compared to a value of 1.017 g/cm^3 for PPPM.

Evidence for the latter comes from the recent study of Ismail et al. where their PPPM results for several water models (TIP3P, TIP4P) are in good agreement with our PME results (see below).[5]

In many simulations it is common to use SHAKE[31] to constrain the water molecule geometry. However, the SHAKE algorithm involves an iterative procedure to satisfy the constraint equations to within a predetermined tolerance. Our simulations used Settle,[27] an analytical version of SHAKE developed for simple water models, which solves the constraint equations exactly. Simulations performed using SHAKE and a relative tolerance of either 10^{-4} or 10^{-5} resulted in surface tensions of 61 and 59 mN/m, respectively, after 5 ns of simulation. Only after a further 5 ns of simulation did the average surface tension decrease to a result consistent with the value obtained using Settle. Hence, to obtain precise values using SHAKE it appears that one requires significantly longer simulations than have been used previously. This probably reflects the large contribution of the constraint forces to the virial, which can only be approximated by the usual implementation of SHAKE, and therefore provides an additional source of noise. The same conclusion was obtained after analysis of the corresponding trajectory using the molecular based virial.

The use of 3D Ewald sums for slab geometries has been investigated by several authors and found to incorrectly reproduce the characteristics of the nonperiodic dimension even when using a relatively large vacuum regions.[33] A simple correction for this problem has been suggested by Yeh and Berkowitz.[33] To our knowledge, the effect of this correction on the surface tension values of water models has not been studied. A simulation performed using the above correction results in an initial surface tension of 56.3 mN/m. This suggests that the error arising from the use of a 3D Ewald summation with tinfoil boundary conditions for slab geometries does not significantly affect the simulated surface tension values of pure water.

4.4 Conclusions

In summary, conflicting values of the surface tension of SPC/E water have been observed most of which can be traced to a variety of numerical issues. Previous studies have either used too few lattice vectors in the z direction,[7,9] or combined the PPPM and PME methodologies,[8] or used SHAKE with a relatively short simulation time.[7-10] Consequently, our final value of 61.1 mN/m for the SPC/E model at 300K is lower than most previous estimates.[7-9] It is, however, still higher than the value of 55.4 mN/m recently determined by Ismail et al. using the PPPM approach.[5] Interestingly, their corresponding values for the TIP3P and TIP4P models are in excellent agreement with the results obtained here (see below). It is currently unclear why the data differ for just the SPC/E model. Our new value for the SPC/E model at 300 K is also in good agreement with a recently quoted value of 62 mN/m.[29] Finally, we also examined the possibility of system size effects by simulating a larger system containing 4340 waters in a box with dimensions of 4 x 4 x12 nm.[3] The value for the surface tension was 62.1 mN/m and in very good agreement with the smaller system size result. Therefore, system size effects seem to be negligible.[10]

TABLE 4.1. Simulated surface tensions (γ in mN/m) of various water models as a function of temperature. Experimental data were taken from Ref(12) and can be represented by the equation $\gamma(T)= 94.74+1.87*10^{-3}T-2.63*10^{-4}T^2$ between 273 and 373 K. Typical estimated errors in the simulated values were 1–2 mN/m.

Model	275 K	300K	325K	350K
TIP3P	54.0	49.5	44.5	41.7
SPC	59.7	53.4	49.0	45.5
SPC/E	64.5	61.3	58.0	52.7
TIP4P	61.0	54.7	50.8	46.7
TIP5P	57.1	52.3	46.1	42.4
TIP6P	64.8	61.8	55.4	52.8
Expt	75.4	71.6	67.6	63.2

In the present study consistent surface tension values were obtained for the PME based electrostatic energy, the Settle algorithm for constraints, and long 2-5 ns simulations. Using this approach the calculated surface tensions of six different water models at four different temperatures were determined and are presented in Table 4.1. The SPC/E and TIP6P models provide the best agreement with experiment at all temperatures. In addition, it is satisfying that our surface tension value for TIP4P at 300K is in agreement with the limiting value obtained by Zakharov et al. [30] for water droplets. which was determined using rigid water molecules and the standard Coulomb potential with no truncation or periodicity effects. The revised value of the surface tension of SPC/E water, while displaying the best agreement with experiment of the simple water models, still underestimates the experimental values by 15%. In our opinion this is to be expected as the SPC/E model was: i) developed for bulk water properties;[6] ii) is nonpolarizable;[31] and iii) overestimates the diffusion constant of water thereby suggesting insufficient hydrogen bonding even in bulk solution.[32] Nevertheless, these computationally

efficient water models are consistently used in simulations of interfacial systems and it is therefore important to know the appropriate surface tension values.

Reference List

1. Wilson, M. A.; Pohorille, A.; Pratt, L. R. Molecular-Dynamics of the Water Liquid Vapor Interface. *Journal of Physical Chemistry* **1987,** *91* (19), 4873-4878.

2. Matsumoto, M.; Kataoka, Y. Study on Liquid Vapor Interface of Water .1. Simulational Results of Thermodynamic Properties and Orientational Structure. *Journal of Chemical Physics* **1988,** *88* (5), 3233-3245.

3. Zhu, S. B.; Fillingim, T. G.; Robinson, G. W. Flexible Simple Point-Charge Water in A Self-Supporting Thin-Film. *Journal of Physical Chemistry* **1991,** *95* (2), 1002-1006.

4. Lie, G. C.; Grigoras, S.; Dang, L. X.; Yang, D. Y.; Mclean, A. D. Monte-Carlo Simulation of the Liquid-Vapor Interface of Water Using An Ab-Initio Potential. *Journal of Chemical Physics* **1993,** *99* (5), 3933-3937.

5. Ismail, A. E.; Grest, G. S.; Stevens, M. J. Capillary waves at the liquid-vapor interface and the surface tension of water. *Journal of Chemical Physics* **2006,** *125* (10), 109902.

6. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.

7. Alejandre, J.; Tildesley, D. J.; Chapela, G. A. Molecular-Dynamics Simulation of the Orthobaric Densities and Surface-Tension of Water. *Journal of Chemical Physics* **1995,** *102* (11), 4574-4583.

8. Shi, B.; Sinha, S.; Dhir, V. K. Molecular dynamics simulation of the density and surface tension of water by particle-particle particle-mesh method. *Journal of Chemical Physics* **2006,** *124* (20), 204715/1-204715/7.

9. Lu, Y. J.; Wei, B. Second inflection point of water surface tension. *Applied Physics Letters* **2006,** *89* (16), 164106/1-164106/3.

10. Alejandre, J.; Tildesley, D. J.; Chapela, G. A. Molecular-Dynamics Simulation of the Orthobaric Densities and Surface-Tension of Water. *Journal of Chemical Physics* **1995,** *102* (11), 4574-4583.

11. Shi, B.; Sinha, S.; Dhir, V. K. Molecular dynamics simulation of the density and surface tension of water by particle-particle particle-mesh method. *Journal of Chemical Physics* **2006,** *124* (20), 204715/1-204715/7.

12. *Handbook of Chemistry and Physics* (CRC, Boca Raton, FL). **1986**.

13. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983,** *79* (2), 926-935.

14. Berendsen, H. J. C.; Postma, J. P. M. van Gunsteren, W. F., Hermans, J. Interaction models for water in relation to protein hydration. *Jerusalem Symposia on Quantum Chemistry and Biochemistry* **1981**, 14, 331-42.

15. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *Journal of Physical Chemistry* **1987,** *91* (24), 6269-6271.

16. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *Journal of Chemical Physics* **1983,** *79* (2), 926-935.

17. Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics* **2000,** *112* (20), 8910-8922.

18. Nada, H.; van der Eerden, J. P. J. M. An intermolecular potential model for the simulation of ice and water near the melting point: A six-site model of H2O. *Journal of Chemical Physics* **2003,** *118* (16), 7401-7413.

19. Wemhoff, A. P.; Carey, V. P. Surface tension prediction using characteristics of the density profile through the interfacial region. *International Journal of Thermophysics* **2006,** *27* (2), 413-436.

20. Ismail, A. E.; Grest, G. S.; Stevens, M. J. Capillary waves at the liquid-vapor interface and the surface tension of water. *Journal of Chemical Physics* **2006,** *125* (10), 014702.

21. Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Gromacs - A Message-Passing Parallel Molecular-Dynamics Implementation. *Computer Physics Communications* **1995,** *91* (1-3), 43-56.

22. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001,** *7* (8), 306-317.

23. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular-Dynamics with Coupling to An External Bath. *Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

24. Harris, J. G. Liquid Vapor Interfaces of Alkane Oligomers - Structure and Thermodynamics from Molecular-Dynamics Simulations of Chemically Realistic Models. *Journal of Physical Chemistry* **1992,** *96* (12), 5077-5086.

25. Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - An N.Log(N) Method for Ewald Sums in Large Systems. *Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.

26. Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids* (Clarendon, Oxford, 1987). 1987.

27. Miyamoto, S.; Kollman, P. A. Settle - An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992,** *13* (8), 952-962.

28. Blokhuis, E. M.; Bedeaux, D.; Holcomb, C. D.; Zollweg, J. A. Tail Corrections to the Surface-Tension of A Lennard-Jones Liquid-Vapor Interface. *Molecular Physics* **1995,** *85* (3), 665-669.

29. Wynveen, A.; Bresme, F. Interactions of polarizable media in water: A molecular dynamics approach. *Journal of Chemical Physics* **2006,** *124* (10), 1045021-1045028

30. Zakharov, V. V.; Brodskaya, E. N.; Laaksonen, A. Surface tension of water droplets: A molecular dynamics study of model and size dependencies. *Journal of Chemical Physics* **1997,** *107* (24), 10675-10683.

31. Rivera, J. L.; Starr, F. W.; Paricaud, P.; Cummings, P. T. Polarizable contributions to the surface tension of liquid water. *Journal of Chemical Physics* **2006,** *125* (9), 0947121-0947128.

32. Chitra, R.; Smith, P. E. A comparison of the properties of 2,2,2-trifluoroethanol and 2,2,2-trifluoroethanol/water mixtures using different force fields. *Journal of Chemical Physics* **2001,** *115* (12), 5521-5530.

33. I. C. Yeh and M. L. Berkowitz, Ewald summation for systems with slab geometry. *Journal of Chemical Physics* **1999**, *111*, 3155.

CHAPTER 5 - Theory and Computer Simulation of Solute Effects on the Surface Tension of Liquids

The Journal of Physical Chemistry B **2008**, 112, 8975–8984

Abstract

A complete description of the thermodynamics of planar mixed solute-solvent interfaces suitable for the analysis of computer simulation data is provided. The approach uses surface probability distributions to characterize the interface regions, coupled with radial distribution functions and the Kirkwood-Buff theory of solutions to characterize the bulk solution properties. The approach is then used to understand the relationship between changes in the surface tension, the degree of surface adsorption or depletion, and the bulk solution properties of several aqueous solute systems. Here we will, provide examples of a surface excluded solutes, such as NaCl, GdmCL, and urea. Aqueous methanol, solutions will be investigated as example of a surface adsorbed solute. The numerical results support the theoretical relationships described here and provide a consistent picture of the thermodynamics of solution interfaces involving any number of components which can be applied to a wide variety of systems.

5.1 Introduction

The distribution of solutes at the solution/vapor interface has important consequences in chemistry and chemical engineering. Consequently, a variety of experimental and theoretical approaches have been used to help understand surface adsorption or exclusion.[1–5] It is well established that an increase in the surface tension of a solution due to the addition of a solute indicates exclusion of that solute from the interface region, and vice versa.[6] However, atomic

level detail concerning the surface distributions has been more difficult to achieve. In principle, molecular simulations can provide such detail. Simulations require an accurate force field and an adequate degree of sampling of the property of interest if the results are to be interpreted with any confidence. Recently, a wealth of information has been provided by simulations of solutes at the water vapor (or vacuum) interface.[7–18] We will not discuss all the results here as they have been summarized in several recent reviews.[19–22] The main issues arising from a simulation perspective are the observed dependence of the surface structure on the force field used,[8,14] and the long times required to obtain precise values for the surface tension and surface structure probability distributions.[8,23] In particular, the use of polarizable force fields has generally indicated an increased probability for the location of larger, more polarizable, anions at the surface.[8] However, the surface adsorption must be accompanied by a sub surface depletion in order for the Gibbs adsorption equation to be obeyed.[15,19] Unfortunately, it has been difficult to fully quantify such structural and thermodynamic changes so that they may be compared with experimental data on surface adsorption. This is a major focus of the current study.

5.2 Background and Theory

5.2.1 Thermodynamics of Surfaces.

The general theory of surface adsorption has been established for over a century. However, the analysis of computer simulation data on surface adsorption is rather new. We will briefly derive the Gibbs adsorption to the Kirkwood-Buff (KB) theory of solutions, which we will utilize later. KB theory has been used previously to analyze experimental surface adsorption data,[24] albeit in a slightly different form. To our knowledge it has not been used to analyze simulation data on the thermodynamics of surface tension changes due to additives. Throughout

this article we shall refer to a mixture of a solvent (1), usually water, and any number of additional solutes (2, 3, 4, etc.). Consider the planar surface region of a solution in contact with a vacuum or vapor region containing a negligible number of molecules on one side and bulk solution on the other. The surface lies in the xy plane and is perpendicular to the *z* axis. The Gibbs-Duhem equations for the surface and the bulk regions at constant *T* and *P* are given by

$$A d\gamma + N_1 d\mu_1 + N_2 d\mu_2 + N_3 d\mu_3 + \quad \cdots \quad = 0 \qquad \text{surface} \qquad (5.1a)$$

$$n_1 d\mu_1 + n_2 d\mu_2 + n_3 d\mu_3 + \quad \cdots \quad = 0 \qquad \text{bulk} \qquad (5.1b)$$

respectively. Here, A is the surface area of the interface, γ is the surface tension, and μ_i is the chemical potential of species *i*. The N_i values represent the number of solvent and solute molecules in the interface region, while the n_i values represent the corresponding number in the bulk solution. At equilibrium, the chemical potentials of each species are the same in the surface and bulk regions. The number of molecules in the surface region differs from that in the bulk due to the perturbing effect of the interface which requires a redistribution of solution components in order to maintain equilibrium with the bulk solution. One can use the bulk equation to eliminate the solvent chemical potential change ($d\mu_1$) from the surface equation to give

$$A d\gamma = -\left(N_2 - \frac{n_2}{n_1} N_1 \right) d\mu_2 - \left(N_3 - \frac{n_3}{n_1} N_1 \right) d\mu_3 - \quad \cdots \qquad (5.2)$$

or,

$$d\gamma = -\Gamma_{2,1}\, d\mu_2 - \Gamma_{3,1}\, d\mu_3 - \quad \cdots \qquad (5.3)$$

where $\Gamma_{i,1}$ is the usual Gibbs excess surface adsorption (per unit surface area) of each solute *i* relative to the solvent (1). Taking derivatives of Equation 5.3 one finds,

$$\left(\frac{\partial \gamma}{\partial \rho_2} \right)_{T,P} = -\Gamma_{2,1} \left(\frac{\partial \mu_2}{\partial \rho_2} \right)_{T,P} - \Gamma_{3,1} \left(\frac{\partial \mu_3}{\partial \rho_2} \right)_{T,P} - \quad \cdots \qquad (5.4)$$

which is the same result that has been derived previously.[6,25]

An expression for N_i can be obtained from the (one dimensional) surface probability distribution functions $g_i(z)$,

$$N_i = A\rho_i \int_0^Z g_i(z)\, dz \tag{5.5}$$

where $\rho_i = n_i/V$ is the corresponding bulk number density (or molar concentration) of species *i*. The distance Z defines the extent of the surface region from some (arbitrary) origin in the vacuum or vapor phase. Beyond Z, the solute and solvent distributions are the same as that in the bulk solution. The excess surface adsorption of solute *i* per unit surface area is then given by,

$$\Gamma_{i,1} = \frac{1}{A}\left(N_i - \frac{n_i}{n_1} N_1 \right) = \rho_i \int_0^Z [g_i(z) - g_1(z)]\, dz = \rho_i I_{i,1} \tag{5.6}$$

and the integration limit can be formally extended to Z $\to\infty$, i.e. one does not need to know the extent of the surface region. The change in surface tension and the excess adsorption of a species (as defined by Equation 5.6) can be obtained relatively easily from computer simulations.[26,27] Even so, we note that no such integration has been performed in recent simulation studies.[8,10,20,28] The integral ($I_{i,1}$) over the surface probability distributions can be viewed as a measure of the surface structure. The derivatives on the r.h.s. of Equation 5.4 are seemingly more difficult to obtain and have not traditionally been used to analyze computer simulation data on surface adsorption. However, we note that they are properties of the bulk solution.

5.2.2 The Kirkwood-Buff Theory of Solutions

A central aspect of the current approach involves the application of Kirkwood-Buff (KB) theory to relate the simulation data obtained for bulk solution properties to the corresponding activity derivatives.[29,30] Previous studies have used KB theory to investigate a variety of

experimental and theoretical solvation effects.[31-33] It is an exact theory of solutions. In particular, it is important to emphasize that KB theory does not involve any approximations or limitations concerning the size or character of the molecules to which it can be applied.[32] The theory relates several properties of solution mixtures (containing any number of components) to KB integrals which are defined by,[29]

$$G_{ij} = G_{ji} = 4\pi\int_0^{\infty}[g_{ij}^{\mu VT}(r)-1]r^2dr \approx 4\pi\int_0^{R}[g_{ij}^{NPT}(r)-1]r^2dr \qquad (5.7)$$

where $g_{ij}(r)$ is the radial distribution function (rdf) in the grand canonical (*:VT*) ensemble between species *i* and *j* as a function of their intermolecular separation *r*. The negligible approximation in Equation 5.7 is used to enable the determination of KB integrals in closed (NPT) systems,[26,29,30] as this is the more typical system to be studied. In this case, the integral is truncated at a distance R beyond which the rdfs are essentially unity. KB integrals can be obtained from experimental data (densities, compressibilities and activities) on solution mixtures,[25] or directly from computer simulations.[28,31,32] In principle, expressions for the required chemical potential derivatives can be obtained for any number of solute components.[24,26] In practice, the expressions become rather cumbersome as the number of components increases and so we will focus on the expressions provided for a binary system of a single solute (2) in a solvent (1). Expressions for ternary systems are available.[27,33,34.]

One property of solutions that will prove particularly useful during the present discussion is the change in solute activity (a_2) with solute concentration in a binary solution as defined by,

$$a_{22} = \beta\left(\frac{\partial\mu_2}{\partial\ln\rho_2}\right)_{T,P} = \left(\frac{\partial\ln a_2}{\partial\ln\rho_2}\right)_{T,P} = \frac{1}{1+\rho_2(G_{22}-G_{12})} \qquad (5.8)$$

where $\beta = 1/RT$, and R is the gas constant. To enable the transformation between different concentration scales we will also require the KB expression for the partial molar volume (pmv) of the solvent,[24]

$$\overline{V_1} = \frac{1 + \rho_2 (G_{22} - G_{12})}{\eta} = \frac{1}{\eta\, a_{22}} \tag{5.9}$$

where $\eta = \rho_1 + \rho_2 + \rho_1\, \rho_2\, \Delta G_{12}$ and $\Delta G_{12} = G_{11} + G_{22} - 2\, G_{12}$. All the integrals in Equations 5.8 and 9 can be obtained from simulations performed at the appropriate bulk solution compositions.

The application of KB theory to neutral salt solutions is somewhat more involved.[35] We will use the indistinguishable ion approach where the solute number density in Equations 5.8 and 5.9 refers to the total ion concentration, and the KB integrals are determined by ignoring the identity of the individual ions.[32,35] The consequences of this approach for determining the surface excess or deficit of salts, as provided by Equation 5.6, will be discussed later. Hence, we shall distinguish between the normal molar (or molal) salt concentration (c_s or m_s), and the total ion concentration ($\rho_2 = n_{\pm}\, c_s$ or $m_2 = n_{\pm}\, m_s$) for a salt that releases a total of $n_{\pm} = n_+ + n_-$ ions in solution. Therefore, we also have $d\mu_s = n_{\pm} d\mu_2$, $d\ln c_s = d\ln \rho_2$, and $d\ln a_2 = d\ln(y_2 \rho_2) = d\ln(y_{\pm} c_s)$. More details can be found in the literature.[32,36,37.]

5.2.3 Combined Approach for Binary Systems

Here we combine the usual surface adsorption approach with the results provided by the KB theory of solutions to provide a consistent picture of the thermodynamics of surfaces in terms of distribution functions in binary systems – all of which can be obtained from computer simulations. The surface probability distribution functions are normalized by reference to the corresponding bulk distribution. Therefore, one can express the change in surface tension with solute concentration as,

$$\beta\left(\frac{\partial\gamma}{\partial\rho_2}\right)_{T,P} = -a_{22}\frac{\Gamma_{2,1}}{\rho_2} = -a_{22}\int_0^z \left[g_2(z) - g_1(z)\right]dz = -a_{22}I_{2,1} \qquad (5.10)$$

where a_{22} is determined for the corresponding bulk solution, and is usually positive for real solutions (see later). We have used a_{22} rather than $\partial\mu_2/\partial\rho_2$ as this removes the inherent singularity in the latter at low solute concentrations. Hence, if the change in surface tension is proportional to the solute concentration, as observed for most salts,[38] then the changes in the surface distribution of solute and solvent molecules mirrors the changes in a_{22}. Equation 5.10 is exact and all contributions can be obtained from computer simulations. Therefore, an accurate simulated value of a_{22} will be important for a correct description of surface adsorption. Unfortunately, many common force fields we have tested do not (a priori) reproduce the experimental KB integrals and a_{22} values for binary solutions.[39-43] We note that Equation 5.10 is directly analogous to equations derived for the treatment of solute effects on peptides and proteins.[32,44,45]

In many cases it is more convenient to express the surface tension changes using other solute concentration scales. To achieve this one can use the following standard thermodynamic transformations between the molar and molal (m), or mole fraction (x), concentrations in binary solutions,

$$\left(\frac{\partial\rho_2}{\partial m_2}\right)_{T,P} = \rho_1^2\overline{V_1} = \rho_1\phi_1 \qquad (5.11)$$

and

$$\left(\frac{\partial\rho_2}{\partial x_2}\right)_{T,P} = \rho^2\overline{V_1} \qquad (5.12)$$

where ϕ_1 is the volume fraction of the solvent, and $\rho = \rho_1 + \rho_2$ is the total number density. Consequently, we have the expression,

$$\beta\left(\frac{\partial\gamma}{\partial x_2}\right)_{T,P} = -\frac{\rho^2}{\eta}\frac{\Gamma_{2,1}}{\rho_2} = -\frac{\rho}{\eta}\frac{\Gamma_{2,1}}{x_2} \tag{5.13}$$

for the mole fraction derivative and,

$$\beta\left(\frac{\partial\gamma}{\partial m_2}\right)_{T,P} = -\frac{\rho_1^2}{\eta}\frac{\Gamma_{2,1}}{\rho_2} = -\frac{\rho_1}{\eta}\frac{\Gamma_{2,1}}{m_2} \tag{5.14}$$

for the molality derivative. We note that the bulk solute molality is given by $m_2 = n_2/n_1$ to within a conversion factor of $1000/M_1 = 55.51$ mol/Kg for water.

Analysis of the stability criteria for binary solution mixtures indicates that the value of η must be positive for miscible solutions.[26] Consequently, the value of a_{22} will also be positive for compositions where the pmv of the solvent is positive (see Equation 5.9). This is the case for the majority of water mixtures. Hence, the sign of the surface excess in Equations 5.10, 5.13 and 5.14 is given by the slope of the surface tension against solute concentration using any concentration scale. The conditions for ideal bulk solution behavior depend on the concentration scale being used. On the molar concentration scale an ideal solution ($a_2 = \rho_2$) is characterized by $G_{22} - G_{12} = 0$ (or $a_{22} = 1$) for all compositions. Alternatively, on the mole fraction scale one has $\Delta G_{12} = 0$ (or $\eta = \rho$), while ideal behavior on the molality scale is provided by $\rho_1\Delta G_{12} = -1$ (or $\eta = \rho_1$).

The above expressions have several consequences for the analysis of computer simulation data which, to our knowledge, have not been considered in previous studies. First, to fully and accurately describe the thermodynamics and structure of surfaces (as provided by g_i) the force fields used must reproduce the change in surface tension with solute concentration, the values of g_i (distribution of solute and solvent at the surface), and a_{22} (or η) a property of the bulk solution. Actually, the above equation is exact and therefore only two of the three need be correct. Second, the integration over $g_2 - g_1$ should be performed to a distance at which the integral

remains unchanged, i.e. $g_2 = g_1 = 1$, which may involve many solvation shells away from the surface. The

5.3 Some Simple Cases

Here we will investigate several simplified expressions that are obtained in a few limiting cases. These situations provide some useful reference points for real solutions. The first arises when the surface tension change is proportional to the solute molarity – which is common for many salt solutions.[38] Therefore, one has $\Delta\gamma = \gamma - \gamma_1 = b\, c_s$, where γ_1 is the surface tension of the pure solvent and $b = (\partial\gamma/\partial c_s)_{T,P}$. Consequently, the surface excess can be obtained from Equation 5.10 which reduces to,

$$\Gamma_{2,1} = -\frac{\beta\, b\, c_s}{a_{22}} = -\frac{\beta\, \Delta\gamma}{a_{22}} \tag{5.15}$$

The above expression is also valid for salt solutions. In this case the surface excess refers to salt "molecules" if the chemical potential (activity) used in the definition of a_{22} refers to the salt ($\mu_s = n_+\mu_+ + n_-\mu_-$), whereas the excess is in terms of ions if the indistinguishable ion approximation is used (as is adopted here). The latter is equivalent to using the mean ionic activity coefficient and the total salt concentration in determining a_{22}. The above expression can be reduced further if one assumes or observes ideal behavior for the bulk solution mixture ($a_{22} = 1$).

The second type of simplification arises for symmetric ideal (SI) solutions. SI solutions have been discussed in detail by Ben-Naim in the context of KB theory.[27] Here, the mole fraction activity coefficients are unity for both species at all compositions in a binary solution. In this case one has $\Delta G_{12} = G_{11} + G_{22} - 2\, G_{12} = 0$ and therefore $\eta = \rho$ for all compositions. Therefore, Equation 5.13 reduces to,

$$\Gamma_{2,1} = -\beta\, x_2 \left(\frac{\partial \gamma}{\partial x_2}\right)_{T,P} \tag{5.16}$$

One can also further assume that the surface tension of an SI mixture of 1 and 2 is given by $\gamma = x_1\gamma_1 + x_2\gamma_2$ where γ_i is the surface tension of pure *i* at the same T and P. The above equation then simplifies to give,

$$\Gamma_{2,1} = -\beta\, x_2\, (\gamma_2 - \gamma_1) \tag{5.17}$$

which provides another useful reference point for solution mixtures. We will label this type of system as SI^2. It is clear that SI^2 solution mixtures still give rise to an excess surface adsorption unless the pure solutions also have similar surface tensions.

Finally, we note that the limiting slope obtained as the solute concentration tends to zero is provided by,

$$\beta\left(\frac{\partial \gamma}{\partial \ln y_2}\right)_{T,P,y_2\to 0} = -\Gamma^{o}_{2,1} \tag{5.18}$$

with y = ρ, x, or m, and where the zero superscript indicates a property of the system at infinite dilution of the solute. Some interesting cases appear in the literature and provide information concerning the structure of the interface at very low solute concentrations as defined by the limiting value of the integral $I_{2,1}$. For salts displaying a linear dependence of the surface tension on the salt concentration one finds,

$$I^{o}_{2,1} = -\frac{\beta\, b}{n_{\pm}} \tag{5.19}$$

In this case the integral refers to the average probability distribution for the ions, i.e. $g_2 = (n_+ g_+ + n_- g_-)/n_{\pm}$. Alternatively, for surface adsorbed solutes an empirical relationship has been developed for low ($x_2 < 0.01$) solute concentrations,[47,48]

$$\gamma = \gamma_1\left[1 - B\log_{10}(1 + x_2 / C)\right] \tag{5.20}$$

where B and C are positive constants. Using this expression one can show that at infinite dilution of the solute we have,

$$I_{2,1}^{o} = \frac{\beta \gamma_1}{2.303\rho_1^{o}} \frac{B}{C} \tag{5.21}$$

and ρ_1^{o} is the number density of pure water. Hence, knowledge of the parameter b, or B and C, can provide quantitative estimates for the nature of the surface structure need to include the exact value of a_{22} or η in experimental studies has been noted previously.[46]

5.4 Methods

All molecular dynamics simulations were performed using the KBFF force fields for NaCl and methanol,[41,49] together with the SPC/E water model,[50] as implemented in the GROMACS program (v3.3.1).[51,52] A time-step of 2 fs was used and the methanol bond lengths were constrained using Lincs,[53] while the water geometry was constrained using SETTLE.[54] A twin range cut-off of 1.0 nm/1.5 nm was employed with a nonbonded pair list update of every 10 steps. Long range electrostatic interactions were evaluated using the PME approach.[55] Simulations performed to determine the surface tension and surface adsorption involved approximately 12,000 atoms in a fixed rectangular box of dimensions 4x4x12 nm with the atoms located in a slab geometry occupying the central 4x4x8 nm and generating two surfaces of area A = 16 nm^2. The slab simulations were performed at constant volume and a constant temperature of 300 K using the weak coupling approach.[56] The surface tensions were then determined from the pressure tensor elements via the following relationship,

$$\gamma = \tfrac{1}{2} L_z [P_{zz} - \tfrac{1}{2}(P_{xx} - P_{yy})] \tag{5.22}$$

as described previously.[57,58] Here, L_z = 12 nm is the box dimension in the extended (z) direction and $P_{\alpha\alpha}$ are the diagonal elements of the pressure tensor. No long range dispersion corrections to

the surface tension values were included as this is nontrivial for mixed solutions. Fortunately, the contribution to the surface tension of the neglected dispersion interactions beyond 1.5 nm is usually negligible, especially as we are focusing on changes in surface tension. Typically, systems were equilibrated for 5 ns and then simulate for a further 15-30 ns to ensure precise results for both the surface tension and surface adsorption data.

The values of a_{22} for the bulk solutions were determined previously during the initial parametrization procedure.[41,49] Both the simulated and experimental data for bulk solutions were taken directly from these earlier studies. The excess adsorption values were determined from the simulations using a simple counting procedure. One can define a local number of solute and solvent molecules which depends on the distance from our arbitrary origin in the vacuum phase, i.e. $N_i(Z)$. The excess adsorption can then be determined as a function of Z. When Z is large enough to approach the bulk solution distribution, the value of $\Gamma_{2,1}$ will tend to the constant value required for the current analysis. In order to evaluate this precisely we have used the following expression for the excess adsorption,[59]

$$A\Gamma_{2,1}(Z) = N_2(Z) - \frac{n_2}{n_1} N_1(Z) = N_2(Z) - \frac{n_2^o - N_2(Z)}{n_1^o - N_1(Z)} N_1(Z) \qquad (5.23)$$

where n_i^o are the initial number of *i* ions or molecules in the system. The values of n_i^o and n_i differ due to the redistribution of molecules to and from the surface in our finite size system. Therefore, the final bulk solution composition n_2/n_1 is somewhat different to the initial composition n_2^o/n_1^o, and this difference can be significant as N_1 can be large. An alternative viewpoint is that any ions or molecules which contribute to the value of N_i should not be considered part of the bulk solution. Using Equation 5.23 one does not have to generate the corresponding rdfs in advance. This can be a problem as the normalization procedure requires the bulk solution concentrations which are unknown if one doesn't know the extent of the

surface region, i.e. Z in Equation 5.5. The bulk solute and solvent concentrations were therefore determined after equilibration by averaging over the central 4 nm of the slab. The total excess surface adsorption for salts involves both the distribution of the cations and anions. Hence, Equation 5.6 becomes,

$$\Gamma_{2,1} = \Gamma_{+,1} + \Gamma_{-,1} = \frac{1}{A}\left(N_{+} - \frac{n_{+}}{n_1} N_1 \right) + \frac{1}{A}\left(N_{-} - \frac{n_{-}}{n_1} N_1 \right) \tag{5.24}$$

or in terms of the probability distribution functions,

$$\Gamma_{2,1} = \rho_{+} \int_0^Z [g_{+}(z) - g_1(z)]\, dz + \rho_{-} \int_0^Z [g_{-}(z) - g_1(z)]\, dz \tag{5.25}$$

and the value of ρ_2 is taken as $\rho_+ + \rho_-$ in Equation 5.6. In all cases the results represent an average over both surfaces within the system. No evaporation was observed for any of the systems.

5.5 Results

To illustrate the relationship between the different terms in Equations 5.10 and 5.13 we have performed molecular dynamics simulations to investigate the effect of several different solutes on the surface thermodynamics of water solutions. The first group of solutes solute were NaCl, GdmCL and urea which produce an increase in the surface tension of the solution from that of pure water and is therefore excluded from the surface. Other solute such as methanol will decrease the surface tension and are therefore adsorbed at the surface. In studying these systems our primary objective was to illustrate and establish the relationships provided by Equations 5.10 and 5.13. We have also compared the results to the corresponding experimental data. However, it should be noted that the agreement with experiment is a function of the quality of force field being used. Being exact, Equations 5.10 and 5.13 should hold even for force fields which provide a poor description of the solution and/or surface thermodynamics.

The above solutes were also chosen as we have developed corresponding force fields which were specifically designed to reproduce the experimental densities and KB integrals (and thereby a_{22} and η) for the solution mixtures.[41,49] The degree of agreement with experiment one can obtain for a_{22} and η is provided in Figures 5.1 – 5.4. The details of the calculations can be found elsewhere.[41,49] The agreement with experiment for the value of a_{22} in NaCl solutions is very good for all but the highest > 4M) concentrations, particularly in comparison with other force fields.[41] Our simulated activity derivatives of GdmCl showed the correct trend with increase of salt concentration but were considerably overestimated over experiment, especially at high concentrations. However, all calculated derivatives were normally within errors of standard derivation of experimental measurements. Furthermore, the GdmCl solution densities are in very good agreement with experiment. KBFF model also reproduced experimental trend and magnitude of a_{22} of urea solution with increasing of solution concentration, which indicates urea solution activities was correctly reproduced by using our KBFF. Even though, values of a_{22} were slightly underestimated at low and very high concentrations. The agreement for the methanol η values is less quantitative compared with NaCl, although the appropriate trend is clearly reproduced. The methanol solution densities are reproduced well at low methanol concentrations but are somewhat lower than experiment for high methanol compositions.[49]

Table 5.1 Simulated surface tensions of aqueous solutions of NaCl, Urea, GdmCl and methanol.

Soln	n_1^o	n_2^o	c_s^o M	c_s M	x_2^o	x_2	γ	T_{sim}
Water	4340		0.00	0.00	0.000	0.000	59.7	30
NaCl	4158	152	0.99	1.09			60.7	20
	4096	308	2.00	2.21			64.4	30
	3900	616	4.01	4.34			67.3	20
Urea	3500	308		3.91			60.6	20
	2724	616		8.05			62.2	20
GdmCl	3650	152		2.20			62.0	20
	3034	306		4.25			64.6	20
	2418	460		5.91			67.0	20
CH_3OH	3808	192			0.048	0.036	55.4	20
	3500	500			0.125	0.106	42.0	20
	2484	828			0.250	0.223	38.4	20
	1836	1102			0.375	0.350	34.7	20
	1320	1320			0.500	0.479	27.7	20
	900	1502			0.625	0.610	29.1	20
	546	1634			0.750	0.731	26.5	20
	250	1744			0.875	0.873	25.1	20
	0	2000			1.000	1.000	26.2	20

Table 5.2. Simulated and experimental surface adsorption data for aqueous solutions of NaCl , Urea, GdmCl and methanol.

Soln	c_s	x_2	a_{22}		η		$\partial\gamma/\partial y$		$\Gamma_{2,1}$	
			MD	exp	MD	exp	MD	exp	MD	exp
Water	0.00	0.000	1.00	1.00	55.1	55.1				
NaCl	1.09		1.03	1.01			1.93	2.08	-0.49	-0.54
	2.21		1.19	1.19			1.93	2.08	-0.86	-0.93
	4.34		1.55	1.72			1.93	2.08	-1.30	-1.26
Urea	3.91		1.02	0.99					-2.5	
	8.05		0.91	1.15					-11	
GdmCl	2.20		0.92	0.78					-17	
	4.25		1.30	0.85					-20	
	5.91		1.40	1.04					-7	
CH_3OH		0.036			54.8	53.2	-129	-215	1.15	1.86
		0.106			54.8	51.7	-81	-117	2.30	3.12
		0.223			55.4	53.3	-50	-54	3.40	3.43
		0350			53.2	56.4	-30	-27	3.47	3.21

The simulated and experimental surface tension values for the different solutes are displayed in Table 5.1 and in Figures 5.1 – 5.4. As expected, NaCl, Urea and GdmCl solutions displayed increases in surface tension from that of pure water, while methanol solutions exhibited a decrease. Unfortunately, the simulated surface tension of pure SPC/E water is low (59.7 mN/m) compared to the experimental value of 71.6 mN/m.[60] We note that the exact value

for SPC/E water has been the subject of several recent studies.[20,58,61-63] Our value is in agreement with the most recent findings.[64-66] On the other hand, the pure methanol value is slightly too high (26.2 mN/m) compared to the value of 21.6 mN/m observed experimentally.[60] However, the correct trends in surface tension are observed with increasing solute concentration. The change in surface tension with NaCl concentration is almost quantitatively reproduced by the current simulations. The changes for other solutes are reasonable. The correct bulk solution compositions were therefore determined from the slab simulations by averaging the solute and solvent concentrations over the central 4 nm of the approximately 8 nm slab. These results are provided in Table 5.2 and used in all subsequent Figures.

Figure 5.1. Thermodynamics of aqueous NaCl solutions. Top: The change in surface tension ($\Delta\gamma$ in mN/m) with solute molarity (c_s). The symbols represent the raw simulation data and the thin line represents the corresponding fit provided by $\gamma = 59.3 + 1.93 c_s$. The thick line represents the corresponding experimental data (extrapolated beyond 1M). Bottom: Bulk solution activity derivative (a_{22}) as a function of solute molarity. The symbols represent the raw simulation data and the thin line represents the corresponding fit. The thick line represents the corresponding experimental data.

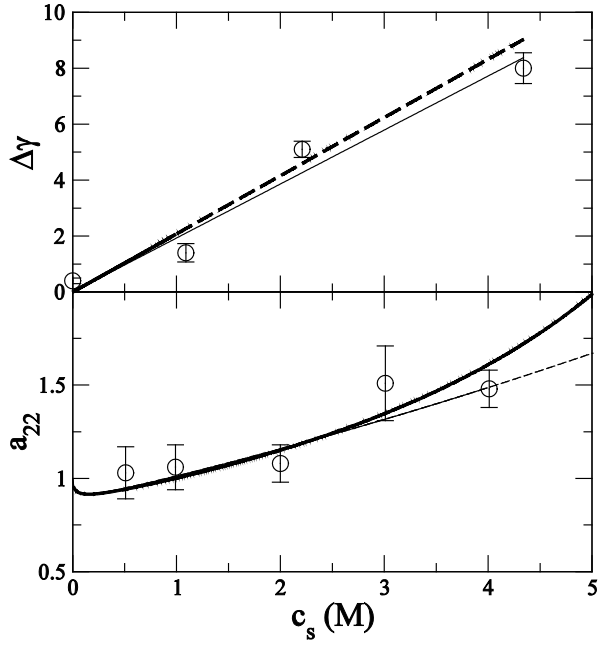


Figure 5.2. Thermodynamics of aqueous urea solutions. Top: The change in surface tension ($\Delta\gamma$ in mN/m) with solute molarity (c_s). The symbols represent the raw simulation data c_s. The thick line represents the corresponding experimental data. Bottom: Bulk solution activity derivative (a_{22}) as a function of solute molarity. The symbols represent the raw simulation data and the thin line represents the corresponding fit. The thick line represents the corresponding experimental data.

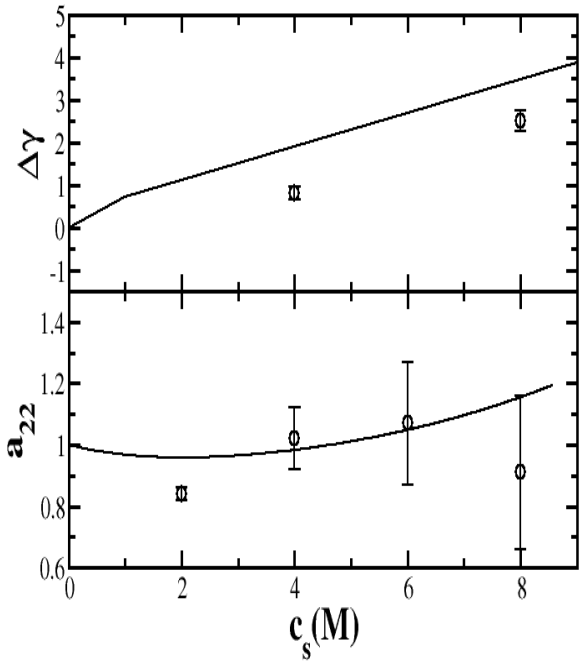


Figure 5.3. Thermodynamics of aqueous GdmCl solutions. Top: The change in surface tension ($\Delta\gamma$ in mN/m) with solute molarity (c_s). The symbols represent the raw simulation data c_s. The thick line represents the corresponding experimental data. Bottom: Bulk solution activity derivative (a_{22}) as a function of solute molarity. The symbols represent the raw simulation data and the thin line represents the corresponding fit. The thick line represents the corresponding experimental data.

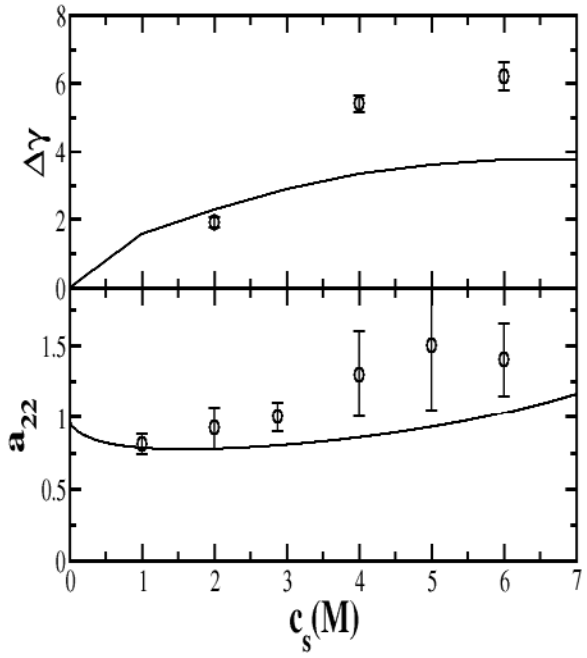
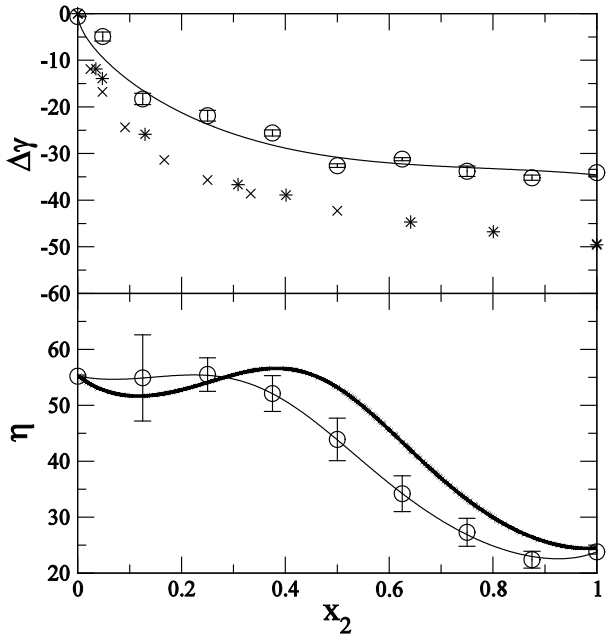


Figure 5.4. Thermodynamics of aqueous methanol solutions. Top: The change in surface tension ($\Delta\gamma$ in mN/m) with solute mole fraction (x_2). The circles represent the raw simulation data and the thin line represents the corresponding fit provided by $\gamma = 60.3 - 33.80x^{0.5} - 45.43x + 81.85x^2 - 37.20x^3$. Other symbols represent two corresponding experimental data sets. Bottom: Bulk solution η values (in M) as a function of solute mole fraction. The symbols represent the raw simulation data and the thin line represents the corresponding fit. The thick line represents the corresponding experimental data.



In the section, NaCl and CH_3OH solutions have been chosen to further evaluate of equation 5.12. Using the fitting equation $\Delta\gamma = a\ x^{0.5} + b\ x + c\ x^2 + d\ x^3$, together with our previous determinations of a_{22} and η, the corresponding experimental values of $\partial\gamma/\partial y$ ($y = c_s$ or x_2) and $\Gamma_{2,1}$ were determined for both solutes. The analogous simulated values corresponding to the bulk solvent compositions studied here were obtained by using the same fitting equations applied to the simulated surface tension data. In addition, the surface adsorption was determined directly from the trajectories using Equations 5.5 and 5.6. The data are provided in Table 5.2.

The surface depletion data for NaCl solutions is displayed in Figure 5.5. It can be seen that the surface depletion predicted from the surface tension changes and the values of a_{22}

(Equation 5.10), are in complete agreement with the surface depletion obtained directly from Equation 5.5 and 5.6. Hence, we have numerical proof of the validity of Equation 5.10 and the associated analysis of computer simulation data. Furthermore, the simulated and experimental data are in excellent agreement providing a high degree of confidence in the underlying surface distributions. Assuming an ideal solution (a_{22} = 1) is clearly acceptable at low (1M) salt concentrations, but produces significant errors at higher concentrations. Both the simulated and experimental adsorption values display an initial linear dependence on solute concentration, which then levels off at higher solute concentrations. Interestingly, the curve resembles an inverted Langmuir binding isotherm.

Figure 5.5. Surface adsorption ($\Gamma_{2,1}$ in ions/nm^2) of aqueous NaCl solution/vacuum interfaces as a function of solute molarity (c_s). The symbols correspond to the results obtained after integrating the surface probability distributions (Equation 5.6). The solid line is the expected surface exclusion as determined from Equation 5.10 using the simulated values of the surface tension derivative and a_{22}. The thick line is the corresponding experimental result after extrapolation (dotted line). The straight line corresponds to the surface exclusion expected using the experimental surface tension derivative with a_{22} = 1 (an ideal solution).

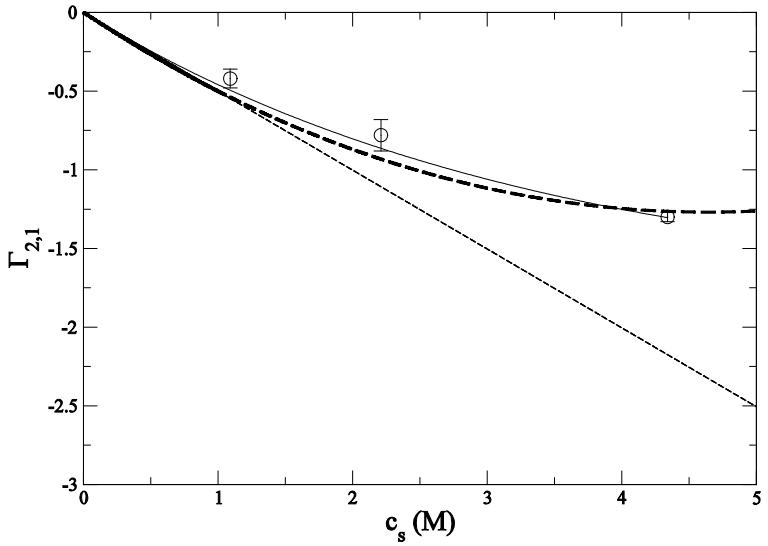
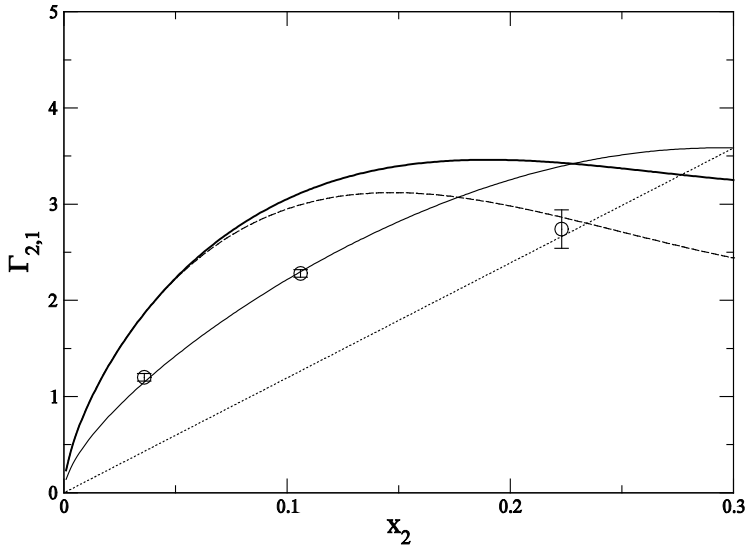


Figure 5.6. Surface adsorption ($\Gamma_{2,1}$ in molecules/nm^2) of aqueous methanol solution/vacuum interfaces as a function of solute mole fraction (x_2). The symbols correspond to the results obtained after integrating the surface probability distributions (Equation 5.6). The solid line is the expected surface exclusion as determined from Equation 5.12 using the simulated values of the surface tension derivative and η. The thick line is the corresponding experimental result after averaging over two data sets. The dashed line corresponds to the experimental surface exclusion expected for SI solutions (Equation 5.16), while the straight dotted line corresponds to the experimental surface exclusion expected for SI^2 solutions (Equation 5.17).



The same approach has been applied to the methanol solutions and the results are displayed in Figure 5.6. Methanol displays the expected positive surface adsorption. Again, the validity of Equation 5.13 is verified by the simulated data obtained at low methanol mole fractions. The simulated data for $x_2 = 0.223$ displays some disagreement. In our opinion, this is a minor problem which is probably a consequence of either: i) errors associated with the fitting procedure; and/or ii) an inability to accurately model the rather large surface excesses using our relatively small system sizes. Hence, the agreement between experiment and simulation is less satisfactory for this system. The data presented in Table 5.2 indicate that the disagreement with

experiment is essentially due to an incorrect description of the surface tension changes, and not due to differences between the η values. Assuming an ideal solution (η = ρ) is clearly acceptable at low (< 0.1) methanol mole fractions, but leads to errors at higher concentrations.

If the simulated values of $\Delta\gamma$ and a_{22} are correct then Equation 5.10 indicates that the value of $\Gamma_{2,1}$ must also be correct. This is the case for the NaCl solutions studied here. The simulations, however, also allow us to decompose the integral in Equation 5.10 to investigate the surface structure via $g_i(r)$. Some of the surface distribution functions are displayed in Figure 5.7, together with the corresponding values of $\Gamma_{2,1}$ as a function of the integration distance Z. The water distribution for 2.2M NaCl indicates a sharp interface region. The corresponding sodium and chloride ion distributions confirm exclusion of the ions from the immediate surface with no other significant structural features. This exclusion is expected due to the increase in surface tension on addition of salt and presumably arises due to the large desolvation penalty required to locate the ions at the surface. Interestingly, the sodium and chloride ions display similar behavior. If we define the region of exclusion as that between $g_2(z_2) = 0.5$ and $g_1(z_1=0) = 0.5$, then $z_2 - z_1 =$ 0.3 nm for both ions. The degree of exclusion can be quantified from the distribution functions as indicated in Equations 5.5 and 5.6. The results as a function of integration distance are also displayed in Figure 5.7. Exclusion of -0.9 ions/nm^2 was observed for 2.2M NaCl solutions and is in quantitative agreement with the experimental value. The values of $\Gamma_{2,1}(Z)$ reach a constant value after 0.5 nm from the interface region – corresponding to the point where the surface distribution equals the bulk distribution. The contribution from each ion to the final value of $\Gamma_{2,1}$ was the same and is required to maintain electroneutrality in the surface region.[36,37]

Figure 5.7. Surface properties of an aqueous 2.2M NaCl solution/vacuum interface. The origin has been shifted so that $g_1(z = 0) = 0.5$ for convenience. Top: Surface probability distributions (g_i) for water, sodium, and chloride ions as a function of distance from the interface (z). Middle: Surface adsorption ($\Gamma_{2,1}$ in ions/nm^2) as a function of integration distance (Z) for all ions (black line and circle), sodium, and chloride ions. The thin dashed line is the expected surface exclusion as determined from Equation 5.10 using the simulated values of the surface tension derivative and a_{22}. The thick dashed line corresponds to the experimental surface exclusion provided by Equation 5.10. The adsorptions observed at Z = 1 nm were taken as the final simulated values. Bottom: The average salt molality (m_s) as a function of distance from the interface (z) obtained from the simulations. The dashed line is the average bulk molality (2.32m).

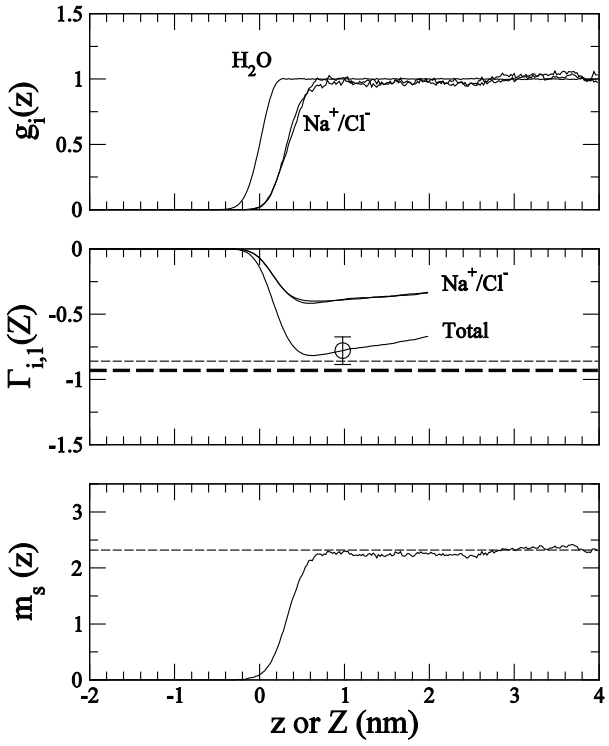
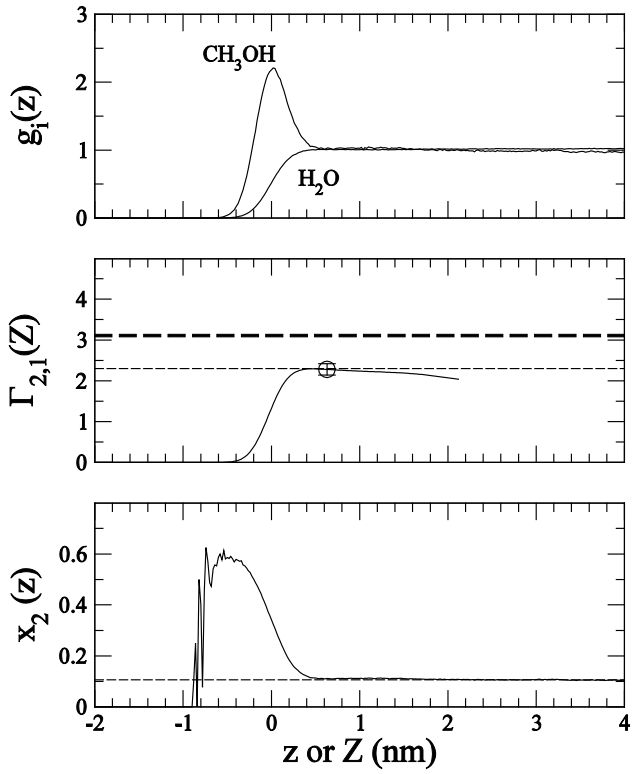


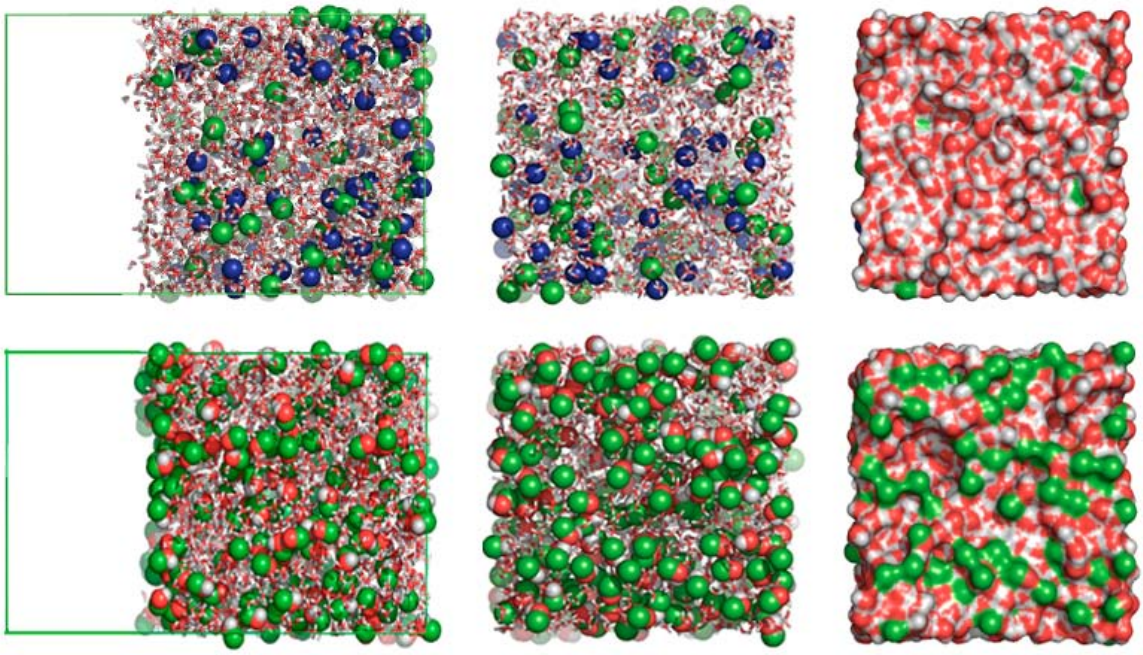
Figure 5.8. Surface properties of an aqueous x_2 = 0.106 methanol solution/vacuum interface. The origin has been shifted so that $g_1(0)$ = 0.5 for convenience. Top: Surface probability distributions (g_i) for water and methanol as a function of distance from the interface (z). Middle: Surface adsorption ($\Gamma_{2,1}$ in molecules/nm^2) of methanol (black line and circle) as a function of integration distance (Z). The thin dashed line is the expected surface adsorption as determined from Equation 5.12 using the simulated values of the surface tension derivative and η. The thick dashed line corresponds to the experimental surface adsorption provided by Equation 5.12. The adsorptions observed at Z = 1 nm were taken as the final simulated values. Bottom: The average methanol mole fraction (x_2) as a function of distance from the interface (z) obtained from the simulations. The dashed line is the average bulk mole fraction (0.106).



The corresponding distributions obtained for one of the methanol compositions (x_2 = 0.106) are presented in Figure 5.8. Here, there is a significant methanol peak situated at the interface, although there is very little structure beyond this immediate interface region. The methanol peak resides at z = 0 or when $g_1(z=0)$ = 0.5, and extends beyond the water interface

into the vacuum region. The corresponding excess surface adsorptions are also shown in Figure 5.8. Again, the distributions approach their bulk values beyond 0.5 nm from the interface. Defining the region of adsorption as that between $g_2(z_2) = 0.5$ and $g_1(z_1=0) = 0.5$ one obtains $z_2 - z_1 = -0.3$ nm for methanol, the same as for NaCl. The simulations suggest that the local mole fraction concentration of methanol at the interface can be significantly higher (x6) than that in the bulk. Snapshots of the NaCl and methanol interfaces are presented in Figure 5.9 for comparison.

Figure 5.9. Snapshots from the MD trajectories illustrating the surface distribution of solutes. Side (left) and interface views (middle) are displayed together with a surface representation of the interface (right). Top: Side and surface views of a 2.2M NaCl solution. Sodium ions (blue) and chloride ions (green) are displayed as spheres with water molecules as sticks. Bottom: Side and surface views of an $x_2 = 0.106$ aqueous methanol solution. Methanol molecules are displayed as spheres with water molecules as sticks. Figures were made with Pymol.



5.6 Discussion

There are several advantages and disadvantages of using the present approach. The advantages are that one has a complete and consistent picture of the simulated thermodynamics for interfacial systems. This provides confidence in the corresponding results if the appropriate properties are well reproduced, or indicators of specific force field deficiencies (surface vs bulk) if they are not. The disadvantages include system size and simulation time. In general, rather long simulation times (> 10ns) are required in order to precisely determine the surface tension changes, and the surface adsorption via Equations 5.5 and 5.6. In addition, somewhat larger system sizes are also required to ensure that one has an accurate description of the bulk solution. Hence, our use of nonpolarizable models in the current study. One observes in Figures 5.7 and 5.8 that the integrated adsorption values do not strictly remain constant beyond the surface region. This behavior has been observed before during our bulk solution simulations and typically disappears as the system size increases.[29]

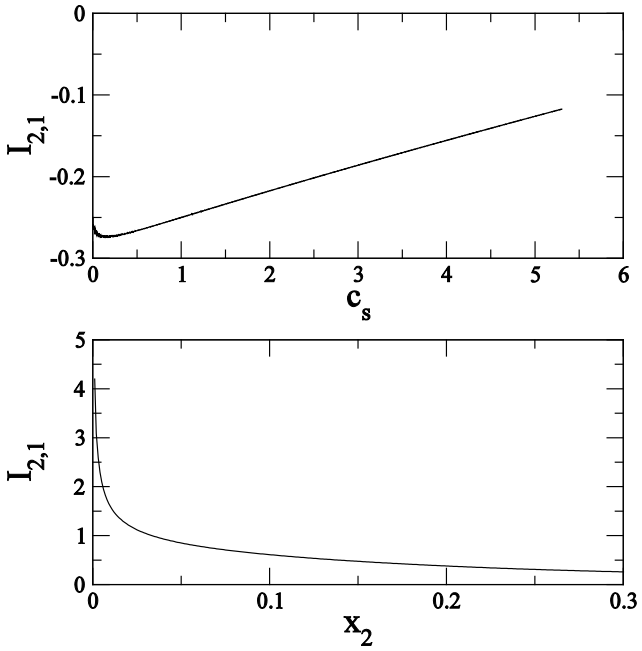
As noted earlier, simulations which provide quantitative agreement with experiment for the surface tensions (derivatives) and the bulk solution values of a_{22} or η, must also then be in quantitatively agreement with the experimental surface adsorption data ($\Gamma_{2,1}$). Unfortunately, the values of $\Gamma_{2,1}$ represent an integral over the surface probability distributions $g_i(z)$ and therefore it is possible that many different surface distributions can produce the same value of the surface adsorption (see Equation 5.6). This is also true for the bulk solution KB integrals G_{ij}. However, in our previous experience with bulk solution simulations this is generally not the case.[40,41,69,70] In fact, it is often difficult to reproduce the density and experimental KB integrals with any parameter set (force field), and we have not observed two different force fields providing the same set of three KB integrals. Hence, we expect the same to be true for the surface probability distributions.

The results displayed here for NaCl provide essentially quantitative agreement with experiment for $\Delta\gamma$, a_{22}, and $\Gamma_{2,1}$. This has been achieved using a force field without explicit polarization effects, although the normal combination rules for Na-O interactions were broken. The current force field was capable of reproducing both bulk solution and interfacial thermodynamic data for a system where changes in polarization would be expected to be significant. While this was not the objective of the present simulations, it does suggest that nonpolarizable force fields can be used with some confidence to study surface distributions as long as a well parametrized force field is used. Other common force fields for NaCl solutions can produce values of a_{22} which are 3-4 times too low, primarily because the G_{22} values are too large and positive.[41] Hence, as Equation 5.6 is exact, they cannot simultaneously reproduce both the change in surface tension and the correct degree of surface adsorption or exclusion. For instance, a previous NaCl force field comparison provides values of 0.34, 0.31, and 0.82 for a_{22} at 2M NaCl.[41] The current force field provides a value of 1.08 compared to the experimental result of 1.15. Hence, even if all these force fields reproduced the change in surface tension reasonably accurately, the resulting surface exclusion of ions obtained from Equation 5.10 would be a factor of 3.38, 3.71, 1.40, and 1.06 too large, respectively. Clearly, the use of some of these force fields would result in non negligible errors. This type of behavior (the degree of exclusion being dependent on a_{22}) has also been observed in our previous studies of cavity formation in mixed solvents.[71]

The NaCl results appear to be consistent with the experimental studies of Raymond and Richmond.[4] In particular, the above study concluded that there were ions at the interface, but not at the surface. The simulated surface probability distributions presented here displayed a simple exclusion of both salt ions with very few features. This contrasts somewhat with previous simulation results which suggest a more structured interface.[8,10] The reasons for this are

undoubtedly due to the different force fields employed in each study. We have confidence in our distributions as we have a force field which reproduces both the changes in surface tensions and the activity derivative (a_{22}) for bulk solutions. Finally, we note that for 1:1 salts one must have $\Gamma_{+,1} = \Gamma_{-,1}$ in order to preserve an electrically neutral interface. However, this does not mean that there is no surface potential.[67] The latter will depend on the asymmetry between the anion and cation distributions.

Figure 5.10. The variation in the surface structure (in units of ions or molecules/nm^2/M) for aqueous NaCl solutions (top) and methanol solutions (bottom) as a function of composition obtained from the experimental data.



In the theory section we investigated some simple cases where the structure of the surface at low solute concentrations, as characterized by $I_{2,1}$, could be related to the fitting parameters describing the changes in surface tension. Changes in the surface structure at larger concentrations determined from the experimental data are displayed in Figure 5.10. If one

ignores the initial cusp region, due to the Debye-Hueckel contribution to the activity coefficient, then the structure of the interface for NaCl solutions varies essentially linearly with solute concentration. The trend is towards a less structured interface as the salt concentration is increased. Hence, while the surface exclusion increases continually with concentration, this is due to an increase in the salt concentration, not an increase in the surface structure. We note that using Equation 5.15 one can write,

$$I_{2,1} = -\frac{\beta\, b}{n_{\pm}} \frac{1}{a_{22}} \tag{5.26}$$

and that an excellent fit to the experimental data between 0.5 and 5.3M is provided by $1/a_{22} = 1.118 - 0.124\ c_s$, i.e. the value of $G_{22} - G_{12}$ is approximately constant (see Equation 5.8). Hence, an observed linear dependence of the above integral on salt concentration. For low concentrations of surface adsorbed solutes Equations 5.18 and 5.20 provide,

$$I_{2,1} \approx \frac{\beta\, \gamma_1}{2.303\rho} \frac{B}{C}\left[1 + x_2(\rho_1 \Delta G_{12} - 1/C)\right] \tag{5.27}$$

as B and C are positive constants, and methanol displays a negative deviation from Raoult's Law, one finds that ΔG_{12} is negative for methanol solutions (see Figure 5.4) and therefore the surface structure decreases with solute concentration. This is the behavior observed in Figure 5.10 where methanol solutions exhibit a sharp decrease in the surface structure even though the surface adsorption is increasing (see Figure 5.6). In summary, the structure of the interface decreases with increasing solute concentration for both NaCl and methanol solutions and therefore the most structured interface is observed at low solute concentrations.

It should be noted that the approach used here does not require any definition of where the surface is, or how diffuse it may be. Hence, it is not necessary to locate the Gibbs dividing surface or its variants. The integration can originate at any z coordinate in the vacuum phase and continues until both surface distribution functions reach their bulk solution values. In practice,

this requires the simulation of reasonably large systems to ensure the bulk phase distribution can be simulated accurately and the values of $\Gamma_{2,1}(Z)$ converge to a constant value. Using Equation 5.22 one can reduce the integration step to a simple counting procedure.

Equation 5.4 can be applied to systems with any number of components and KB theory can be used to provide expressions for $\partial\mu_2/\partial\rho_2$, $\partial\mu_3/\partial\rho_2$, etc in terms of KB integrals.[27,33] This illustrates several further advantages of the current approach. First, while the required chemical potential derivatives are often difficult to obtain experimentally for ternary (and higher) systems, the process is relatively easy for simulated systems as all rdfs and corresponding KB integrals are available. Second, the chemical potential derivatives can be viewed as (composition dependent) scaling factors relating the different relative adsorptions ($\Gamma_{2,1}$, $\Gamma_{3,1}$, etc) to the change in surface tension. If one does not know the values of these derivatives then it is impossible to say which solute contributes the most to the overall surface tension change even when all the $\Gamma_{i,1}$ values are known.

5.7 Conclusions

We have provided an approach for the description of solute effects on the surface tension of solution/vapor interfaces that is suitable for the analysis of computer simulations involving any number of solute components. The approach uses surface probability distributions and the KB theory of solutions to provide a complete and consistent thermodynamic and structural description of the interface. It is illustrated that the relationships derived here are observed in simulations of binary solutions where the aqueous solute is either adsorbed or excluded from the surface. Furthermore, it is strongly suggested that simulations of the corresponding bulk solutions be performed in order to determine the required chemical potential (activity) derivatives, as these can significantly affect the observed degree of surface adsorption or

exclusion, even when the surface tension changes are accurately reproduced. Unfortunately, one cannot assume that current force fields accurately reproduce these derivatives.

Reference List

1. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists;* Plenum Press: New York, **1992**.

2. Chitra, R.; Smith, P. E. Properties of 2,2,2-trifluoroethanol and water mixtures. *Journal of Chemical Physics* **2001,** *114* (1), 426-435.

3. Weerasinghe, S.; Pettitt, M. B. Ideal chemical potential contribution in molecular dynamics simulations of the grand canonical ensemble. *Molecular Physics* **1994,** *82*, 897-912.

4. Ben-Naim, A. Inversion of the Kirkwood -Buff theory of solutions: Application to the water-ethanol system. *Journal of Chemical Physics* **1977,** *67* (11), 4884-4890.

5. Matteoli, E.; Mansoori, G. A. *Fluctuation Theory of Mixtures. Vol 2.;* Taylor & Francis: New York, **1990**.

6. Chitra, R.; Smith, P. E. Preferential interactions of cosolvents with hydrophobic solutes. *Journal of Physical Chemistry B* **2001,** *105*, 11513-11522.

7. Smith, P. E. Cosolvent interactions with biomolecules: Relating computer simulation data to experimental thermodynamic data. *Journal of Physical Chemistry B* **2004,** *108*, 18716-18724.

8. Kirkwood, J. G.; Buff, F. P. The statistical mechnical theory of solutions. I. *Journal of Chemical Physics* **1951,** *19* (6), 774-777.

9. Ben-Naim, A. *Molecular Theory of Solutions;* Oxford University Press: New York, **2006**.

10. Smith, P. E. Chemical potential derivatives and preferential interaction parameters in biological systems from Kirkwood-Buff theory. *Biophysical Journal* **2006,** *91*, 849-856.

11. Ruckenstein, E.; Shulgin, I. Effect of a third component on the interactions in a binary mixture determined from the fluctuation theory of solutions. *Fluid Phase Equilibria* **2001,** *180* (1-2), 281-297.

12. Kirkwood, J. G.; Buff, F. P. The statistical mechnical theory of solutions. I. *Journal of Chemical Physics* **1951,** *19* (6), 774-777.

13. Friedman, H.; Ramanathan, P. S. Theory of mixed electrolyte solutions and application to a model for aqueous lithium chloride- cesium chloride. *Journal of Physical Chemistry* **1970,** *74* (21), 3756-3765.

14. Kusalik, P. G.; Patey, G. N. The thermodynamic properties of electrolyte solutions: Some formal results. *Journal of Chemical Physics* **1987,** *86* (9), 5110-5116.

15. Chitra, R.; Smith, P. E. Molecular association in solution: A Kirkwood-Buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water. *Journal of Physical Chemistry B* **2002,** *106* (6), 1491-1500.

16. Weissenborn, P. K.; Pugh, R. J. Surface-Tension and Bubble Coalescence Phenomena of Aqueous-Solutions of Electrolytes. *Langmuir* **1995,** *11* (5), 1422-1426.

17. Chitra, R.; Smith, P. E. A comparison of the properties of 2,2,2-trifluoroethanol and 2,2,2-trifluoroethanol/water mixtures using different force fields. *Journal of Chemical Physics* **2001,** *115* (12), 5521-5530.

18. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of urea and water. *Journal of Physical Chemistry B* **2003,** *107*, 3891-3898.

19. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for sodium chloride in water. *Journal of Chemical Physics* **2003,** *119*, 11342-12349.

20. Perera, A.; Sokolic, F. Modeling nonionic aqueous solutions: The acetone-water mixture. *Journal of Chemical Physics* **2004,** *121* (22), 11272-11282.

21. Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. Spontaneous formation of KCl aggregates in biomolecular simulations: A force field issue? *Journal of Chemical Theory and Computation* **2007,** *3* (5), 1851-1859.

22. Parsegian, V. A.; Rand, R. P.; Rau, D. C. Osmotic stress, crowding, preferential hydration, and binding: A comparison of perspectives. *Proceedings of the National Academy of Sciences of the United States of America* **2000,** *97*, 3987-3992.

23. Baynes, B. M.; Trout, B. L. Proteins in Mixed Solvents: A Molecular-Level Perspective. *Journal of Physical Chemistry B* **2003,** *107*, 14058-14067.

24. Weissenborn, P. K.; Pugh, R. J. Surface-Tension and Bubble Coalescence Phenomena of Aqueous-Solutions of Electrolytes. *Langmuir* **1995,** *11* (5), 1422-1426.

25. Meissner, H. P.; Michaels, A. S. Surface Tensions of Pure Liquids and Liquid Mixtures. *Industrial and Engineering Chemistry* **1949,** *41* (12), 2782-2787.

26. Poling, B. E.; Prausnitz, J. M.; O'Connell, J. P. *The properties of gases and liquids;* 5th ed.; McGraw-Hill: New York, 2001.

27. Strey, R.; Viisanen, Y.; Aratono, M.; Kratohvil, J. P.; Yin, Q.; Friberg, S. E. On the necessity of using activities in the Gibbs equation. *Journal of Physical Chemistry B* **1999,** *103* (43), 9112-9116.

28. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

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

30. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. Gromacs - A Message-Passing Parallel Molecular-Dynamics Implementation. *Computer Physics Communications* **1995,** *91* (1-3), 43-56.

31. Lindahl, E.; Hess, B.; van der Spoel, D. GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling* **2001,** *7* (8), 306-317.

32. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.

33. Miyamoto, S.; Kollman, P. A. Settle - An Analytical Version of the Shake and Rattle Algorithm for Rigid Water Models. *Journal of Computational Chemistry* **1992,** *13* (8), 952-962.

34. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **1995,** *103* (19), 8577-8593.

35. Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *Journal of Chemical Physics* **1984,** *81*, 3684-3690.

36. Harris, J. G. Liquid Vapor Interfaces of Alkane Oligomers - Structure and Thermodynamics from Molecular-Dynamics Simulations of Chemically Realistic Models. *Journal of Physical Chemistry* **1992,** *96* (12), 5077-5086.

37. Alejandre, J.; Tildesley, D. J.; Chapela, G. A. Molecular-Dynamics Simulation of the Orthobaric Densities and Surface-Tension of Water. *Journal of Chemical Physics* **1995,** *102* (11), 4574-4583.

38. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

39. Kang, M.; Smith, P. E. Preferential interaction parameters in biological systems by Kirkwood-Buff theory and computer simulation. *Fluid Phase Equilibria* **2007,** *256* (1-2), 14-19.

40. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

41. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

42. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for methanol and aqueous methanol solutions. *Journal of Physical Chemistry B* **2005,** *109* (31), 15080-15086.

43. Weast, R. C. *Handbook of Chemistry and Physics*, CRC Press: Boca Raton, Florida, 1985.

44. Alejandre, J.; Tildesley, D. J.; Chapela, G. A. Molecular-Dynamics Simulation of the Orthobaric Densities and Surface-Tension of Water. *Journal of Chemical Physics* **1995,** *102* (11), 4574-4583.

45. Ismail, A. E.; Grest, G. S.; Stevens, M. J. Capillary waves at the liquid-vapor interface and the surface tension of water. *Journal of Chemical Physics* **2006,** *125* (1), 014702.

46. Shi, B.; Sinha, S.; Dhir, V. K. Molecular dynamics simulation of the density and surface tension of water by particle-particle particle-mesh method. *Journal of Chemical Physics* **2006,** *124* (20), 204715.

47. Lu, Y. J.; Wei, B. Second inflection point of water surface tension. *Applied Physics Letters* **2006,** *89* (16), 164106.

48. Wemhoff, A. P.; Carey, V. P. Surface tension prediction using characteristics of the density profile through the interfacial region. *International Journal of Thermophysics* **2006,** *27* (2), 413-436.

49. Wynveen, A.; Bresme, F. Interactions of polarizable media in water: A molecular dynamics approach. *Journal of Chemical Physics* **2006,** *124* (10), 104502.

50. Chen, F.; Smith, P. E. Simulated surface tensions of common water models. *Journal of Chemical Physics* **2007,** *126* (22), 221101.

51. in't Veld, P. J.; Ismail, A. E.; Grest, G. S. Application of Ewald summations to long-range dispersion forces. *Journal of Chemical Physics* **2007,** *127*, 144711.

52. Chitra, R.; Smith, P. E. Molecular association in solution: A Kirkwood-Buff analysis of sodium chloride, ammonium sulfate, guanidinium chloride, urea, and 2,2,2-trifluoroethanol in water. *Journal of Physical Chemistry B* **2002,** *106* (6), 1491-1500.

53. Weerasinghe, S.; Smith, P. E. A Kirkwood-Buff derived force field for mixtures of acetone and water. *Journal of Chemical Physics* **2003,** *118*, 10663-10670.

54. Kang, M.; Smith, P. E. A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* **2006,** *27* (13), 1477-1485.

55. Weerasinghe, S.; Smith, P. E. Cavity formation and preferential interactions in urea solutions: Dependence on urea aggregation. *Journal of Chemical Physics* **2003,** *118* (13), 5901-5910.

56. Raymond, E. A.; Richmond, G. L. Probing the molecular structure and bonding of the surface of aqueous salt solutions. *Journal of Physical Chemistry B* **2004,** *108* (16), 5051-5059.

57. Bhatt, D.; Newman, J.; Radke, C. J. Molecular dynamics simulations of surface tensions of aqueous electrolytic solutions. *Journal of Physical Chemistry B* **2004,** *108* (26), 9077-9084.

58. Jungwirth, P.; Tobias, D. J. Ions at the air/water interface. *Journal of Physical Chemistry B* **2002,** *106* (25), 6361-6373.

59. Jarvis, N. L.; Scheiman, M. A. Surface Potentials of Aqueous Electrolyte Solutions. *Journal of Physical Chemistry* **1968,** *72* (1), 74-78.

60. Wilson, M. A.; Pohorille, A. Interaction of Monovalent Ions with the Water Liquid Vapor Interface - A Molecular-Dynamics Study. *Journal of Chemical Physics* **1991,** *95* (8), 6005-6013.

CHAPTER 6 - Summary and Future Work

As the power of available computational resources keeps growing, classical based molecular dynamics simulation studies of interested system will continue to play an important role as researchers develop improved models and methodology. The rate of improvement is remarkable. Datasets that were used to be collected in several months may now be collected in a week, which makes it possible for us to study system with bigger size with longer MD runs. Kirkwood-Buff (KB) theory has been proved its ability to provide a direct linkage between the probability distribution of species and the thermodynamics quantities in solution mixtures. The preceding chapters have discussed our attempts to apply molecular dynamics ensembles combined with KB theory to investigate properties of bulk solution, liquid/air interface and protein backbone structures. Results have shown the promising future of our newly developed KB derived force field. It is obvious that our force field together with KB theory can provide detail insights of variety problems in today's research work. On their way towards to high quality accurate KB derived force field, smith and his colleagues have shown their courage and confidence to be able to accomplish high level academic demands. In the future, we will continue work on improving the quality of our current parameters set, and hopefully our work will eventually lead us to clearer view on future tasks, such as protein folding, ligand binding, and drug design.

Appendix A - Copy of Permission Letter from the Publisher

AMERICAN INSTITUTE OF PHYSICS LICENSE TERMS AND CONDITIONS

Nov 19, 2009

This is a License Agreement between feng chen ("You") and American Institute of Physics ("American Institute of Physics") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by American Institute of Physics, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2312570648087
License date	Nov 19, 2009
Licensed content publisher	American Institute of Physics
Licensed content publication	Journal of Chemical Physics
Licensed content title	Simulated surface tensions of common water models
Licensed content author	Feng Chen, Paul E. Smith
Licensed content date	Jun 14, 2007
Volume number	126
Issue number	22
Type of Use	Thesis/Dissertation
Requestor type	Author (original article)
Format	Print and electronic
Portion	Excerpt (> 800 words)
Will you be translating?	No
Title of your thesis / dissertation	MOLECULAR DYNAMICS SIMULATIONS AND THEORY OF INTERMOLECULAR
Expected completion date	Dec 2009
Estimated size (number of pages)	110
Total	0.00 USD

Terms and Conditions

American Institute of Physics -- Terms and Conditions: Permissions Uses

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

1. You agree to include the following copyright and permission notice with the reproduction of the Material: "Reprinted with permission from [FULL CITATION]. Copyright [PUBLICATION YEAR], American Institute of Physics." For an article, the copyright and permission notice must be printed on the first page of the article or book chapter. For photographs, covers, or tables, the copyright and permission notice may appear with the Material, in a footnote, or in the reference list.
2. If you have licensed reuse of a figure, photograph, cover, or table, it is your

responsibility to ensure that the material is original to AIP and does not contain the copyright of another entity, and that the copyright notice of the figure, photograph, cover, or table does not indicate that it was reprinted by AIP, with permission, from another source. Under no circumstances does AIP, purport or intend to grant permission to reuse material to which it does not hold copyright.

3. You may not alter or modify the Material in any manner. You may translate the Material into another language only if you have licensed translation rights. You may not use the Material for promotional purposes. AIP reserves all rights not specifically granted herein.
4. The foregoing license shall not take effect unless and until AIP or its agent, Copyright Clearance Center, receives the Payment in accordance with Copyright Clearance Center Billing and Payment Terms and Conditions, which are incorporated herein by reference.
5. AIP or the Copyright Clearance Center may, within two business days of granting this license, revoke the license for any reason whatsoever, with a full refund payable to you. Should you violate the terms of this license at any time, AIP, American Institute of Physics, or Copyright Clearance Center may revoke the license with no refund to you. Notice of such revocation will be made using the contact information provided by you. Failure to receive such notice will not nullify the revocation.
6. AIP makes no representations or warranties with respect to the Material. You agree to indemnify and hold harmless AIP, American Institute of Physics, and their officers, directors, employees or agents from and against any and all claims arising out of your use of the Material other than as specifically authorized herein.
7. The permission granted herein is personal to you and is not transferable or assignable without the prior written permission of AIP. This license may not be amended except in a writing signed by the party to be charged.
8. If purchase orders, acknowledgments or check endorsements are issued on any forms containing terms and conditions which are inconsistent with these provisions, such inconsistent terms and conditions shall be of no force and effect. This document, including the CCC Billing and Payment Terms and Conditions, shall be the entire agreement between the parties relating to the subject matter hereof.

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

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a check or money order referencing your account number and this license number 2312570648087.
If you would prefer to pay for this license by credit card, please go to http://www.copyright.com/creditcard to download our credit card payment authorization form.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2312570648087 and noting the reason for cancellation.