COMPUTER SIMULATION AND THEORY OF AMINO ACID INTERACTIONS IN SOLUTION

By

MOON BAE GEE

B.A., Chungnam National University, Korea, 2000
M.S., Chungnam National University, Korea, 2002

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

The force fields used in computer simulations play an important role in describing a particular system. In order to estimate the accuracy of a force field, physical or thermodynamic properties are usually compared with simulation results. Recently, we have been developing a force field which is called the Kirkwood-Buff Force Field (KBFF). This force field is established by transforming experimental data into Kirkwood-Buff (KB) integrals and then attempting to reproduce those KBIs with molecular dynamic (MD) simulations. Here we investigate a variety of intermolecular interactions in aqueous solutions through KB theory and molecular simulations. First, we describe a force field for the simulation of alkali halide aqueous solutions. These models are developed specifically to reproduce the experimentally determined Kirkwood-Buff integrals and solution activities as a function of molality. Additionally, other experimentally known properties including ion diffusion constants, relative permittivities, the densities and heats of mixing are reproduced by these models. Second, In an effort to understand the interactions which occur between amino acids in solution we have developed new force fields for simple amino acids and their analogs including glycine, betaine, *β*-alanine, *dl*-alanine, NH_4Cl, NH_4Br, $N(CH_3)_4Cl$, $N(CH_3)_4Br$, CH_3NH_3Cl, and CH_3COONa. The new force fields reproduce the experimental Kirkwood-Buff integrals which describe the relative distribution of all the species in a solution mixture. Furthermore, it is shown that these simple amino acids can be understood in terms of the interactions of their functional groups and that, to a very good approximation, the transferability and additivity usually assumed in the development of biomolecular force fields appear to hold true. Third, an analysis of the effect of a cosolvent on the association of a solute in solution is presented by using the Kirkwood-Buff theory of solutions. The derived expressions provide a foundation for the investigation of cosolvent effects on molecular and biomolecular

equilibria, including protein association, aggregation, and cellular crowding. Finally, in an effort to understand peptide aggregation at the atomic level we have performed simulations of polyglycine ($(gly)_n$) using our recently developed force fields. Experimentally, the association of glycine polypeptides increases with n. Our force fields reproduce this behavior, and we investigated the reasons behind this trend. In addition to studying closed ensembles, we also simulate these systems in a semi-open ensemble that was designed to mimic cellular environments typically open to water, using a simple direct approach. The differences between the two ensembles are investigated and compared with our recent theoretical descriptions of aggregating systems using Kirkwood-Buff theory.

COMPUTER SIMULATION AND THEORY OF AMINO ACID INTERACTIONS IN SOLUTION

by

MOON BAE GEE

B.A., Chungnam National University, Korea, 2000
M.S., Chungnam National University, Korea, 2002

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
Paul E. Smith

Abstract

The force fields used in computer simulations play an important role in describing a particular system. In order to estimate the accuracy of a force field, physical or thermodynamic properties are usually compared with simulation results. Recently, we have been developing a force field which is called the Kirkwood-Buff Force Field (KBFF). This force field is established by transforming experimental data into Kirkwood-Buff (KB) integrals and then attempting to reproduce those KBIs with molecular dynamic (MD) simulations. Here we investigate a variety of intermolecular interactions in aqueous solutions through KB theory and molecular simulations. First, we describe a force field for the simulation of alkali halide aqueous solutions. These models are developed specifically to reproduce the experimentally determined Kirkwood-Buff integrals and solution activities as a function of molality. Additionally, other experimentally known properties including ion diffusion constants, relative permittivities, the densities and heats of mixing are reproduced by these models. Second, In an effort to understand the interactions which occur between amino acids in solution we have developed new force fields for simple amino acids and their analogs including glycine, betaine, *β*-alanine, *dl*-alanine, NH_4Cl, NH_4Br, $N(CH_3)_4Cl$, $N(CH_3)_4Br$, CH_3NH_3Cl, and CH_3COONa. The new force fields reproduce the experimental Kirkwood-Buff integrals which describe the relative distribution of all the species in a solution mixture. Furthermore, it is shown that these simple amino acids can be understood in terms of the interactions of their functional groups and that, to a very good approximation, the transferability and additivity usually assumed in the development of biomolecular force fields appear to hold true. Third, an analysis of the effect of a cosolvent on the association of a solute in solution is presented by using the Kirkwood-Buff theory of solutions. The derived expressions provide a foundation for the investigation of cosolvent effects on molecular and biomolecular

equilibria, including protein association, aggregation, and cellular crowding. Finally, in an effort to understand peptide aggregation at the atomic level we have performed simulations of polyglycine ($(gly)_n$) using our recently developed force fields. Experimentally, the association of glycine polypeptides increases with n. Our force fields reproduce this behavior, and we investigated the reasons behind this trend. In addition to studying closed ensembles, we also simulate these systems in a semi-open ensemble that was designed to mimic cellular environments typically open to water, using a simple direct approach. The differences between the two ensembles are investigated and compared with our recent theoretical descriptions of aggregating systems using Kirkwood-Buff theory.

Table of Contents

List of Figures

List of Tables

Acknowledgements

My deepest gratitude is to my advisor, Dr. Paul E. Smith. I have been amazingly fortunate to have an advisor who gave me the freedom to explore on my own. Paul taught me how to question thoughts and express ideas. His patience and support helped me overcome many crisis situations and finish this dissertation. Without his guidance and persistent help this dissertation would not have been possible.

I am also grateful to all of my Ph.D committee members, Dr. Christine Aikens, Dr. Daniel A. Higgins, and Dr. Ramaswamy Krishnamoorthi for their valuable time and efforts, commenting on my views and helping me understand and enrich my ideas

I am also thankful to members of Dr. Smith's group present and departed, who have supported me through years of graduate study. Their support and care helped me overcome setbacks and stay focused on my graduate study. I greatly value their friendship and I deeply appreciate their belief in me.

Most importantly, none of this would have been possible without the love and patience of my family. My family, to whom this dissertation is dedicated to, has been a constant source of love, concern, support and strength all these years. I would like to express my heart-felt gratitude to my family.

Thank you!

Dedication

To my parents

CHAPTER 1 - Introduction

General Introduction

Aqueous solution of proteins and salts are some of the most interesting systems in cell biology because salts can affect the thermodynamics and structure of proteins. Small perturbations in protein structure may expose hydrophobic surfaces that lead to aggregation which can influence the normal functions of proteins in our bodies. Therefore, many scientists have tried to understand the roles of salts in aqueous solutions. For example, it has recently been revealed that many diseases including Alzheimer's, Huntington's, Amyotrophic Lateral Sclerosis (ALS), and prion diseases,[1-2] are related to protein aggregation which is affected by the salt concentration or type, temperature, and pH of the cellular environment. Unfortunately, it is difficult to describe diseases at the atomic level using typical experimental procedures. We need alternative methods to understand the altered physicochemical properties of proteins which lead to disease at atomic detail. Molecular dynamics (MD) simulation is one of the great tools which can be used to investigate these diseases because it provides information on the physical properties and the dynamics of aqueous solutions including proteins and salts. In order to develop a bridge between the model system used in MD simulations and the real system, a proper theory needs to be employed. Otherwise, the results from MD simulations are meaningless. For example, it has been questioned if the mainstream force fields used for MD simulation can correctly describe the properties of solution mixtures.[3-5] Here, the first goal of this project is to develop a correct force field for systems including a combination of protein, salts, and water molecules. The second is to extend the theory used in the force field development (Kirkwood-Buff theory) to describe interesting phenomena of aggregating peptide systems in

both closed and semi-open ensembles. Finally, we will present a new model of peptide aggregation in both closed and semi-open systems using our new force fields.

Molecular Simulation

Computer simulations are increasingly playing a more important role in the investigation of complicated biological systems. The development of algorithms and theories, combined with the increase in computing power, has made it possible to investigate the properties of complex biomolecular systems at different levels of resolution. The resolution should be chosen based upon the type of property or process under investigation. Computational techniques are the best way to investigate complicated biomolecular system at this stage even though recent progress in experimental analysis has also been developed.[6-9]

There are two types of computer simulations based on statistical mechanics. In Monte Carlo (MC) simulation, new coordinates of the particles in the box are generated by small random moves, and then the change in total potential energy of the system is calculated.[10] The acceptance of a new configuration is dependent on the Boltzmann distribution. MC has been used for the investigation of peptide oligomerization and an advantage of this method is the ability to mimic the actual assembly of several peptide chains. A disadvantage is the limited sampling.[11]

Molecular dynamics (MD) is another computer simulation technique based on statistical mechanics, in which the net force arising from all atoms and molecules in a system is allowed to move the molecules by Newton's equation of motion via interactions following empirical pair-additive potentials, for a period of time. This type of simulation is frequently used for the investigation of proteins, biomolecules, and materials science, because it provides dynamical

properties of the system through statistical ensemble averages which are equal to time averages of the system. In Figure 1.1, the relationship between macroscopic and microscopic system properties is described.

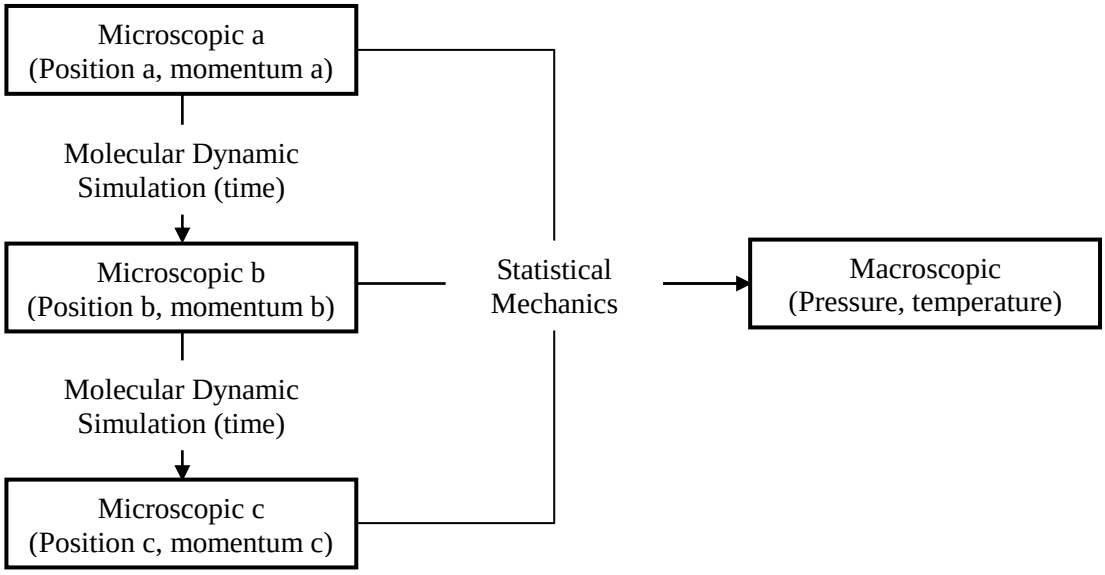


Figure 1.1 The relationship between macroscopic and microscopic system properties. The macroscopic properties can be determined by statistical mechanics calculations from the microscopic ensemble averages.

MD simulation is useful to study protein dynamics and to address peptide self-assembly. In the case of all atom approaches, each atom in the system, including the solvent, is described explicitly. Hence, MD simulation affords a high degree of resolution for complicated environments. Most biomolecular simulations involve non-polarizable force fields, where the atoms in a protein are modeled as rigid spheres with a fixed charge that interact through a series of bonded and non-bonded potentials. The positions and velocities of the atoms are then obtained by numerical integration of the classical equations of motion, and a direct link between the atomic descriptions and macroscopic properties is established using statistical mechanics.

When a molecular dynamics simulation is designed, the computational costs should be considered. The proper time steps, box size for the simulations, and number of particles in the box should be selected to provide reasonable computational times, while the simulation time should be long enough to sample the time scales of the interesting natural processes under investigation. Otherwise, it is difficult to obtain statistically valid results from simulation. These days, in order to obtain the lengths of time required for statistically meaningful simulation results, people are using parallel algorithms which make use of multiple CPUs.[12]

Force Field Development

The quality of a molecular dynamics simulation is determined by the force field used in the simulation. Each particle in the system behaves as described by the input parameters of a force field, thus it is critical to use accurate simulation parameters for the atomic properties and their interactions in order to obtain a correct description of the system.

Force Field

A force field involves the functional form and parameter sets used to describe the potential energy for a particular system. The basic functional form of a force field includes bonded terms and nonbonded terms. The bonded terms are composed of bond, angle, and dihedral angle terms. Nonbonded terms describe electrostatic and van der Waals forces. The total energy is given by:

$$E_{total} = E_{bond} + E_{angle} + E_{proper} + E_{improper} + E_{electrostatic} + E_{vanderWaals.}$$

These functions and parameter sets can be obtained from both experimental work and quantum calculations. The bonded terms including bond, angle, improper, and proper dihedral functions are well established.[13] The nonbonded terms are the most computationally intensive

because they include many more interactions per atom. A popular way to minimize the computational cost is to limit interactions to pairwise energies. The van der Waals terms usually follow a Lennard-Jones potential and the electrostatic term is described by Coulomb's law.

There are two different types of force fields: all atom, which explicitly describe every atom in a system, and united atom, which treat the hydrogen and carbon atoms in methyl groups as a single atom. Coarse-grained force fields are a subset of the united atom force fields. They are useful for long-time simulations of proteins because they reduce computational cost. However, the atomic details are neglected.[14-16]

The Strategy for the Force Field Development

The force fields used to describe aqueous solution containing proteins and salts have significant flaws.[5] Hence, we need a new strategy for force field development which will allow us to describe those systems with greater accuracy. Relying on the principle of additivity, which says that a protein is the sum of its parts, we take the following approach. First, we separate the whole system in to a series of salts in water and small peptides in water. Second, small peptides are broken down into amino acids. Third, amino acids are broken down into their building blocks which are studied in detail to develop the force field parameters.

Figure 1.2 illustrates the strategy for the study of systems including proteins, salts, and water molecules. As a model for salts and water, alkali halide aqueous solutions are not only one of the simplest model systems, but these ions also play an important role in many biological systems e.g., by influencing the structure and dynamics of biomolecules. They stabilize important biomolecules like proteins, nucleic acids, and lipids, and they are important for biological catalysis.[17-19] In the 1880s, Hofmeister and co-workers investigated the relative ability of

different salts to precipitate proteins.[19] His work established the ranking for an anion or cation's ability to precipitate a protein. It is widely held that the Hofmeister series reflects specific ion effects on the long-range structure of water.

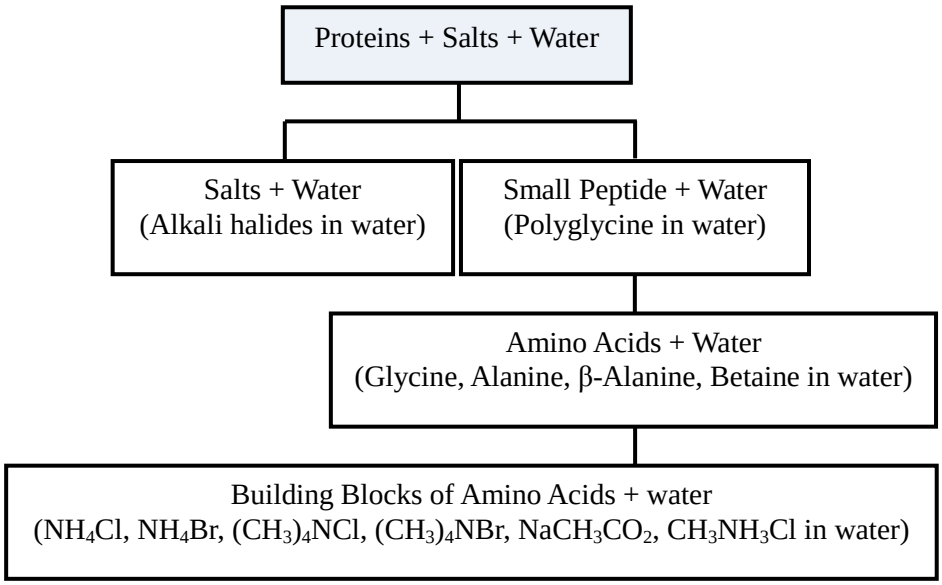


Figure 1.2 The strategy for the force field development of aqueous solutions including proteins and salts in biological environments.

As a model of a small peptide in water, we have chosen the polyglycine and water system because glycine is the simplest amino acid. Before the investigation of polyglycines, several amino acid systems: glycine, *dl*-alanine, *β*-alanine, betaine in water were also studied. The propensity for different amino acid sequences to induce protein aggregation could then be investigated. Different short peptide sequences can self-assemble into different structures of nanoscale dimensions.[20]

Furthermore, the studies of components such as amine groups, carboxylic acid groups, and side chains, which are the building blocks of amino acids, are particularly important in the study

of proteins because the chemical properties of the amino acids determine the biological activity of the protein. The various amino acids differ in which side chain is attached to their α-carbon. Therefore, before we can understand the structure and properties of amino acids we need to investigate the components of amino acids first. In this project, the building blocks of those amino acids namely: NH_4Cl, NH_4Br, $(CH_3)_4NCl$, $(CH_3)_4NBr$, $NaCH_3CO_2$, CH_3NH_3Cl in water have been studied.

Polarizability in Force Fields

Polarization is the changes in a charge distribution around a molecule due to a change in its environment. Real systems are polarized when placed in a high-dielectric medium such as water. This polarization strongly affects the geometry and energy of molecules in the system. Unfortunately, simulations performed using fixed charges cannot describe polarization or charge-transfer effects, which may be critical for systems of biopolymers such as protein and nucleic acid molecules. In order to overcome limitations with fixed charge models, some models that include explicit polarization or charge transfer effects have been proposed.[21-24]

To introduce explicit polarization is one of the ways to improve force fields. This method typically includes the use of induced dipoles, and thereby the effects of changes in environment such as the electronic structure of ions and molecules. The advantages of polarizable force fields in molecular dynamics simulations are an increase in accuracy due to the inclusion of electronic structure changes. However, the disadvantages of polarizability calculations are they are the computationally expensive compared to nonpolarizable fixed charge methods. For example, complicated biological systems including a large number of molecules will require significant computation time.

Kirkwood-Buff Theory

The Kirkwood-Buff (KB) theory of solutions was published in 1951. In the original paper, Kirkwood and Buff derived new relationships between thermodynamic quantities and molecular distribution functions for multi-component systems in the μVT ensemble.[25] KB theory can provide general expressions for any type of particle over the entire range of compositions. Ben-Naim developed the useful inversion procedure which makes this theory applied to solutions.[26] His approach provides information concerning the interaction between a pair of species in any solution mixture. Since then, many chemists and physicists have followed his lead to develop KB theory and apply it for the study of various solution mixtures. KB theory has been applied to determine molecular osmolyte effects on macromolecules,[27] to understand structural thermodynamics of protein preferential solvation,[28-29] and to analyze the free energy of molecular binding into lipid membranes.[30]

The radial distribution function (rdf) is an important pair correlation function. It describes the density of species as a function of the distance from one particular central species. This is useful to describe the particle distribution in liquid systems. For example, it can be derived from experimental data such as x-ray or neutron diffraction studies,[31-32] thus providing a direct comparison between experiment and simulation. The physical meaning of radial distribution function (rdf) is described in Figure 1.3.

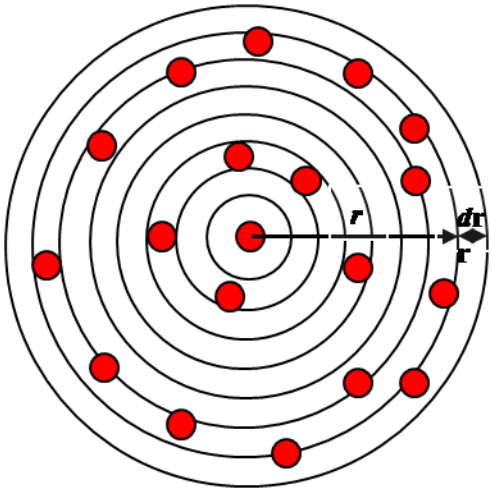


Figure 1.3 The physical meaning of the radial distribution function. The Red balls are particles. *r* indicates the distance from the central particle. *dr* indicates the distance between shells.

The rdf is usually plotted as a function of the distance r between *i* and *j*. The probability function as a function or r is obtained by dividing the average number of atoms in each shell by the volume of the shell, and then normalizing by reference to the bulk density. The rdf indicates the local solution structure, including solvation shells. As shown in Figure 1.4, at short distances less than the effective width of *i* and *j* the rdf has zero value due to a strong repulsive force. Several obvious peaks indicate that the neighbor atoms pack around each other. As the distance between species *i* and *j*, r_{ij}, increases, the rdf goes to unity, indicating the distribution becomes similar to the bulk distribution. As temperature increases the peaks become broader because of thermal motion. In the case of crystals, the peaks are very sharp because atoms are strongly bounded in their given positions.

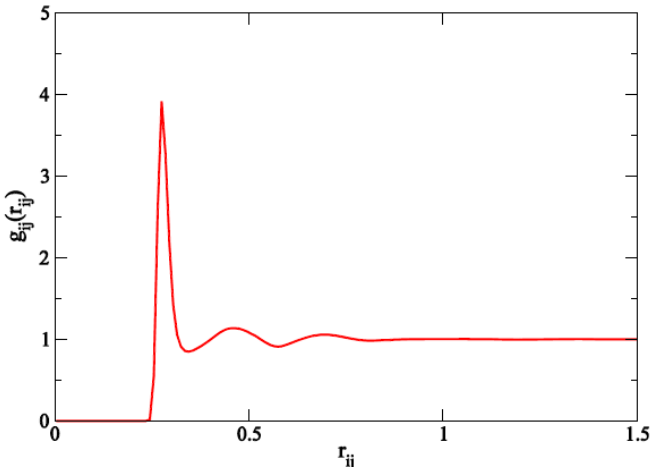


Figure 1.4 Radial distribution function (rdf) is presented.

This KB integral corresponding to the rdf can be obtained either by experimental or simulated data. For a binary solution, three experimental KBIs can be extracted from solution data consisting of chemical potentials, partial molar volumes, and isothermal compressibilities.[26] Simulation data also provide three rdfs which can be integrated to yield the three KBIs from a binary solution mixture including water and cosolvent at constant pressure and temperature. In Figure 1.5 we show an example of a KB integral G_{ij} as a function of integration distance r (nm) between species i and j, and corresponding to the rdf in Figure 1.4. The KBIs are obtained from[26]

$$G_{ij} = 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1 \right] r^2 dr \,. \tag{1.1}$$

Here, G_{ij} is the KB integral between species i and j, g_{ij} is the corresponding radial distribution function (rdf) in the NpT ensemble, and r is the distance between the center of mass (i) and the center of mass (j).

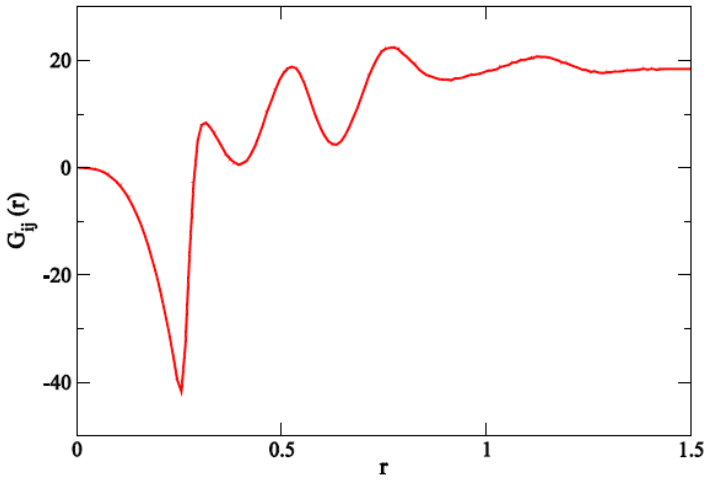


Figure 1.5 An example of KB integral G_{ij} as a function of integration distance R (nm) between species i and j. This KB integral corresponds to the rdf displayed in Figure 1.4.

The thermodynamic properties of a solution mixture can be expressed using the KB integrals between the different solution components. Therefore, KB theory is a nice tool to bridge between an interesting real system, and a model system for computer simulation as shown in Figure 1.6. Furthermore, KB theory has been applied to a number of biomolecular systems, as well as a variety of cosolvent systems.[33-40]

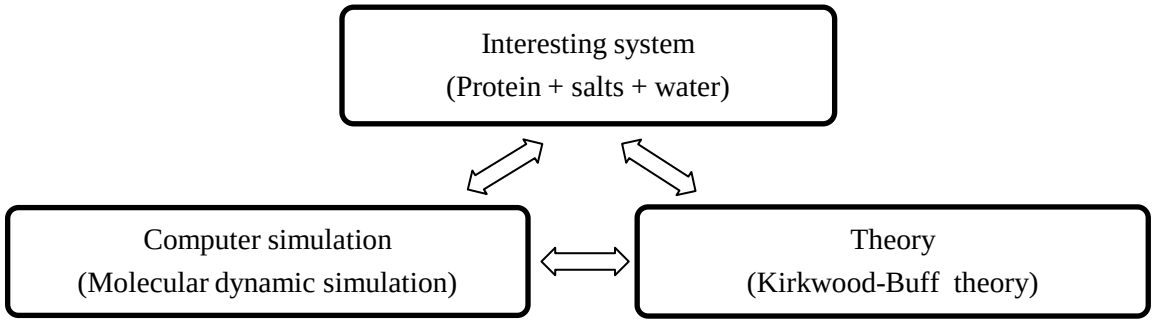


Figure 1.6 The relationship among system, method, and theory used here.

Kirkwood-Buff Derived Force Field

It has been shown that experimental KB integrals can play an important role in the parameterization of a new force field because KBIs are more sensitive to the parameter sets than most other experimental data.[38-43] In addition, they can help us quantify the interaction between a pair of components in solutions. The quality of a force field used in a simulation can be determined by comparing the KBIs and thermodynamic properties obtained from simulations to the KBIs and thermodynamic properties obtained from experimental data. Early tests using commonly available force fields indicated problems[44-46] trying to reproduce KB integrals suggesting that an improved force field is needed. Consequently, a new force field which is specially designed to reproduce experimental KBIs has been developed to describe the delicate balance between solute-solute interactions and solute-solvent interactions in solution.

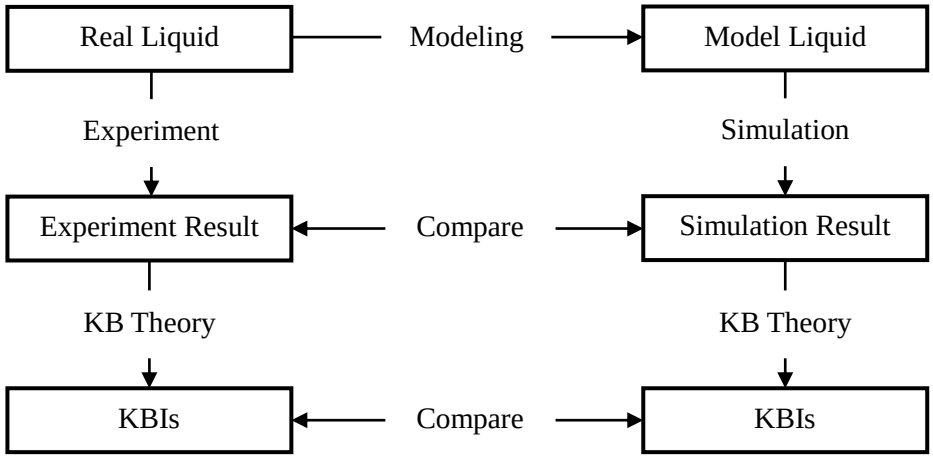


Figure 1.7 The connection between experiment and computer simulation. The experimental results can be compared with the simulation results directly and the experimental KBIs can be compared with the simulated KBIs using KB theory.

It is well known that KB integrals are very sensitive to the input parameters of the force field, in particular to the charge distributions.[37-39, 41] Hence, in the development of a new force field using KB theory, the charge distribution has been focused, while bonded parameters are typically introduced from existing experimental data. The non-bonded van der Waals interactions are also well established. In order to improve the charge distributions on atoms, the KB approach adjusts the charges on the atoms to reproduce the density and KB integrals for solution mixtures at several different compositions.[35-39, 47] Other approaches typically determine the partial charges on atoms by using ab initio calculations of gas phase, which is then followed by scaling in an effort to mimic polarization effect in water. Even though a Kirkwood-Buff force field is not a polarizable force field, accurate effective charges on atoms can be determined due to the sensitivity of the KBIs. It has been shown that the KBFF performs simulation better than many other non-polarizable force fields with the same computational cost.[35-40, 47]

Protein Aggregation

Protein aggregation is not only related to undesired human diseases like Alzheimer's disease, Creutzfeld-Jakob disease, Gerstmann-Straussler syndrome, Huntington's disease, Parkinson's disease, Amyotrophic lateral sclerosis (ALS), and type 2 diabete.[1-2] Figure 1.8 presents the several pathways for protein aggregation. The native state is in equilibration with the nonnative state, but the aggregation steps are usually irreversible reactions.

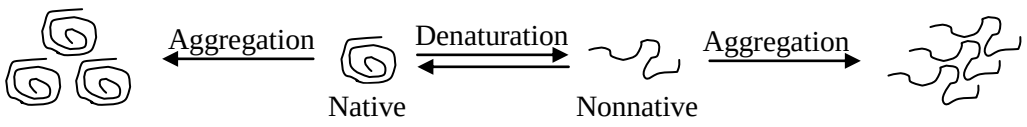


Figure 1.8 General schematic presentation of overall pathway for protein aggregation

Usually, the denaturation of a protein is followed by protein aggregation because of exposure of hydrophobic surfaces which induces favorable protein-protein interactions in aqueous solution. The behavior of many protein aggregates have been found to depend on the properties of the solution environment i.e., temperature, pH, cosolvents, the protein sequences,[20, 48-52] and the relative thermodynamic stability of its native state.[53-56] Recently, the interactions of native protein self-association are getting more consideration since even small changes in the normal interactions between proteins can lead to human diseases.[57-58]

As an initial step towards the study of protein self-associations, small peptides provide excellent models because small peptides are not only easy to be synthesized and modified experimentally, but also require reduced computational costs compared to proteins simulations. In addition, it has been found that short peptides such as pentapeptides or tetrapeptides can form typical fibrils and different short peptide sequences can self-assemble into different structures of nanoscale dimensions.[59] Even dipeptides can form well-ordered assemblies.[60] The investigation of the physical and chemical driving forces of peptide self-assembly is a fundamental step in order to understand protein aggregation.

In order to understand the driving force of peptide aggregation, the intermolecular interactions between a protein and its surroundings i.e., protein-protein interactions, protein-solvent interactions, and protein-cosolvent interactions need to be investigated. Hence, in order to understand the thermodynamics, we must investigate the underlying atomic interactions. The study of cosolvents in solution is especially helpful in understanding peptide association.[61]

Cosolvent Effects on the Stability of Proteins

In order to describe the basic thermodynamic effects of cosolvents on chemical equilibrium, the concept of binding and linkage has been used.[62-63] Timasheff has applied this linkage function and related theories to the stability of protein conformations, which can then be used to describe cosolvent effects.[61] According to the Wyman linkage function, a greater binding to the native state will shift the equilibrium toward the native state.[61] It can be used to describe how weakly interacting cosolvents affect the stability of protein conformation and solubility at relatively high concentrations.[61]

It has been known that high concentrations of cosolvents such as sugars, polyols, and ammonium sulfate stabilize the native state of proteins, whereas other cosolvents such as urea and guanidine hydrochloride act as protein denaturants.[61, 64-65] Denaturants prefer binding to the unfolded state than to the native state. The surface of a protein molecule excludes protein stabilizers, and the exclusion of them increases as its solvent exposed surface area increases.[61, 66-67] Preferential exclusion can thus be described as negative binding. During denaturation, the protein surface area increases, leading to a greater degree of preferential exclusion. The net effect of greater negative binding shifts the equilibrium from the unfolded state to the native state.

In the case of ions, they can control the strength of both intra- and intermolecular electrostatic interactions between the charged groups. At low concentrations, the predominant effect of ions in solution involves charge shielding leading to decreased electrostatic interactions. However, at high concentrations the preferential binding of ions to the protein surface can lead to a decrease in thermodynamic stability of the native conformation.[68] Other salts that are preferentially excluded from protein surfaces show stabilizing effects. In many cases, the effect of a salt on protein stability is too complicated to elucidate the dominant mechanism. Hence, it is

required that there be a combined effort of experimental and theoretical approaches to understand the thermodynamics of these systems.

Scaled Particle Theory

In order to investigate the effects of cosolvents on protein folding, association, and aggregation, the scaled particle theory (SPT) has been used to calculate the excluded volume portion of the solute transfer free energy.[69-72] This is determined from the difference in free energies between two states as a function of the additive volume fraction when hard spheres are inserted into solutions of the initial state and final state. These calculations generally suggest that the native and any associated states are favored on increasing the additive concentration. These trends observed in experiments[69] and in simulations using simple excluded volume crowders.[73-74] The above trends are usually interpreted in terms of a decrease in the free volume, or an increase in the excluded volume, within the solution. However, SPT does not interpret attractive interactions which undoubtedly occur between proteins in real systems.[69] Hence, other models are still required.

Preferential Interactions

Aggregation in a solution mixture involves the balance of intermolecular interactions between solute-solute and solvent-solute. If the solute-solute interactions are larger than solute-solvent interactions, self-association is likely to occur and the tendency for aggregation can be predicted using the difference between solute-solute and solute-solvent interactions. Hence, it is attractive to quantitatively express the difference between solute-solute and solute-solvent interactions. Figure 1.9 indicates that cosolvents will tend to change the chemical potential of a

protein in a cosolvent solution, compared to pure solvent, due to either preferential interactions with, or exclusions from, the protein interface.

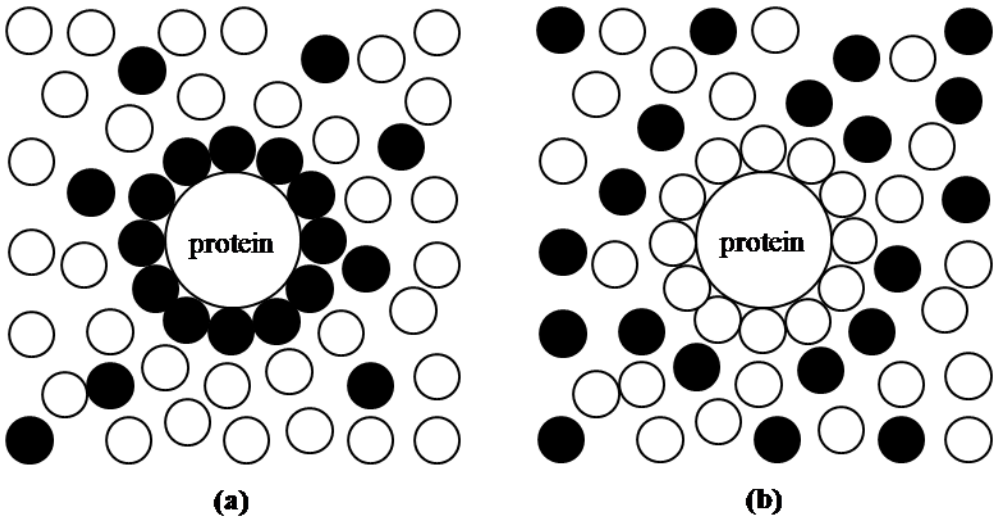


Figure 1.9 The distribution of water molecules (white circles) and cosolvent molecules (black circles) in a mixture of protein, cosolvent, and water mixture (a) **The protein prefers to be surrounded by cosolvent molecules** (b) **The transfer from pure water to the cosolvent solution in unfavourable. The protein prefers to be surrounded by water molecules.**

Smith and coworkers have extended KB theory to the analysis of experimental and computer simulation data relating to the interaction of cosolvents with proteins.[38, 75-77] The preferential interaction (PI) of a cosolvent with a protein measures the change in cosolvent (2) molality (m_2) on changing the biomolecule (3) molality (m_3) in a system open to the cosolvent (2) and water (1). This is also often referred to as the preferential binding parameter.[78] In the infinitely dilute biomolecule limit an exact expression for Γ_{23} in terms of KB integrals can be obtained by using number density (ρ_i), Kirkwood-Buff integrals (G_{ij}), and excess coordination number (N_{ij}).[75, 77]

$$\Gamma_{23} = \left(\frac{\partial m_3}{\partial m_2} \right)_{T,\mu_1,\mu_3} = \rho_3 (G_{23} - G_{21}) = N_{23} - \frac{\rho_3}{\rho_1} N_{21} \tag{1.2}$$

The concept of preferential interaction (PI) has been introduced previously and KB integrals can play an important role in quantifying these PIs.[79] In cellular systems, it is very difficult to estimate the balance of the interactions between solute-solute and solvent-solute. However, it would be useful to be able to quantify these interactions. Hence, we will investigate the effects of a cosolvent on the equilibrium constant for association of a solute in solution using Kirkwood-Buff theory in Chapter 4.

Summary

MD simulations have been used to investigate biological systems by providing details at the atomic level. Kirkwood-Buff theory is a useful tool to interpret and connect experimental and computational data. Here, we apply Kirkwood-Buff theory and computer simulations to various interesting biological environments.

In chapter 2, force fields for the simulation of alkali halide aqueous solutions are developed specifically to reproduce the experimentally determined Kirkwood-Buff integrals and solution activities as a function of molality. Additionally, experimentally known properties: ion diffusion constants, relative permittivities, densities and heats of mixing are also reproduced by these models.

In chapter 3, in an effort to understand the interactions which occur between amino acids in solution we describe new force fields for simple amino acids and their analogs including glycine, betaine, β-alanine, *dl*-alanine, NH_4Cl, NH_4Br, $N(CH_3)_4Cl$, $N(CH_3)_4Br$, CH_3NH_3Cl, and CH_3COONa. The new force fields reproduce the experimental Kirkwood-Buff integrals describing the relative distribution of species in the solution mixture. Furthermore, it can be shown that these simple amino acids can be understood in terms of the interactions of their functional groups and that, to a very good approximation, the transferability and additivity usually assumed in the development of biomolecular force fields appears to hold true.

In chapter 4, the effect of a cosolvent on the association of a solute in solution using the Kirkwood-Buff theory of solutions is presented. The derived expressions provide a foundation for the investigation of cosolvent effects on molecular and biomolecular equilibria including protein association, aggregation, and cellular crowding.

In chapter 5, in an effort to understand peptide aggregation at the atomic level we have performed simulations of polyglycine ($(gly)_n$) using our recently developed force fields. Experimentally, the association of glycine polypeptides increases with n. Our force fields reproduce this behavior, and the reasons behind this trend are investigated. In addition, we also simulate these systems in a semi-open ensemble, designed to mimic cellular environments typically open to water, using a simple approach. The differences between the two ensembles are investigated and compared with our recent theoretical descriptions of aggregating systems using Kirkwood-Buff theory.

References

1. J. D. Batchelor, A. Olteanu, A. Tripathy and G. J. Pielak, J Am Chem Soc **126** (7), 1958-1961 (2004).
2. R. Truant, R. S. Atwal, C. Desmond, L. Munsie and T. Tran, Febs J **275** (17), 4252-4262 (2008).
3. A. A. Chen and R. V. Pappu, Biophys J, 412a-412a (2007).
4. A. A. Chen and R. V. Pappu, J Phys Chem B **111** (41), 11884-11887 (2007).
5. P. Auffinger, T. E. Cheatham and A. C. Vaiana, J Chem Theory Comput **3** (5), 1851-1859 (2007).
6. J. N. Israelachvili, *Intermolecular and surface forces*, 2nd ed. (Academic, London, 1992).
7. E. Evans, K. Ritchie and R. Merkel, Biophys J **68** (6), 2580-2587 (1995).
8. G. Binnig, C. F. Quate and C. Gerber, Phys Rev Lett **56** (9), 930-933 (1986).
9. K. Svoboda and S. M. Block, Annu Rev Bioph Biom **23**, 247-285 (1994).
10. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller and E. Teller, J Chem Phys **21** (6), 1087-1092 (1953).
11. G. Favrin, A. Irback and S. Mohanty, Biophys J **89** (1), 754-754 (2005).
12. S. L. Chaplot, Comp Mater Sci **37** (1-2), 146-151 (2006).
13. W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger and W. F. van Gunsteren, J Phys Chem A **103** (19), 3596-3607 (1999).
14. S. Santini, N. Mousseau and P. Derreumaux, J Am Chem Soc **126** (37), 11509-11516 (2004).

15. N. Mousseau and P. Derreumaux, Accounts Chem Res **38** (11), 885-891 (2005).

16. A. Melquiond, N. Mousseau and P. Derreumaux, Proteins **65** (1), 180-191 (2006).

17. S. Mclaughlin, Annu Rev Biophys Bio **18**, 113-136 (1989).

18. C. F. Anderson and M. T. Record, Annu Rev Phys Chem **46**, 657-700 (1995).

19. R. L. Baldwin, Biophys J **71** (4), 2056-2063 (1996).

20. G. Colombo, P. Soto and E. Gazit, Trends Biotechnol **25** (5), 211-218 (2007).

21. A. K. Rappe and W. A. Goddard, J Phys Chem-Us **95** (8), 3358-3363 (1991).

22. S. W. Rick, S. J. Stuart and B. J. Berne, J Chem Phys **101** (7), 6141-6156 (1994).

23. T. A. Halgren and W. Damm, Curr Opin Struc Biol **11** (2), 236-242 (2001).

24. A. Laio, J. VandeVondele and U. Rothlisberger, J Phys Chem B **106** (29), 7300-7307 (2002).

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

26. A. Bennaim, J Chem Phys **67** (11), 4884-4890 (1977).

27. J. Rosgen, B. M. Pettitt and D. W. Bolen, Protein Sci **16** (4), 733-743 (2007).

28. M. Auton, D. W. Bolen and J. Rosgen, Proteins **73** (4), 802-813 (2008).

29. J. Rosgen, B. M. Pettitt and D. W. Bolen, Biophys J **89** (5), 2988-2997 (2005).

30. N. Matubayasi, W. Shinoda and M. Nakahara, J Chem Phys **128** (19), - (2008).

31. R. Adams, H. H. M. Balyuzi and R. E. Burge, J Appl Crystallogr **10** (Aug1), 256-261 (1977).

32. K. Yamanaka, M. Yamagami, T. Takamuku, T. Yamaguchi and H. Wakita, J Phys Chem-Us **97** (41), 10835-10839 (1993).

33. S. Shimizu and C. L. Boon, J Chem Phys **122** (18), - (2005).

34. S. Shimizu and C. L. Boon, J Chem Phys **121** (18), 9147-9155 (2004).

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

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

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

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

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

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

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

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

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

44. E. M. Duffy, D. L. Severance and W. L. Jorgensen, Israel J Chem **33** (3), 323-330 (1993).

45. O. Engkvist, S. L. Price and A. J. Stone, Phys Chem Chem Phys **2** (13), 3017-3027 (2000).

46. R. Dovesi, M. Causa, R. Orlando, C. Roetti and V. R. Saunders, J Chem Phys **92** (12), 7402-7411 (1990).

47. M. Kang and P. E. Smith, J Chem Phys **128** (24), - (2008).

48. A. Wang, A. D. Robertson and D. W. Bolen, Biochemistry-Us **34** (46), 15096-15104 (1995).

49. J. F. Carpenter, B. S. Kendrick, B. S. Chang, M. C. Manning and T. W. Randolph, Method Enzymol **309**, 236-255 (1999).

50. A. L. Fink, Fold Des **3** (1), R9-R23 (1998).

51. Y. S. Kim, S. P. Cape, E. Chi, R. Raffen, P. Wilkins-Stevens, F. J. Stevens, M. C. Manning, T. W. Randolph, A. Solomon and J. F. Carpenter, J Biol Chem **276** (2), 1626-1633 (2001).

52. Y. F. Maa and C. C. Hsu, Int J Pharm **140** (2), 155-168 (1996).

53. S. Krishnan, E. Y. Chi, J. N. Webb, B. S. Chang, D. X. Shan, M. Goldenberg, M. C. Manning, T. W. Randolph and J. F. Carpenter, Biochemistry-Us **41** (20), 6422-6431 (2002).

54. B. S. Kendrick, J. L. Cleland, X. Lam, T. Nguyen, T. W. Randolph, M. C. Manning and J. F. Carpenter, J Pharm Sci-Us **87** (9), 1069-1076 (1998).

55. Y. S. Kim, J. S. Wall, J. Meyer, C. Murphy, T. W. Randolph, M. C. Manning, A. Solomon and J. F. Carpenter, J Biol Chem **275** (3), 1570-1574 (2000).

56. E. Y. Chi, S. Krishnan, B. S. Kendrick, B. S. Chang, J. F. Carpenter and T. W. Randolph, Protein Sci **12** (5), 903-913 (2003).

57. C. M. Dobson, Nature **426** (6968), 884-890 (2003).

58. C. A. Ross and M. A. Poirier, Nat Med, S10-S17 (2004).

59. M. Reches, Y. Porat and E. Gazit, J Biol Chem **277** (38), 35475-35480 (2002).

60. M. Reches and E. Gazit, Science **300** (5619), 625-627 (2003).

61. S. N. Timasheff, Adv Protein Chem **51**, 355-432 (1998).

62. J. A. Schellman, Biopolymers **17** (5), 1305-1322 (1978).

63. J. Wyman and S. J. Gill, *Binding and linkage : functional chemistry of biological macromolecules*. (University Science Books, Mill Valley, CA, 1990).

64. P. R. Davis-Searles, A. J. Saunders, D. A. Erie, D. J. Winzor and G. J. Pielak, Annu Rev Bioph Biom **30**, 271-306 (2001).

65. M. G. Cacace, E. M. Landau and J. J. Ramsden, Q Rev Biophys **30** (3), 241-277 (1997).

66. T. Arakawa and S. N. Timasheff, Biochemistry-Us **21** (25), 6536-6544 (1982).

67. J. C. Lee and S. N. Timasheff, J Biol Chem **256** (14), 7193-7201 (1981).

68. T. Arakawa and S. N. Timasheff, Biochemistry-Us **23** (25), 5912-5923 (1984).

69. H. X. Zhou, G. N. Rivas and A. P. Minton, Ann Rev Biophys **37**, 375-397 (2008).

70. A. P. Minton, G. C. Colclasure and J. C. Parker, P Natl Acad Sci USA **89** (21), 10504-10506 (1992).

71. S. B. Zimmerman and A. P. Minton, Annu Rev Bioph Biom **22**, 27-65 (1993).

72. H. X. Zhou, Proteins **72** (4), 1109-1113 (2008).

73. M. S. Cheung, D. Klimov and D. Thirumalai, P Natl Acad Sci USA **102** (13), 4753-4758 (2005).

74. M. Wojciehowski and M. Cieplak, Biosystems **94** (3), 248-252 (2008).

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

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

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

78. I. L. Shulgin and E. Ruckenstein, Biophys J **90** (2), 704-707 (2006).

79. M. Kang and P. E. Smith, Fluid Phase Equilibr **256** (1-2), 14-19 (2007).

CHAPTER 2 - A Kirkwood-Buff Derived Force Field for Alkali Halides in Water

We describe a force field for the simulation of aqueous alkali halide solutions. These models are developed specifically to reproduce the experimentally determined Kirkwood-Buff integrals and the solution activities as a function of molality. Additionally, we demonstrate that these models reproduce other experimental properties including ion diffusion constants, relative permittivity, the density and heat of mixing.

Introduction

In an effort to develop force fields for the accurate simulation of biologically interesting solution mixtures, we have recently been developing a specialized force field, the Kirkwood-Buff force field (KBFF).[1-7] The parameters of the KBFF are determined using molecular dynamics simulations, the Kirkwood-Buff (KB) theory of solutions, and experimental data for activity coefficients and solution densities. This approach has several advantages. First, KB theory is exact including no approximations. Second, KB theory can be applied to any solution whose activity coefficients and densities are available. Third, the KB integrals, which are the main quantities that result from the application of the KB theory, are easily obtainable from the radial distribution functions (rdf) through MD simulations and are very sensitive to the force field parameters. And fourth, the KB integrals quantify the relative strength between solute-solute and solute-solvent interactions, and therefore, describe the correct distribution of solutes in solution.[2, 8]

Aqueous solutions of alkali metal halides are not only the simplest models for the application of the KB theory to aqueous electrolyte solutions, but they also play an important role in many biological systems. They stabilize important biomolecules like proteins, nucleic acids, and lipids, and they are often involved in biological catalysis.[9-11] In the case of protein stabilization, for instance, the Hofmeister series predicts that as the molar mass of an alkali metal ion increases, so does its ability to stabilize native a protein.[11]

Because of their importance in biological phenomena, several force fields for alkali metal and halide ions have been reported in the literature.[12] Unfortunately, the force fields have been shown to be incompatible with each other.[12] Since 2008 however, there have been two attempts to develop force fields that are consistent for all alkali metals and all halide ions. First, in 2008, Joung & Cheatham[12] used the free energy of hydration of individual ions, as well as the lattice energies and the lattice constants of alkali metal halides, in order to create force fields for all alkali metal and halide ions. Joung & Cheatham[12] developed parameter sets for the three commonly used nonpolarizable water models, SPC/E, TIP3P, and TIP4P$_{EW}$. Second, in 2009, Horinek *et al.*[13] used both the free energy and the entropy of hydration of the individual ions in order to parameterize their force fields. Horinek *et al.*[13] focused on a single nonpolarizable water model, SPC/E, and argued that their force field would be more applicable in biomolecular simulations where the salt concentrations are low, and that the Joung & Cheatham force fields would be more applicable when the salt concentrations are high.[12]

Although the Joung & Cheatham and the Horinek *et al.* force fields reproduce a series of properties, (including the first peak of the ion-water radial distribution function (rdf), ion-water binding energies, interionic distances, diffusion coefficients, solubilities, and association constants) they were not designed to be applicable over the entire concentration range. Also,

those force fields were parameterized using free energies and entropies of solvation, an approach that does not probe ion-ion interactions. The KB approach, on the other hand, provides force fields applicable over the entire concentration range, as demonstrated in our previous work.[2] Since 2009, two research groups have produced KB-derived force fields for alkali metal halides recently. Hess & van der Vegt used the SPC/E water model to develop KB-derived force fields for Li^+ and K^+ in order to explain the differential binding affinity of alkali metal ions to carboxylate ions.[14] And Klasczyk & Knecht used the SPC water model and our force field for the chloride ion to develop KB-derived force fields for Li^+, K^+, Rb^+, and Cs^+, but not for halide ions.[15] The Klasczyk & Knecht force field is incompatible with ours because we use SPC/E water model. In this paper, we present a KB-derived force field for a variety of alkali metal and halide ions that is applicable over the whole concentration range.

Methods

Kirkwood-Buff Theory

Kirkwood-Buff theory is an exact theory of solution.[16] The central properties of interest are the Kirkwood-Buff integrals (KBIs) which are defined by,

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

Here, G_{ij} is the KB integral between species i and j, $g_{ij}^{NpT}(r)$ is the corresponding radial distribution function (rdf) in the NpT ensemble, r is the distance between the two species, and R represents a correlation region within which the solution composition differs from the bulk composition. All rdfs are assumed to be unity beyond R. Excess coordination numbers are defined as $N_{ij} = \rho_j G_{ij}$, where $\rho_j = N_j/V$ is the number density of j particles. A value of N_{ij}

greater than zero generally 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*.

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}, and $G_{cw} = G_{wc}$, and the number densities (or molar concentrations) ρ_w and ρ_c. The partial molar volumes of the components ($\overline{V}_i$) and the derivative of the cosolvent activity ($a_c = y_c\rho_c$) at a pressure (p) and a temperature (T) are given by,[2]

$$\begin{aligned} &\overline{V}_w = \frac{1+\rho_c(G_{cc}-G_{cw})}{\eta}, \overline{V}_c = \frac{1+\rho_c(G_{ww}-G_{cw})}{\eta}, \\ &\eta = \rho_w + \rho_c + \rho_w\rho_c(G_{ww}+G_{cc}-2G_{cw}) \\ &a_{cc} = \left(\frac{\partial \ln a_c}{\partial \ln \rho_c}\right)_{p,T} = 1+\left(\frac{\partial \ln y_c}{\partial \ln \rho_c}\right)_{p,T} = \frac{1}{1+\rho_c(G_{cc}-G_{cw})} \end{aligned} \quad . \qquad (2.2)$$

There are no approximations made during the derivation of the above equations. Our previous simulations and others have indicated that a combination of KB theory and *NpT* simulations can provide quantitative information concerning the thermodynamics of solutions.[2-4, 6-7] The salt solution needs to be treated as a binary system of indistinguishable ions and water when KB theory is applied to electrolyte solutions.[2, 5] Therefore, we distinguish between the cosolvent (total ion) concentration, ρ_c, and the classical salt concentration, C_s. Consequently, for a 1:1 salt one has $\rho_c = 2C_s, \overline{V}_s = 2\overline{V}_c$, and $y_c = y_\pm$. In addition, the following relationships are also obeyed, $\rho_c\overline{V}_c + \rho_w\overline{V}_w = 1$ and $\rho_c \mathrm{d}\ln a_c + \rho_w \mathrm{d}\ln a_w = 0$, at constant p and T.

Molecular Dynamics Simulations

All molecular dynamic simulations of alkali halide solutions were performed using the SPC/E water model[17] in the isothermal isobaric (*NpT*) ensemble at 300 K and 1 atm as implemented in the GROMACS program (v3.3.1).[18-19] A time-step of 2 fs was used and the geometry of the water molecules was constrained using SETTLE.[20] The weak coupling technique was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively.[21] In order to evaluate electrostatic interactions, the particle mesh Ewald technique (PME) was used.[22] The initial cubic boxes of different solutions have been generated by adding water molecules and ions until the required concentration was obtained. Configurations were saved every 0.1 ps for analysis. Diffusion constants were determined using the mean square fluctuation approach,[23] and relative permittivities were obtained from the dipole moment fluctuations.[24] The excess enthalpy of mixing (ΔH_m) was determined by an established procedure which uses the average potential energies,[25] using configurational energies from the pure SPC/E water and the alkali halide lattice.

Parameter Development

The force field used in this study corresponds to the Lennard-Jones (LJ) 6-12 potential which contains most commonly two adjustable parameters in ionic force developments: the Lennard-Jones diameter (σ) and the interaction strength (ε) plus a Coulomb potential, combined with the SPC/E water model.[17] In this scheme each pair of atoms *i* and *j* interact with an interaction energy given by

$$V_{ij} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \qquad (2.3)$$

Here, all the symbols have their usual meaning.[6] In order to obtain parameters for the LJ term, we have employed the same method we published previously.[2] Hence, we require three pieces of experimental data, i.e the ionic radii of alkali and halide ions which are consistent with the crystal lattice dimension, the crystal lattice unit cell dimension, and the ion to water oxygen contact distances.

Table 2.1 Experimental data for parameter development: *r,* the ionic radii of alkali halide ions which are consistent with the crystal lattice dimension; *a,* the crystal lattice unit cell dimension; *d,* the ion to water oxygen contact distances

	Li^+	K^+	Rb^+	Cs^+	F^-	Cl^-	Br^-	I^-
			Cl^-			Na^+		
r (nm)	0.115	0.138	0.149	0.170	0.133	0.181	0.196	0.220
a (nm)	0.257	0.319	0.332	0.412	0.239	0.282	0.299	0.324
d (nm)	0.213	0.280	0.289	0.314	0.263	0.319	0.338	0.365
Reference	26-27	26-27	26-27	26-27	26-27	2	26-27	26-27

The final parameters have been developed by systemically increasing or decreasing the parameters of each ion, in accordance with the ionic scaling factors used.[2] First, we parameterized the anions (F^-, Br^-, I^-) by studying NaF, NaBr, and NaI. We have used the same values of Na^+ in terms of σ and ε. After the values of σ_{--} were determined by scaling the ionic radii of each ion with the scaling factor (2.43), which was used for the parameter-development of Cl^-,[2] the values of ε_{--} were determined by increasing or decreasing the values of ε_{--} of X^- until the experimental lattice dimensions of the sodium halide were reproduced by simulation. The values determined for each ion were then applied to study the experimental KBIs in aqueous solutions. Unfortunately, in the case of F^-, σ_{FF} and ε_{FF} did not reproduce experimental KBIs in aqueous solutions. Hence, we decided to develop the values of σ_{FF} and ε_{FF} in solution which are different from those in crystal structure. Second, we developed the parameters of cations (K^+, Rb^+, Cs^+) using the properties of KCl, RbCl, and CsCl. After the values of σ_{++} were determined by scaling the ionic radii of each ion, the values of ε_{++} for each cation were determined in the same way as they were determined for anions. Unfortunately, we could not reproduce experimental KBIs in aqueous solution by using standard combination rules with these σ_{++} and ε_{++} in aqueous solutions. Hence, we needed to break combination rules for the determination of the values of ε_{+O} for each cation, as was also required for NaCl solutions. Finally, as a test we applied these parameters to study the cation-anion exchanged solution systems of CsBr and KI in order to demonstrate the transferability of our parameters.

Table 2.2 shows the Lennard-Jones parameters used in our simulations. The LJ parameters for Na^+ and Cl^- were taken from Weerasinghe and Smith,[2] and those for Li^+ from Hess & van der Vegt.[14] As the size of the cation increased, the value of σ decreased and that of ε decreased. A similar trend is observed for the anions, which is expected. This trend in the values of σ was also

observed by both Joung & Cheatham,[12] as well by Horinek et al.[13] However the trend in the values of ε was absent from the work of both research groups.

Table 2.2 Final force field parameters for the KBFF model

Model	Atom	σ_{ii} (nm)	ε_{ii} (kJ/mol)	ε_{iO} (kJ/mol)	q (e)
KBFF	Li	0.182	0.7000	0.2700[a]	+1.0
	Na[2]	0.2450	0.3200	0.3420[b]	+1.0
	K	0.3340	0.1300	0.2327[c]	+1.0
	Rb	0.3620	0.1500	0.2655[d]	+1.0
	Cs	0.4130	0.0065	0.1954[e]	+1.0
	F	0.3700	1.0000		-1.0
	Cl[2]	0.4400	0.4700		-1.0
	Br	0.4760	0.3000		-1.0
	I	0.535	0.2000		-1.0
SPC/E	O[17]	0.3166	0.6506		-0.8476
	H[17]	0.0000	0.0000		+0.4238

The combination rules used were the following: $\sigma_{ij} = \sqrt{\sigma_{ii} \times \sigma_{jj}}$, $\varepsilon_{ij} = \sqrt{\varepsilon_{ii} \times \varepsilon_{jj}}$,

$a: \varepsilon_{OLi} = 0.4 \times \sqrt{\varepsilon_{OO} \times \varepsilon_{LiLi}}$, $b: \varepsilon_{ONa} = 0.75 \times \sqrt{\varepsilon_{OO} \times \varepsilon_{NaNa}}$, $c: \varepsilon_{OK} = 0.8 \times \sqrt{\varepsilon_{OO} \times \varepsilon_{KK}}$,

$d: \varepsilon_{ORb} = 0.85 \times \sqrt{\varepsilon_{OO} \times \varepsilon_{RbRb}}$, $e: \varepsilon_{OCs} = 0.95 \times \sqrt{\varepsilon_{OO} \times \varepsilon_{CsCs}}$

Results

A series of molecular dynamics simulations of alkali halide solutions were performed to validate the models and are summarized in Table 2.3.

Table 2.3 Summary of the MD simulations of alkali halide water mixtures: All simulations were performed at 300 K and 1atm in the *NpT* ensemble. Symbols are N_w, number of water molecules; N_c, number of alkali-halide ions; N_s (= $N_+ = N_- = 1/2N_c$), number of alkali-halide pairs; V, average simulation volume; m_s, salt molality; C_s, salt molarity; ρ, mass density; E_{pot}, average total potential energy per molecule ($N_s + N_w$); and T_{sim}, total simulation time.

	N_s	N_w	m_s (mol/Kg)	V (nm^3)	C_s (mol/l)	ρ (g/cm^3)	E_{pot} (kJ/mol)	T_{sim} (ns)
H_2O	0	2170	0.00	65.265	0.00	0.995	-46.45	2
NaF	20	2150	0.52	64.531	0.52	1.018	-54.94	6
	38	2079	1.01	64.519	1.03	1.040	-63.05	6
NaCl	38	2079	1.01	63.595	0.99	1.036	-60.07	6
	77	2048	2.09	63.829	2.00	1.077	-73.99	4
	115	1987	3.21	63.354	3.01	1.114	-88.00	4
	154	1950	4.38	63.783	4.01	1.149	-102.02	4
NaBr	38	2079	1.01	64.089	0.98	1.072	-59.62	5
	77	2048	2.09	64.810	1.97	1.148	-73.05	5
	115	1987	3.21	64.730	2.95	1.222	-86.59	9
	154	1950	4.38	65.584	3.90	1.291	-100.08	5
	231	1730	7.41	63.426	6.05	1.438	-132.53	9
	308	1600	10.69	64.350	7.95	1.562	-163.90	9
NaI	38	2079	1.01	65.051	0.97	1.101	-58.86	5
	77	2048	2.09	66.683	1.92	1.206	-71.52	5
	115	1987	3.21	67.458	2.83	1.305	-84.27	5
	154	1950	4.38	69.151	3.70	1.398	-96.97	5
	231	1730	7.41	68.599	5.59	1.593	-127.35	5
	308	1600	10.69	71.253	7.18	1.748	-156.79	5
LiCl	127	7065	1.00	216.903	0.97	1.016	-62.37	6
	367	6796	3.00	217.842	2.80	1.052	-92.29	6
	589	6541	5.00	219.083	4.47	1.082	-120.13	6
KCl	126	7002	1.00	216.178	0.97	1.041	-58.36	6

	357	6603	3.00	215.924	2.75	1.120	-80.74	6
	561	6228	5.00	216.204	4.31	1.183	-101.56	6
RbCl	125	6963	1.00	216.055	0.96	1.080	-57.79	6
	352	6512	3.00	215.995	2.71	1.229	-79.14	6
	549	6093	5.00	216.324	4.22	1.352	-99.03	6
CsCl	125	6915	1.00	216.032	0.96	1.119	-57.38	6
	345	6385	3.00	215.737	2.66	1.333	-77.74	6
	533	5920	5.00	216.025	4.10	1.510	-96.76	6
KI	124	6880	1.00	217.183	0.95	1.105	-57.18	6
	340	6300	3.00	218.381	2.59	1.292	-77.20	6
	522	5796	5.00	219.331	3.95	1.447	-95.97	6
CsBr	124	6870	1.00	216.248	0.95	1.153	-56.93	6
	339	6275	3.00	216.377	2.60	1.422	-76.47	6
	519	5761	5.00	216.598	3.98	1.640	-94.77	6

Table 2.4 shows the potential energy, density and lattice constants for the salt crystals studied in this work. For all salts but the iodides, the simulated values exhibit a maximum percent error of 4%. On the other hand, the simulations overestimate the density of NaI by 6% and that of KI by 9%, while they underestimate the lattice constant of NaI by 6%. These effects are due to a slight overestimation of attraction between the iodide anion and the sodium and potassium cations.

Table 2.4 Summary of lattice parameters obtained from crystal simulation: Symbols are E_{pot}, average total potential energy per molecule (N_s + N_w); ρ_{sim}, mass density from simulation; ρ_{exp}, mass density from experiment.

	E_{pot} (kJ/mol)	ρ_{sim} (g/cm^3)	ρ_{exp} (g/cm^3)[26]	a_{sim} (nm)	a_{exp} (nm)[26]
NaF	-969.64	2.646	2.558	0.236	0.235
NaCl	-808.24	2.108	2.165	0.285	0.282
NaBr	-776.08	3.326	3.203	0.295	0.299
NaI	-750.94	3.878	3.667	0.303	0.324
LiCl	-1178.03	1.776	2.068	0.261	0.257
KCl	-725.29	1.980	1.984	0.315	0.315
RbCl	-692.73	2.800	2.800	0.330	0.332
CsCl	-650.12	3.990	3.990	0.419	0.412
KI	-663.23	3.406	3.123	0.343	0.353
CsBr	-628.80	4.580	4.440	0.424	0.429

The radial distribution functions (rdfs) obtained from 1 M simulations are shown in Figure 2.1 for the sodium halides and in Figure 2.2 for the alkali metal chlorides. The sodium to halide rdfs displayed a large first and a significant second peak, in agreement with experiment,[28] and all rdfs approach unity beyond 1 nm.

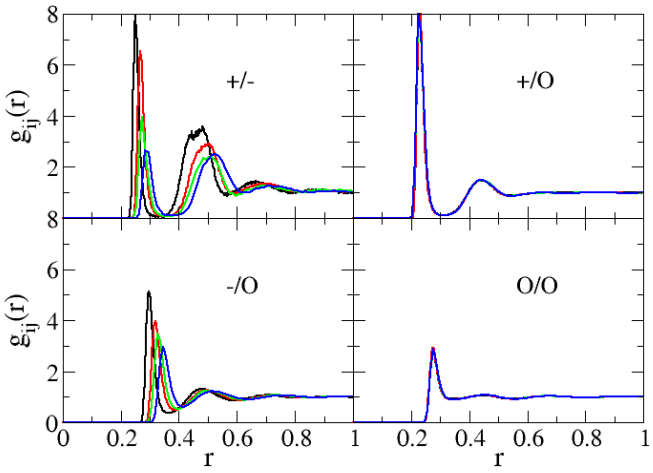


Figure 2.1 Radial distribution functions of 1 M solutions obtained from the NaF (black lines), NaCl (red lines), NaBr (green lines), and NaI (blue lines) simulations. Cations, anions, and the water oxygen are denoted by the symbols +, -, and O, respectively.

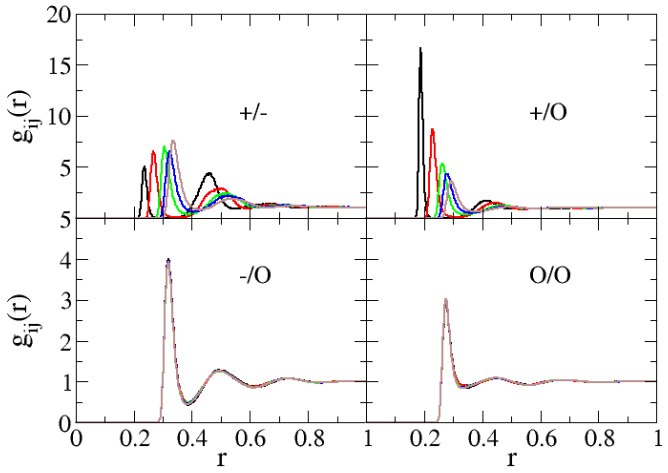


Figure 2.2 Radial distribution functions of 1 M solutions obtained from the LiCl (black lines), NaCl (red lines), KCl (green lines), RbCl (blue lines), and CsCl (brown lines) simulations. Cations, anions, and the water oxygen are denoted by the symbols +, -, and O, respectively.

The first shell coordination numbers, n_{ij}, as well as the distances to the first rdf maximum, R_{max}, and the first rdf minimum, R_{min}, were calculated from the corresponding rdfs as a function of the solution molarity and are presented in Table 2.5. According to Table 2.5, the radii of the first hydration shell of Na^+, K^+, Rb^+, Cs^+, F^-, Cl^-, Br^-, and I^- are 0.23, 0.26, 0.28, 0.29, 0.27, 0.32, 0.33 and 0.35 nm, respectively. As the size of the cation increases, so does the radius of the first hydration shell; the same trend is exhibited by the anions. The predicted values agree with experimental values[27] (0.24, 0.28, 0.29, 0.31, 0.26, 0.32, 0.34, and 0.36, respectively) within a 0.1 nm root mean square (rms) deviation, a deviation that is also exhibited by the force field published by Joung & Cheatham.[12] The first water shell coordination numbers of Na^+, K^+, Rb^+, and Cs^+ in 4 M aqueous solutions are 4.9, 5.9, 6.2, and 6.4 respectively. Similarly to the trend in

the radii of the first hydration shell, the hydration numbers increase as the size of the cation increases. The predicted hydration numbers agree with those those determined by X-ray and neutron scattering[28] (4.9, 5.3, 6.9, and 7.5, respectively) within a 0.2 rms deviation. Table 2.5 shows that the coordination numbers are not only sensitive to the size of the alkali metal ion, but also to concentration. Also, for the 1 M solutions of NaCl, NaBr, NaI, KCl, RbCl, CsCl, KI, and CsBr, the solvation numbers for the ions (refered to as +/o for cations and -/o for the anions in Table 2.5) are higher than those for water, indicating a high solvation of the ions. This trend does not hold for any of the 1M NaF solution, nor for higher concentrations of NaBr.

Table 2.5 First shell coordination numbers (n_{ij}) of as a function of concentration (m) alkali halide aqueous solutions. R_{max} and R_{min} are the distances (nm) to the first maximum and minimum of the radial distribution functions. Cations, anions, and the water oxygen are denoted by the symbols +, -, and o, respectively.

		m	+/-	+/o	-/o	o/o
NaF	R_{max}		0.23	0.23	0.27	0.28
	R_{min}		0.285	0.315	0.335	0.355
	n_{ij}	0.51	0.03	5.62	6.54	4.77
		0.98	0.03	5.65	6.59	6.35
NaCl	R_{max}		0.27	0.23	0.32	0.28
	R_{min}		0.355	0.315	0.405	0.345
	n_{ij}	0.99	0.09	5.52	8.11	5.12
		2.00	0.20	5.39	8.22	5.11
		3.01	0.42	5.11	8.38	5.06
		4.01	0.57	4.93	8.45	5.00
NaBr	R_{max}		0.28	0.23	0.33	0.28
	R_{min}		0.365	0.315	0.415	0.405
	n_{ij}	0.98	0.10	5.49	7.54	5.09
		1.97	0.22	5.34	7.62	6.57
		2.95	0.35	5.17	8.11	7.06
		3.90	0.50	4.97	8.61	8.11
		6.05	0.96	4.33	9.66	8.72
		7.95	1.45	3.67	10.31	8.03
NaI	R_{max}		0.29	0.23	0.35	0.28
	R_{min}		0.375	0.315	0.425	0.395
	n_{ij}	0.97	0.07	5.48	7.91	5.04
		1.92	0.16	5.40	7.98	6.42
		2.83	0.26	5.24	8.04	6.83
		3.70	0.39	5.05	8.59	7.18

		5.59	0.84	4.37	9.60	7.56
		7.18	1.31	3.66	10.80	6.79
LiCl	R_{max}		0.23	0.19	0.32	0.28
	R_{min}		0.315	0.265	0.398	0.360
	n_{ij}	1	0.04	3.96	7.67	5.51
		3	0.12	3.88	7.80	5.40
		5	0.22	3.78	7.91	5.25
KCl	R_{max}		0.31	0.26	0.32	0.28
	R_{min}		0.389	0.342	0.384	0.334
	n_{ij}	1	0.20	6.11	7.03	4.32
		3	0.54	5.77	6.79	4.03
		5	0.89	5.39	6.48	3.74
RbCl	R_{max}		0.32	0.28	0.32	0.27
	R_{min}		0.404	0.361	0.385	0.332
	n_{ij}	1	0.24	6.75	7.03	4.23
		3	0.60	6.35	6.72	3.89
		5	0.95	5.95	6.32	3.57
CsCl	R_{max}		0.34	0.29	0.32	0.27
	R_{min}		0.428	0.376	0.385	0.334
	n_{ij}	1	0.34	7.15	6.91	4.28
		3	0.78	6.63	6.48	3.92
		5	1.18	6.15	6.06	3.59
KI	R_{max}		0.33	0.26	0.34	0.27
	R_{min}		0.418	0.342	0.412	0.332
	n_{ij}	1	0.24	6.02	7.34	4.20
		3	0.68	5.47	7.00	3.83
		5	1.10	4.96	6.61	3.51
CsBr	R_{max}		0.34	0.29	0.33	0.27
	R_{min}		0.438	0.380	0.394	0.332
	n_{ij}	1	0.39	7.18	6.90	4.19

	3	0.93	6.47	6.34	3.81
	5	1.38	5.88	5.85	3.49

The simulated and experimental excess coordination numbers, N_{ij}, are shown in Figure 2.3 for the sodium halides and in Figure 2.4 for the alkali metal chlorides as a function of the salt molality. The KBFF model quantitatively reproduces the experimental data, although the simulated values are not as accurate for NaI and CsCl. The water-water excess coordination number, N_{ww}, represented by green lines and symbols, remains relatively constant at approximately -0.95, independent of the molality of the various salts. The ion-ion excess coordination numbers (black lines) also do not vary significantly from salt to salt, at least compared to the variation in the ion-water excess coordination numbers (red lines). This indicates that the ion-water interactions determine the solution behavior. This suggests that the large activity derivative for Na^+ is due to its strong interaction with water, compared to that of the other halides, and not due to its size. For anions there is not a clear trend in the values of the activity derivatives, shown in Figure 2.5, presumably because of the competing effects of the anion sizes and of their interaction with water, indicated with red lines in Figure 2.3.

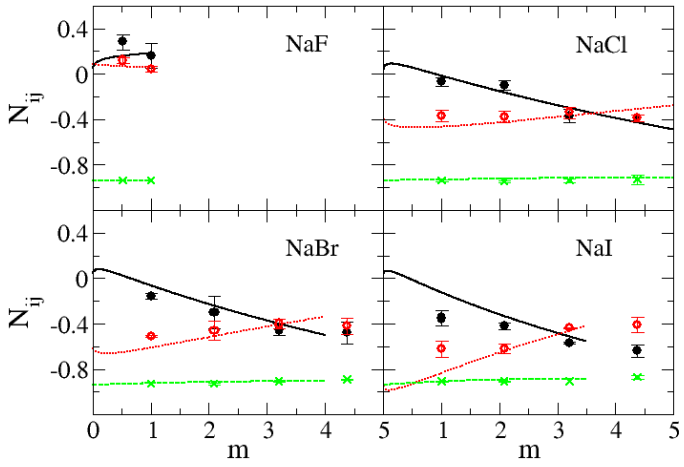


Figure 2.3 Excess coordination numbers as a function of salt molality. The N_{cc} (black lines), N_{cw} (red dotted lines), and N_{ww} (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density.[29-30] The N_{cc} (black ●), N_{cw} (red ○), and N_{ww} (green x) are obtained from simulation.

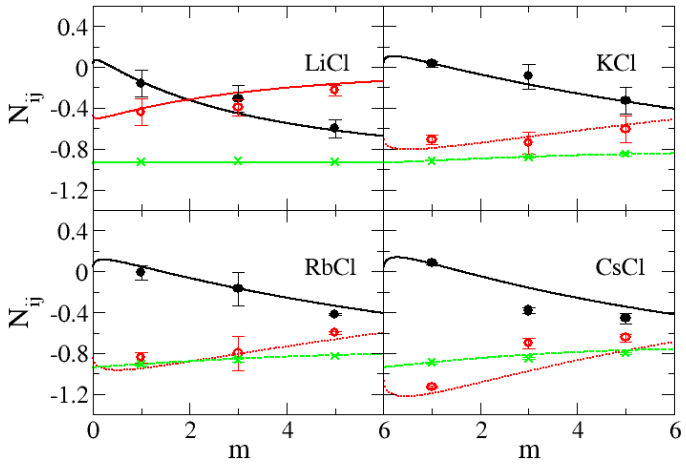


Figure 2.4 Excess coordination numbers as a function of salt molality. The N_{cc} (black lines), N_{cw} (red dot lines), and N_{ww} (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density.[29-30] The N_{cc} (black ●), N_{cw} (red ○), and N_{ww} (green x) are obtained from simulations.

In Figure 2.5 and Figure 2.6 the simulated activity derivatives a_{cc} as a function of molality are compared to the experimental values.[29] The KBFF model reproduced the correct increase in a_{cc} with concentration as displayed by the experimental data. An expression for the molar activity coefficient ($y_c = y_{\pm}$) was obtained by extending the fitting equations described in Robinson and Stokes[29]. We note that a_{cc} play an important role for describing the accurate force field.[2]

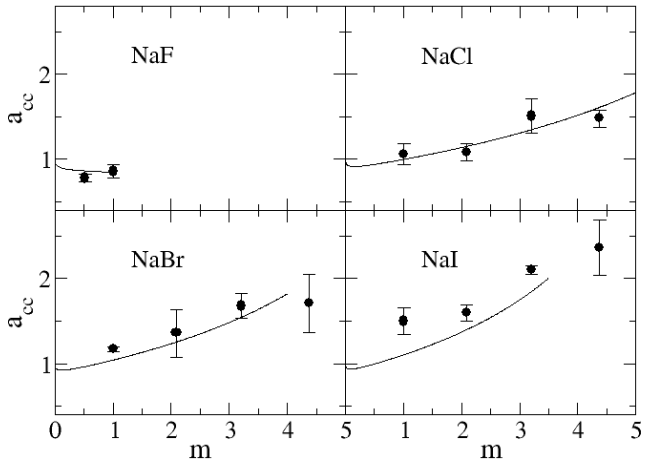


Figure 2.5 Activity derivatives as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient experimental data[29] and dots correspond to the KBFF model.

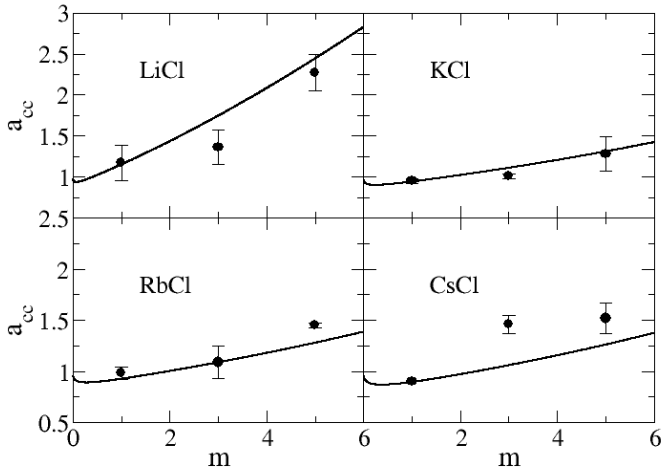


Figure 2.6 Activity derivatives as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient experimental data[29] and dots correspond to the KBFF model

Figure 2.7 and Figure 2.8 show the partial molar volumes of water and of the salts. The partial molar volume of the salts increases monotonically, and that of water decreases monotonically, as the salt concentration increases. Also, as the size of the ions increases the partial molar volume of the salt increases. The KBFF reproduces the experimental data quantitatively apart from the case of NaI, presumably due to an overestimation of ion size.

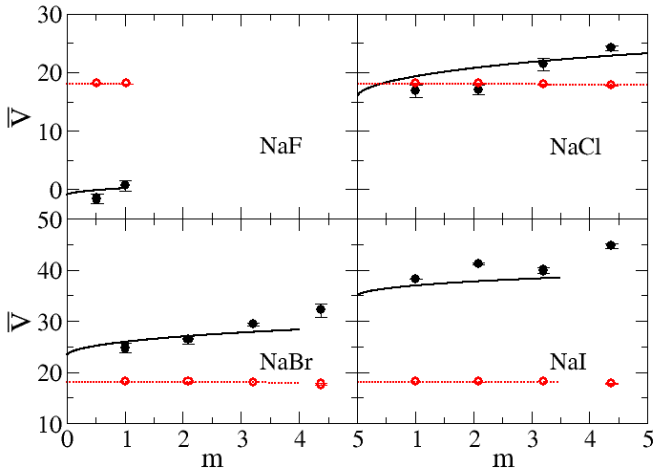


Figure 2.7 Partial molar volumes as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient and experimental density data,[29-30] and dots correspond to the KBFF model. The black lines represent the partial molar volume of salts and the red dotted lines indicate partial molar volume of water. The dots (●) represent partial molar volume of salts and the dots (red ○) indicate partial molar volume of water obtained from simulation.

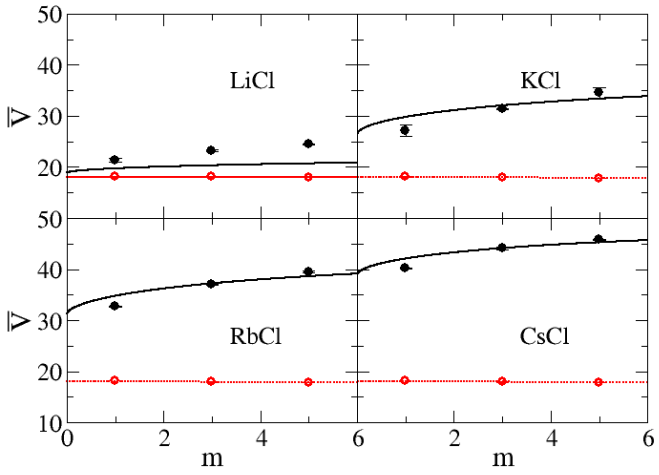


Figure 2.8 Partial molar volumes as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient and experimental density data[29-30] and dots correspond to the KBFF model. The black lines represent the partial molar volume of salts and the red dotted lines indicate partial molar volume of water. The dots (●) represent partial molar volume of salts and the dots (red ○) indicate partial molar volume of water obtained from simulation.

The self-diffusion constants, calculated using the mean square fluctuation approach,[23] are displayed in Figure 2.9 and Figure 2.10 as a function of molality of the alkali halide. They all exhibit an essentially linear decrease with molality. The self-diffusion constants of alkali ions increase even though the mass of the ions increases. It suggests that the solvation of the cation is a more important effect on the diffusion constant. On the contrary, the self-diffusion constants of halide ions do not display any correlation with the size of the ion. This suggests that the effect of solvation is compensated by the increasing mass of the ion.

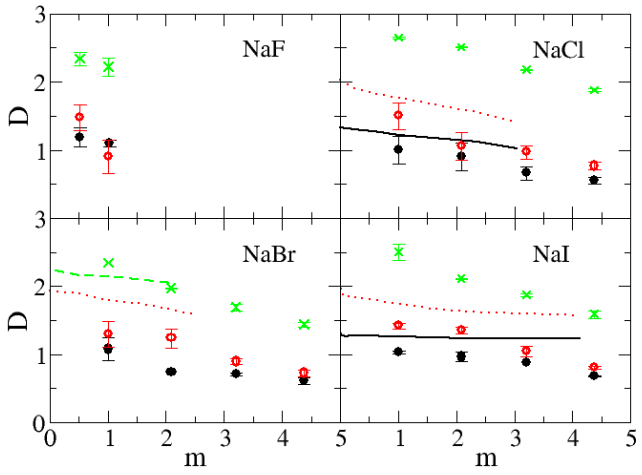


Figure 2.9 Diffusion constants as a function of salt molality. The D_+ (black lines), D_- (red dotted lines), and D_w (green dash lines) are obtained from experimental diffusion constant data[31-34] and the D_+ (black ●), D_- (red ○), and D_w (green x) are obtained from simulation.

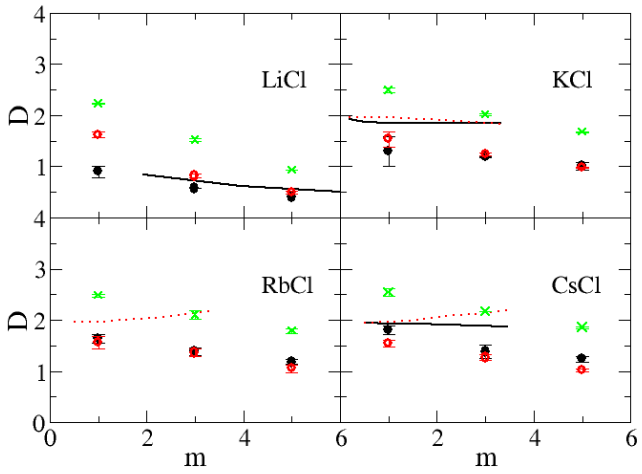


Figure 2.10 Diffusion constants as a function of salt molality. The D_+ (black lines), D_- (red dotted lines), and D_w (green dash lines) are obtained from experimental diffusion constant data[35] and the D_+ (black ●), D_- (red ○), and D_w (green x) are obtained from simulation.

The relative permittivity of alkali halides, calculated from the dipole moment fluctuations,[24] are displayed in Figure 2.11. They all exhibit an essentially linear decrease with salt concentration. The relative permittivity for all solutions decreases as a function of molarity. The KBFF models reproduce the experimental data well, partly because of the SPC/E model used for water. Here, ε_0 indicates the dielectric constants of pure water. Smith and van Gunsteren have determined the dielectric constant of the SPC/E model of liquid water which has been used this study[36]. However, the value of SPC/E models (62.7) is low compared to the experimental value (78).[37] Hence, we have applied 62.7 to calculate $\varepsilon - \varepsilon_0$ for the relative permittivity of simulations and 78 to calculate $\varepsilon - \varepsilon_0$ for the relative permittivity of experiments.

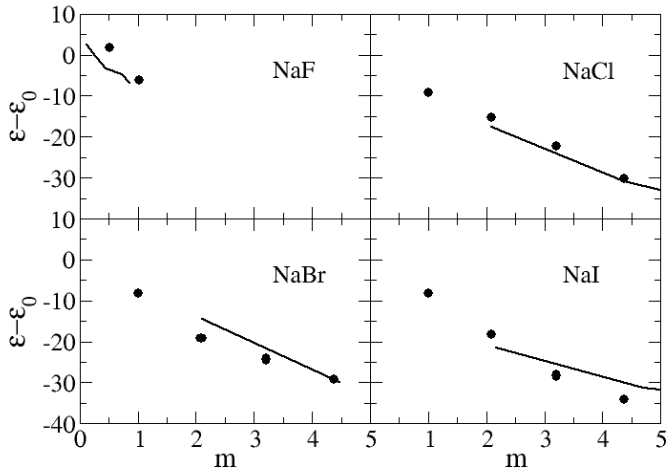


Figure 2.11 Relative permittivities as a function of salt molality. Lines are obtained from experimental dielectric constant data[38-40] and dots obtained from simulation.

The excess enthalpy of mixing, determined from the average potential energies,[25] as a function of salt molality of the sodium halides are displayed in Figure 2.12. The excess enthalpy of mixing for each sodium halide solution is calculated by the difference between the molar potential energy in the solution phase and in the crystal phase. The data indicate that the model is reproducing experimental data concerning interaction energies in solution very well.

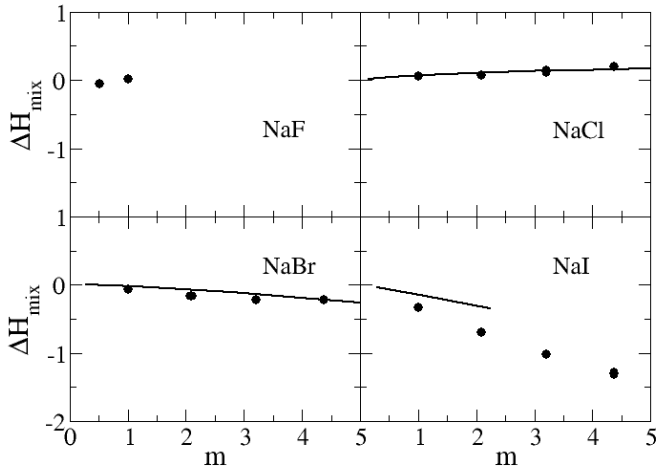


Figure 2.12 Excess enthalpy of mixing as a function of salt molality. Lines are obtained from experimental data[41] and dots obtained from simulations.

In the previous sections we have developed parameters for a series of sodium halides and alkali metal chlorides by using Kirkwood-Buff theory. In order to demonstrate the transferability of the parameters for the alkali halides, we used the same parameters to the study two other systems, KI and CsBr, which were not included in the previous parameterization and for which there are no free parameters. Figure 2.13 and Figure 2.14 clearly suggest that, to a high degree of accuracy, the parameters developed here for the sodium and chloride salts are transferable to other alkali halide salts.

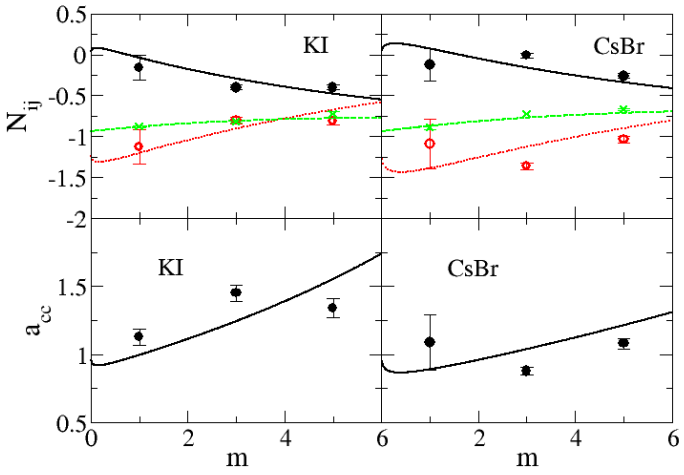


Figure 2.13 Excess coordination numbers as a function of salt molality (top): The N_{cc} (black lines), N_{cw} (red dot lines), and N_{ww} (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density. The N_{cc} (black ●), N_{cw} (red ○), and N_{ww} (green x) are obtained from simulations. Activity derivatives as a function of salt molality (bottom): Lines are obtained from a KB analysis using experimental activity coefficient experimental data and dots correspond to the KBFF models.

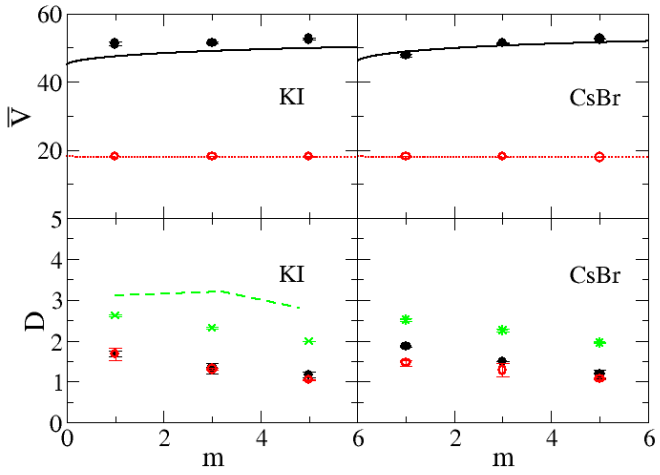


Figure 2.14 Partial molar volumes as a function of salt molality (top): Lines are obtained from a KB analysis using experimental activity coefficient experimental data[29] and dots correspond to the KBFF model. The black lines represent the partial molar volume of salts and the red dot lines indicate partial molar volume of water. The dots (●) represent partial molar volume of salts and the dots (red ○) indicate partial molar volume of water obtained from simulation. Diffusion constants as a function of salt molality (bottom): The D_+ (black lines), D_- (red dot lines), and D_w (green dash lines) are obtained from experimental diffusion constant data and the D_+ (black ●), D_- (red ○), and D_w (green x) are obtained from simulation.

Conclusions

A model for alkali halide aqueous solutions has been developed by reproducing the experimentally derived Kirkwood-Buff integrals using molecular dynamic simulations thereby providing a reasonably accurate representation of the balance between solute-solute and solute-solvent interactions. Other physical and thermodynamic properties such as ion diffusion constants, relative permittivity, the density and heat of mixing have been also reproduced as well. Finally, it has been demonstrated that the parameters developed for sodium and chloride salts are transferable to other type of alkali halide salts by examining the results obtained for KI and CsBr aqueous solutions.

References

1. N. Bentenitis, N. R. Cox and P. E. Smith, J Phys Chem B **113** (36), 12306-12315 (2009).
2. S. Weerasinghe and P. E. Smith, J Chem Phys **119** (21), 11342-11349 (2003).
3. S. Weerasinghe and P. E. Smith, J Phys Chem B **107** (16), 3891-3898 (2003).
4. S. Weerasinghe and P. E. Smith, J Chem Phys **118** (23), 10663-10670 (2003).
5. S. Weerasinghe and P. E. Smith, J Chem Phys **121** (5), 2180-2186 (2004).
6. S. Weerasinghe and P. E. Smith, J Phys Chem B **109** (31), 15080-15086 (2005).
7. M. Kang and P. E. Smith, J Comput Chem **27** (13), 1477-1485 (2006).
8. R. Chitra and P. E. Smith, J Chem Phys **115** (12), 5521-5530 (2001).
9. S. Mclaughlin, Annu Rev Biophys Bio **18**, 113-136 (1989).
10. C. F. Anderson and M. T. Record, Annu Rev Phys Chem **46**, 657-700 (1995).
11. R. L. Baldwin, Biophys J **71** (4), 2056-2063 (1996).
12. I. S. Joung and T. E. Cheatham, J Phys Chem B **112** (30), 9020-9041 (2008).
13. D. Horinek, S. I. Mamatkulov and R. R. Netz, J Chem Phys **130** (12), - (2009).
14. B. Hess and N. F. A. van der Vegt, P Natl Acad Sci USA **106** (32), 13296-13300 (2009).
15. B. Klasczyk and V. Knecht, The Journal of Chemical Physics **132**, 024109 (2010).
16. J. G. Kirkwood and F. P. Buff, J Chem Phys **19** (6), 774-777 (1951).
17. H. J. C. Berendsen, J. R. Grigera and T. P. Straatsma, J Phys Chem-Us **91** (24), 6269-6271 (1987).
18. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, J Comput Chem **26** (16), 1701-1718 (2005).
19. E. Lindahl, B. Hess and D. van der Spoel, J Mol Model **7** (8), 306-317 (2001).
20. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, J Comput Phys **23** (3), 327-341 (1977).

21. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, J Chem Phys **81** (8), 3684-3690 (1984).

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

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

24. P. E. Smith and W. F. Vangunsteren, J Chem Phys **100** (1), 577-585 (1994).

25. R. Walser, A. E. Mark, W. F. van Gunsteren, M. Lauterbach and G. Wipff, J Chem Phys **112** (23), 10450-10459 (2000).

26. R. C. Weast, *CRC Handbook of Chemistry and Physics*, 66 ed. (CRC Press, Inc., Boca Raton, Florida, 1985).

27. Y. Marcus, Chemical Reviews **88** (8), 1475-1498 (1988).

28. S. Ansell, A. C. Barnes, P. E. Mason, G. W. Neilson and S. Ramos, Biophys Chem **124** (3), 171-179 (2006).

29. R. A. Robinson and R. H. Stokes, *Electrolyte solutions*, 2 ed. (London Butter Worths, 1959).

30. O. Sohnel and P. Novotny, *Densities of aqueous solutions of inorganic substances*. (Amsterdam - Oxford - New York - Tokyo, 1985).

31. J. H. Wang and J. W. Kennedy, J Am Chem Soc **72** (5), 2080-2083 (1950).

32. F. Nelson, A. E. Marcinkowsky and K. A. Kraus, *In Research and development progress report/office of saline water*, 302 ed. (1968).

33. H. J. V. Tyrrell and K. R. Harris, *Diffusion in Liquids*. (Butterworths, London, 1984).

34. A. J. Easteal and L. A. Woolf, J Phys Chem-Us **90** (11), 2441-2445 (1986).

35. E. Kumamoto and H. Kimizuka, B Chem Soc Jpn **52** (7), 2145-2146 (1979).

36. P. E. Smith and W. F. Vangunsteren, J Chem Phys **100** (4), 3169-3174 (1994).

37. K. Heger, M. Uematsu and E. U. Franck, Ber Bunsen Phys Chem **84** (8), 758-762 (1980).

38. G. H. Haggis, J. B. Hasted and T. J. Buchanan, J Chem Phys **20** (9), 1452-1465 (1952).

39. F. E. Harris and C. T. Okonski, J Phys Chem-Us **61** (3), 310-319 (1957).

40. R. Buchner, G. T. Hefter and P. M. May, J Phys Chem A **103** (1), 1-9 (1999).

41. G. Beggerow, *in Landolt-Boernstein*. (Springer-Verlag, Berlin, 1976).

CHAPTER 3 - Understanding Amino Acid Interactions in Aqueous Solutions

In an effort to understand the interactions which occur between amino acids in aqueous solutions we have developed new force fields for simple amino acids and their analogs including glycine, *dl*-alanine, betaine, β-alanine, and simple salts: NH_4Cl, NH_4Br, $N(CH_3)_4Cl$, $N(CH_3)_4Br$, CH_3NH_3Cl, and CH_3COONa. The new force fields reproduce the experimental Kirkwood-Buff integrals describing the relative distribution of species in the solution mixture. Furthermore, it can be shown that these simple amino acids can be understood in terms of the interactions of their constituent functional groups and that, to a very good approximation, the transferability and additivity usually assumed in the development of biomolecular force fields appears to hold true.

Introduction

Amino acids are critical to life. Particularly, the most important function of amino acids is their role as the building blocks of proteins, which are linear chains of amino acids. Every protein is chemically defined by this primary structure, i.e. its unique sequence of amino acid residues. Amino acids can be linked together to form various proteins.

The development of accurate force field for amino acids plays an important role in understanding the interactions between amino acids in aqueous solution as provided by molecular dynamic simulations. One of the most important terms in the force field is the electrostatic term, because electrostatic force exists even over very long ranges and constitutes a major part of any intermolecular interactions. The majority of standard molecular simulations use a simple effective point charge model for the electrostatic term. In these models, fixed partial charges are assumed for each component atom. For example, the AMBER[1] force field assigns

fixed partial charges defined by a restricted electrostatic potential fit to the molecular electrostatic potential.[2]

However, the atomic charges should vary depending on the environment and the geometry of the molecule. Thus a simulation with fixed charges describes an average effective polarization and charge-transfer effect, which are involved in the system. This includes biopolymer simulations of proteins and nucleic acid molecules. Attempts have been made to overcome limitations in the fixed charge model by including explicit polarization or charge transfer effects.[3-6] However, these efforts have often provided less satisfying results, and require lots of computational costs as well.

Rappe and Goddard have proposed an approach based on a density functional for estimating the atomic partial charges according to the molecular geometry called the charge equilibration method[3]. Based on this approach, Ogawa and coworkers have developed the consistent charge equilibration method,[7] which uses an identical energy expression for the calculations of both partial charges and the electrostatic energy gradient. The consistent charge equilibration energy term is then combined with a generic force field, the universal force field,[8] to develop the consistent charge equilibration with universal force field.

Typically, other force fields for amino acids have been developed by reproducing small molecule solvation free energies.[9-10] However, by focusing only on solute-solvent interactions, the model may fail to represent the correct distribution of solutes in solutions. These studies cannot guarantee the complicated balance between solute-solute and solute-solvent interactions is reproduced. The description of this delicate balance can play an important role in determining the charge distributions which are the most commonly adjusted parameters during the development of amino acid force fields.[11]

In order to improve the effective charge distributions for molecules, Smith and coworkers have developed new force field which is specially designed to reproduce experimental Kirkwood-Buff integrals.[11-22] It has been designed to accurately describe the delicate balance between solute-solute interactions and solute-solvent interactions. The charges on the atoms are typically adjusted to reproduce the density and KB integrals for solution mixtures at several different compositions using the Kirkwood-Buff theory of solutions as a guide. Even though a Kirkwood-Buff derived force field is not a polarizable force field the best effective charge on each atom can be determined by using the sensitivity of KBIs to the molecular charge distribution.[11, 20-21] It has been shown that KBFF perform simulation better than many other non-polarizable force fields with the same computational cost.

The KB integrals can play an important role in the parameterization of a new force field because KBIs are more sensitive to the parameter sets than many other experimental data.[11, 22-24] In addition, it can help us quantify the interaction between a pair of components in solution. The quality of a force field used in simulation can be determined by comparing the KBIs or thermodynamic properties obtained from simulations to the KBIs or thermodynamic properties obtained from experimental data. Here, a KB analysis of the properties of several amino acids and electrolytes. Typical building blocks of proteins in aqueous solutions as a function of molality is used to develop a force field for the description of amino acid solutions using the extended simple point charge (SPC/E) water model.[25]

Methods

Kirkwood-Buff Theory

The application of Kirkwood-Buff theory follows the same outline as presented in Chapter 2.

Molecular dynamics simulations

All molecular dynamic simulations of ammonium salts, sodium acetate, and amino acid aqueous solutions were performed using the SPC/E water model[25] in the isothermal isobaric (*NpT*) ensemble at 300 K and 1 atm as implemented in the GROMACS program (v3.3.1).[26-27] A time-step of 2 fs was used and the geometry of the water molecules was constrained using SETTLE.[28] All bonds were constrained using SHAKE[29] for the salts and LINCS[30] for the amino acids. A twin range cut-off of 0.8 nm/1.5 nm was employed with a nonbonded pair list update of every 10 steps. The weak coupling technique was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively.[31] In order to evaluate electrostatic interactions, the particle mesh Ewald technique (PME) was used.[32] The initial cubic boxes of different solutions have been generated by adding water molecules and molecules/ions until the required concentration was obtained. Configurations were saved every 0.1 ps for analysis. Diffusion constants were determined using the mean square fluctuation approach,[33] and relative permittivities were obtained from the dipole moment fluctuations.[34]

Parameter Development

We need to develop parameters for the N-terminal, C-terminal, and peptide bond groups in their zwitterionic form. For the development of parameter we need models for each group. Our strategy is shown in Figure 3.1. For the peptide bonds, Smith and his coworker have released the parameters of acetamide.[16]

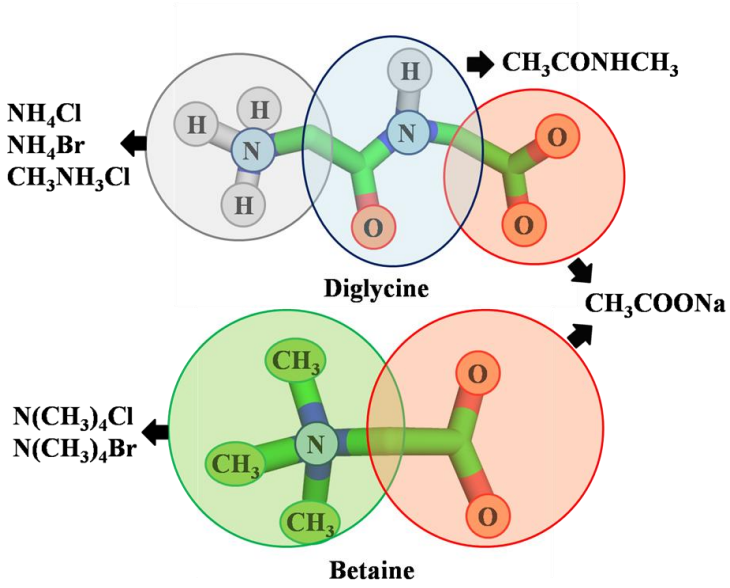


Figure 3.1 The strategy for the development of force field parameters. NH_4Cl, NH_4Br, and CH_3NH_3Cl are analogues for the N-terminus, while CH_3COONa corresponds to a model for the C-terminus. The $CH_3CONHCH_3$ molecule describes the peptide group and has been developed previously by Smith and Kang.[16]

The nonbonding force field used in this study corresponds to the Lennard-Jones (LJ) 6-12 potential which contains most commonly two adjustable parameters: the Lennard-Jones diameter and the interaction strength plus Coulomb potential with the SPC/E water model.[25] In this scheme each pair of atoms *i* and *j* interact with an interaction energy given by

$$V_{ij} = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} + 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \quad (3.1)$$

Here, all the symbols have their usual meaning.[6] Hence, we need three parameters for each atom; σ, ε, and q. The σ and ε have been obtained elsewhere.[26-27] However, the charge distribution (q) for each molecule has to be determined. Hence, we have focused on the charge distribution of each atom for our molecules during the parameterization.

Table 3.1 and table 3.2 show the final Lennard-Jones parameters used in our simulations. The LJ parameters for ions; Na^+, Cl^-, and Br^- were taken from Weerasinghe and Smith,[22] and our previous research (Chapter 2). The charge distribution was optimized by initial charge distribution for each atom in the molecule based on the polarity and the electro negativity. Then the final parameters have been refined by increasing or decreasing the charge of each atom until the simulated value of KBIs reproduced the experimental KBIs for each molecule in aqueous solution.

The best charge distribution for some of the salts gave us unexpected results. For the tetrahedral ions, NH_4^+ and $(CH_3)_4N^+$ the total charge (+1) is distributed to five component atoms evenly, while the charge for $CH_3NH_3^+$ is very different. We believe this is due to the different solvent environments around these ions, which gives rise to different polarization effects. For the charge distribution of $CH_3CO_2^-$, we found the best agreement with a zero charge on the CH_3 group.

We then attempted to combine the charge distributions of methyl ammonium and acetate ions to determine the charge distribution for amino acids. Unfortunately, this did not work. It appears we cannot apply the same charge of these salts to amino acids because the charge of CH_3 in CH_3CO_2 is not zero, and the methyl ammonium CH_3 charge is 0.5. In the case of betaine and

β-alanine, the charge of the original salts can be used for amino acid without any modification combining, while the C-terminal charge of glycine and *dl*-alanine needed to be made more polar. It is reasonable to expect some charge transfer from neighbor atoms, i.e. CH_3 (H_3NCH_3) to C (CO_2). The bonding force field used in this study corresponds to GROMOS96 data, and is included in Table 3.3 to Table 3.7.

Table 3.1 Final nonbonded parameters for ammonium salt and sodium acetate aqueous solutions for the KBFF model

Model	Salts	Atom	σ (nm)	ε (kJ/mol)	q (e)
KBFF	NH_4Cl	N	0.3370	0.5620	+0.2
		H	0.1580	0.0880	+0.2
		Cl	0.4400	0.4700	-1.0
	NH_4Br	N	0.3370	0.5620	+0.2
		H	0.1580	0.0880	+0.2
		Br	0.4760	0.3000	-1.0
	$(CH_3)_4NCl$	N	0.3370	0.5620	+0.2
		CH_3	0.3748	0.8672	+0.2
		Cl	0.4400	0.4700	-1.0
	$(CH_3)_4NBr$	N	0.3370	0.5620	+0.2
		CH_3	0.3748	0.8672	+0.2
		Br	0.4760	0.3000	-1.0
	CH_3NH_3Cl	N	0.3370	0.5620	+0.5
		CH_3	0.3748	0.8672	+0.5
		H	0.1580	0.0880	0.0
		Cl	0.4400	0.4700	-1.0
	CH_3CO_2Na	CH_3	0.3748	0.8672	0.0
		C	0.3360	0.3300	+0.3
		O	0.3500	0.6047	-0.65
		Na	0.2450	0.3200	+1.0
SPC/E	H_2O	O^{25}	0.3166	0.6506	-0.8476
		H^{25}	0.0000	0.0000	+0.4238

Table 3.2 Final nonbonded parameters for amino acid aqueous solutions for the KBFF model

Model	Molecules	Atom	σ (nm)	ε (kJ/mol)	q (e)
KBFF	Glycine	H	0.1580	0.0880	0.0
		N	0.3370	0.5620	+0.5
		CH_2	0.4070	0.4105	+0.5
		C	0.3360	0.3300	+1.0
		O	0.3500	0.6047	-1.0
	dl-Alanine	H	0.1580	0.0880	0.0
		N	0.3370	0.5620	+0.5
		CH_1	0.5019	0.09489	+0.5
		CH_3	0.3748	0.8672	0.0
		C	0.3360	0.3300	+1.0
		O	0.3500	0.6047	-1.0
	Betaine	CH_3	0.3748	0.8672	+0.2
		N	0.3370	0.5620	+0.2
		CH_2	0.4070	0.4105	+0.2
		C	0.3360	0.3300	+0.3
		O	0.3500	0.6047	-0.65
	β-Alanine	H	0.1580	0.0880	0.0
		N	0.3370	0.5620	+0.5
		CH_2	0.4070	0.4105	+0.5
		CH_2	0.4070	0.4105	+0.0
		C	0.3360	0.3300	+0.3
		O	0.3500	0.6047	-0.65
SPC/E	H_2O	O^{25}	0.3166	0.6506	-0.8476
		H^{25}	0.0000	0.0000	+0.4238

Table 3.3 Final bonded parameters for ammonium salt and sodium acetate aqueous solutions for the KBFF model: Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Model	Salts	Atom	σ (nm)	ε (kJ/mol)	q (e)
KBFF	NH_4Cl	N	0.3370	0.5620	+0.2
		H	0.1580	0.0880	+0.2
		Cl	0.4400	0.4700	-1.0
	NH_4Br	N	0.3370	0.5620	+0.2
		H	0.1580	0.0880	+0.2
		Br	0.4760	0.3000	-1.0
	$(CH_3)_4NCl$	N	0.3370	0.5620	+0.2
		CH_3	0.3748	0.8672	+0.2
		Cl	0.4400	0.4700	-1.0
	$(CH_3)_4NBr$	N	0.3370	0.5620	+0.2
		CH_3	0.3748	0.8672	+0.2
		Br	0.4760	0.3000	-1.0
	CH_3NH_3Cl	N	0.3370	0.5620	+0.5
		CH_3	0.3748	0.8672	+0.5
		H	0.1580	0.0880	0.0
		Cl	0.4400	0.4700	-1.0
	CH_3CO_2Na	CH_3	0.3748	0.8672	0.0
		C	0.3360	0.3300	+0.3
		O	0.3500	0.6047	-0.65
		Na	0.2450	0.3200	+1.0
SPC/E	H_2O	O^{25}	0.3166	0.6506	-0.8476
		H^{25}	0.0000	0.0000	+0.4238

Table 3.4 Final bonded parameters for aqueous glycine solutions for the KBFF model: Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Model	Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
Glycine	H-N	0.1000	Constraint	
	N-CH_2	0.1468		
	CH_2-C	0.1520		
	C-O	0.1250		
	Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
	H-N-H	109.5	334.72	
	H-N-CH_2	109.5	376.56	
	N-CH_2-C	114.1	502.10	
	CH_2-C-O	117.0	502.08	
	O-C-O	126.0	502.08	
	Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
	H-N-CH_2-C	0.0	4.0002	3
	N-CH_2-C-O	0.0	1.0002	6
	Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
	C-CH_2-O-O	0.0	167.36	

Table 3.5 Final bonded parameters for aqueous *dl*-alanine solutions for the KBFF model: Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Model	Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
dl-Alanine	H-N	0.1000	Constraint	
	$N-CH_1$	0.1468		
	CH_1-CH_3	0.1530		
	CH_1-C	0.1520		
	C-O	0.1250		
	Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
	H-N-H	109.5	334.72	
	H-N-CH_1	109.5	376.56	
	N-CH_1-C	114.1	502.10	
	N-CH_1-CH_3	109.5	376.56	
	CH_3-CH_1-C	109.5	376.56	
	CH_1-C-O	117.0	502.08	
	O-C-O	126.0	502.08	
	Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
	H-N-CH_2-C	0.0	4.0002	3
	N-CH_2-C-O	0.0	1.0002	6
	Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
	C-CH_1-O-O	0.0	167.36	
	CH_1-N-CH_3-C	35.26439	334.72	

Table 3.6 Final bonded parameters for aqueous betaine solutions for the KBFF model: Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Model	Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
Betaine	CH_3-N	0.1470	Constraint	
	N-CH_2	0.1468		
	CH_2-C	0.1520		
	C-O	0.1250		
	Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
	CH_3-N- CH_3	109.5	334.72	
	CH_3-N-CH_2	109.5	376.56	
	N-CH_2-C	114.1	502.10	
	CH_2-C-O	117.0	502.08	
	O-C-O	126.0	502.08	
	Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
	CH_3-N-CH_2-C	0.0	4.0002	3
	N-CH_2-C-O	0.0	1.0002	6
	Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
	C-CH_2-O-O	0.0	167.36	

Table 3.7 Final bonded parameters for aqueous β-Alanine solutions for the KBFF model: Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Model	Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
β-Alanine	H-N	0.1000	Constraint	
	N-CH_2	0.1468		
	CH_2-CH_2	0.1530		
	CH_2-C	0.1520		
	C-O	0.1250		
	Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
	H-N-H	109.5	334.72	
	H-N-CH_2	109.5	376.56	
	N-CH_2-CH_2	111.0	460.24	
	CH_2-CH_2-C	111.0	460.24	
	CH_2-C-O	117.0	502.08	
	O-C-O	126.0	502.08	
	Proper Dihedrals	δ (deg)	k_φ (kJ mol^{-1})	n
	H-N-CH_2-CH_2	0.0	4.0000	3
	N-CH_2-CH_2-C	0.0	6.0000	3
	CH_2-CH_2-C-O	0.0	6.0000	6
	Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
	C-CH_2-O-O	0.0	167.36	

Results

Table 3.8 and Table 3.9 describe the number of ions or solutes and the number of water molecules used in our solution simulations, as well as the molality, and molality of the resulting solutions. Also in the same table, the simulation time, volume, density, and potential energy are shown. For all solutions, the density increases and the potential energy decreases as the molality increases, as expected.

Table 3.8 Summary of the MD simulations of aqueous salt solutions: All simulations were performed at 300 K and 1 atm in the *NpT* ensemble. Symbols are N_w, number of water molecules; N_s (= $N_+ = N_- = 1/2Nc$), number of salts pairs; V, average simulation volume; m_s, salt molality; C_s, salt molarity; ρ, mass density; E_{pot}, average total potential energy per molecule ($N_s + N_w$); and T_{sim}, total simulation time.

	N_s	N_w	m_s (mol/Kg)	V (nm^3)	C_s (mol/l)	ρ (g/cm^3)	E_{pot} (kJ/mol)	T_{sim} (ns)
H_2O	0	2170	0.00	65.265	0.00	0.995	-46.45	2
NH_4Cl	125	6936	1	215.270	0.96	1.015	-57.80	11
	349	6458	3	214.703	2.70	1.044	-79.06	11
	630	5832	6	215.317	4.86	1.070	-108.32	11
NH_4Br	125	6924	1	216.547	0.96	1.050	-57.37	11
	347	6421	3	217.725	2.65	1.141	-77.74	11
	621	5749	6	219.365	4.70	1.244	-105.82	11
$(CH_3)_4NCl$	125	6924	1	214.072	0.97	1.008	-53.01	11
	349	6458	3	211.468	2.74	1.026	-63.56	11
	424	4706	5	210.126	3.35	1.037	-92.66	11
$(CH_3)_4NBr$	116	6457	1	214.165	0.90	1.041	-55.99	11
	290	5358	3	211.475	2.28	1.109	-73.94	11
	413	4578	5	209.831	3.27	1.156	-90.56	11
CH_3NH_3Cl	122	6790	1	214.929	0.94	1.009	-56.77	11
	328	6069	3	212.751	2.56	1.026	-76.14	11
	495	5494	5	212.360	3.87	1.035	-94.17	11
CH_3COONa	124	6907	1	215.345	0.96	1.038	-59.85	11
	344	6368	3	215.117	2.66	1.103	-84.86	11
	530	5888	5	216.025	4.08	1.150	-108.12	11

Table 3.9 Summary of the MD simulations of amino acid water mixtures. All simulations were performed at 300 K and 1atm in the *NpT* ensemble. Symbols are N_w, number of water molecules; N_s, number of amino acids; V, average simulation volume; m_s, amino acid molality; C_s, amino acid molarity; ρ, mass density; E_{pot}, average total potential energy per molecule ($N_s + N_w$); and T_{sim}, total simulation time

	N_s	N_w	m_s (mol/Kg)	V (nm^3)	C_s (mol/l)	ρ (g/cm^3)	E_{pot} (kJ/mol)	T_{sim} (ns)
H_2O	0	2170	0.00	65.265	0.00	0.995	-46.45	2
Glycine	124	6898	1	215.793	0.95	1.028	-58.22	11
	238	6611	2	215.177	1.84	1.057	-69.50	11
	343	6341	3	214.817	2.65	1.082	-80.45	11
dl-Alanine	62	6988	0.5	216.245	0.48	1.009	-52.12	16
	122	6795	1.0	216.469	0.94	1.022	-57.68	16
	178	6592	1.5	216.061	1.37	1.035	-63.13	16
Betaine	118	6553	1	215.589	0.91	1.016	-54.59	16
	300	5556	3	214.520	2.32	1.047	-69.78	16
	435	4830	5	214.368	3.37	1.069	-84.00	16
β-Alanine	122	6795	1	217.462	0.93	1.018	-57.57	11
	329	6096	3	219.458	2.49	1.052	-78.54	11
	496	5505	5	220.758	3.73	1.078	-98.20	11

The radial distribution functions (rdfs) obtained from the simulations are shown on Figure 3.2 ~ 3.7 for the salts in water. Figure 3.8 and 3.9 represent rdfs for the amino acid aqueous solutions. All rdfs approach unity beyond 1.5 nm.

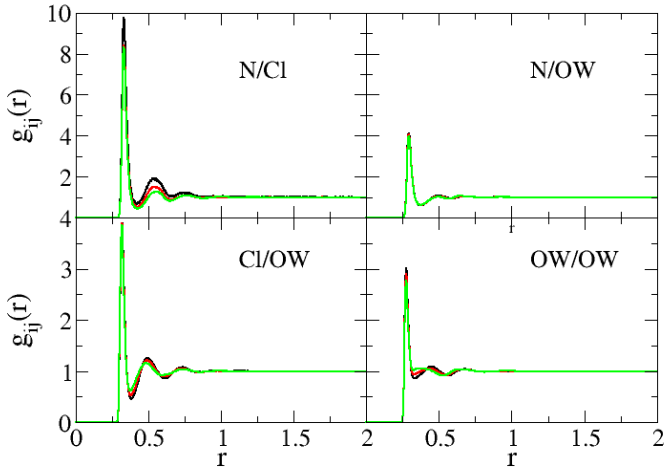


Figure 3.2 Radial distribution functions of NH_4Cl obtained from the 1 m (black lines), 3 m (red lines), and 6 m (green lines) simulations. Nitrogen of NH_4^+, Cl^-, and the water oxygen are denoted by the symbols N, Cl, and OW, respectively.

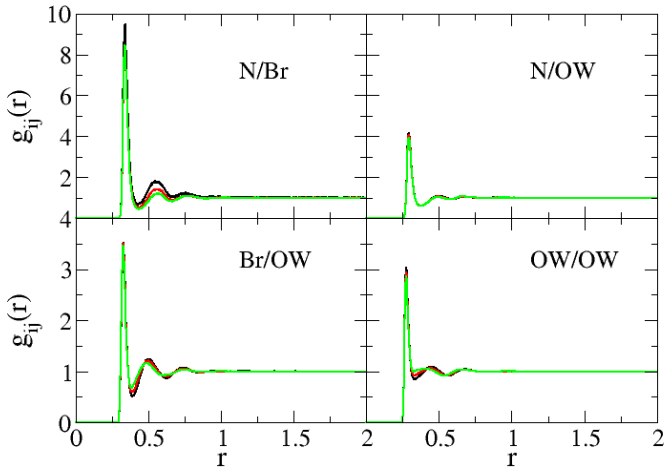


Figure 3.3 Radial distribution functions of NH_4Br obtained from the 1 m (black lines), 3 m (red lines), and 6 m (green lines) simulations. Nitrogen of NH_4^+, Br^-, and the water oxygen are denoted by the symbols N, Br, and OW, respectively.

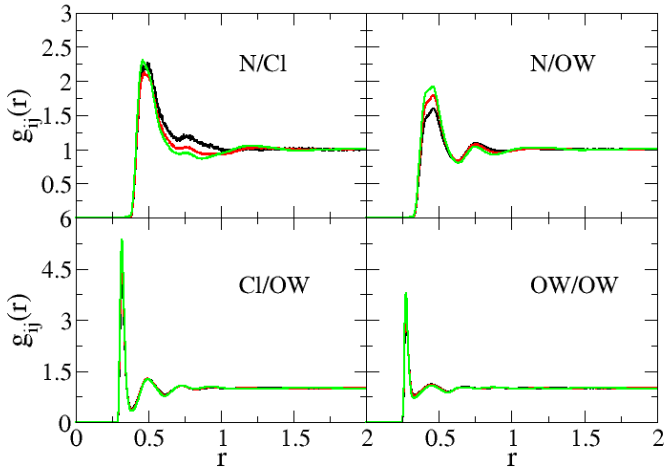


Figure 3.4 Radial distribution functions of $(CH_3)_4NCl$ obtained from the 1 m (black lines), 3 m (red lines), and 5 m (green lines) simulations. Nitrogen of $(CH_3)_4N^+$, Cl^-, and the water oxygen are denoted by the symbols N, Cl, and OW, respectively.

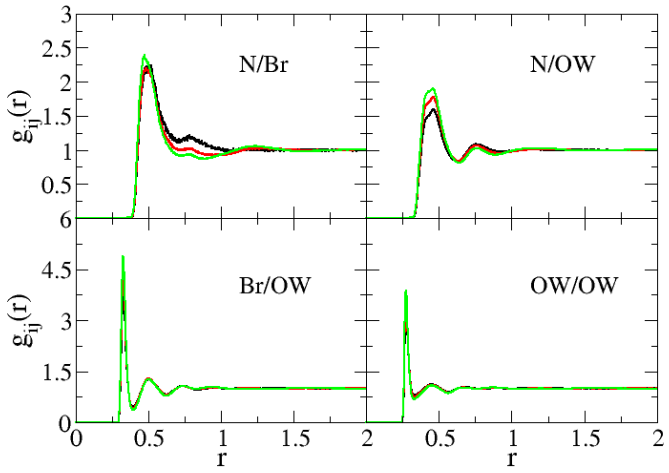


Figure 3.5 Radial distribution functions of $(CH_3)_4NBr$ obtained from the 1 m (black lines), 3 m (red lines), and 5 m (green lines) simulations. Nitrogen of $(CH_3)_4N^+$, Br^-, and the water oxygen are denoted by the symbols N, Br, and OW, respectively

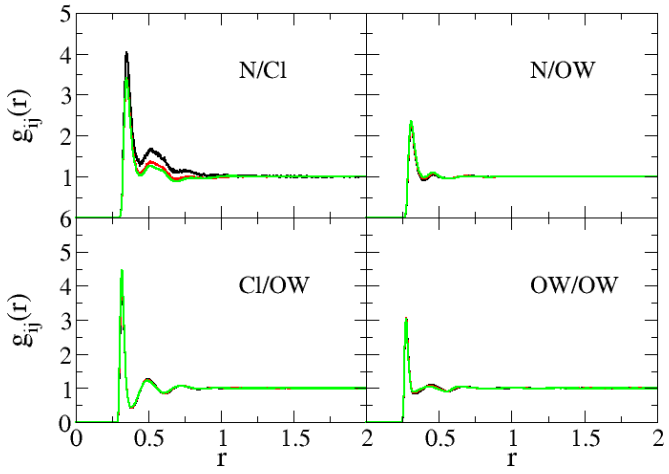


Figure 3.6 Radial distribution functions of CH_3NH_3Cl obtained from the 1 m (black lines), 3 m (red lines), and 5 m (green lines) simulations. Nitrogen of $CH_3NH_3^+$, Cl^-, and the water oxygen are denoted by the symbols N, Cl, and OW, respectively.

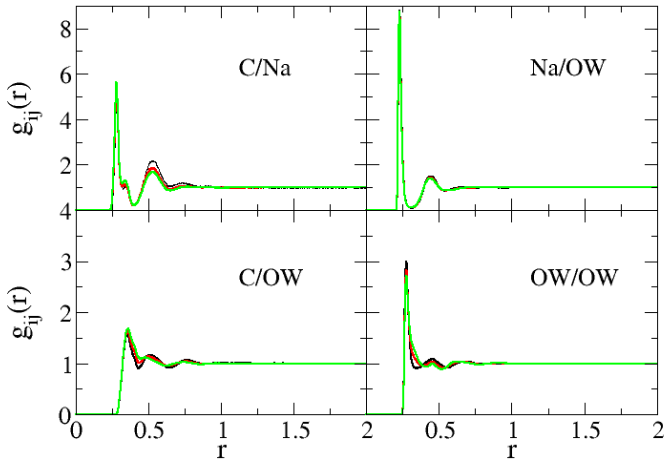


Figure 3.7 Radial distribution functions of CH_3COONa obtained from the 1 m (black lines), 3 m (red lines), and 5 m (green lines) simulations. Carbon of CH_3COO^-, Na^+, and the water oxygen are denoted by the symbols C, Na, and OW, respectively.

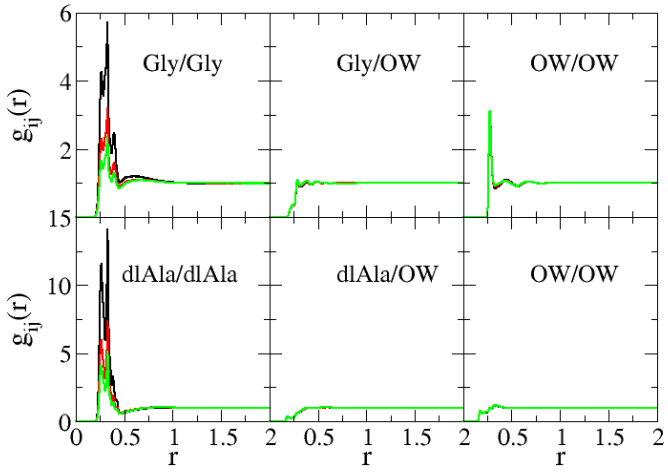


Figure 3.8 Radial distribution functions for aqueous glycine (top) obtained from the 1 m (black lines), 2 m (red lines), and 3 m (green lines) simulations, and aqueous *dl*-alanine (bottom) obtained from the 0.5 m (black lines), 1.0 m (red lines), and 1.5 m (green lines) simulations: Center of mass for glycine and *dl*-alanine are denoted by Gly and dlAla. Water oxygens are denoted by the symbols OW.

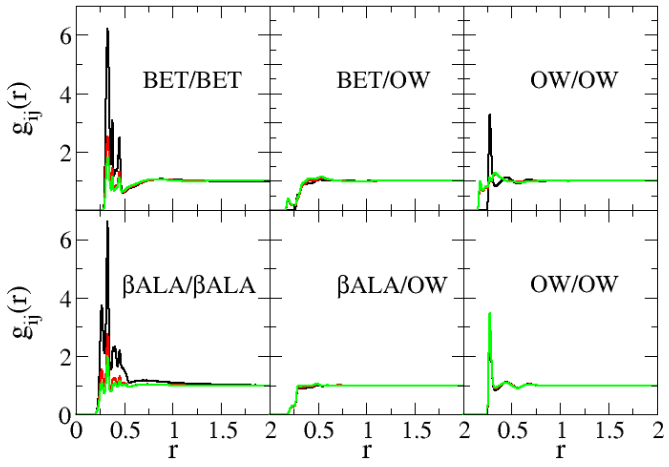


Figure 3.9 Radial distribution functions for aqueous betaine (top) obtained from the 1 m (black lines), 3 m (red lines), and 5 m (green lines) simulations, and aqueous *β*-alanine (bottom) obtained from the 1 m (black lines), 2 m (red lines), and 3 m (green lines) simulations: Center of mass for betaine and *β*-alanine are denoted by BET and βALA. Water oxygens are denoted by the symbols OW.

The first shell coordination numbers, n_{ij}, as well as the distances to the first rdf maximum, R_{max}, and the first rdf minimum, R_{min}, were calculated from the corresponding rdfs as a function of the solution molality and are presented in Table 3.10. As the size of the cation increases, so does the radius of the first hydration shell; the same trend is exhibited by the anions which have been shown in the previous chapter (chapter 2). Similar to the trend in the radii of the first hydration shell, the hydration numbers increase as the size of the cation increases. Table 3.10 shows that coordination numbers are not only sensitive to the size of the salt ion, but also the salt concentration.

Table 3.10 First shell coordination numbers (n_{ij}) for aqueous salt solutions. R_{max} and R_{min} are the distances (nm) to the first maximum and minimum of the radial distribution functions. Cations, anions, and the water oxygen are denoted by the symbols +, -, and o, respectively.

		m_s	+/-	+/o	-/o	o/o
	R_{max}		0.33	0.29	0.32	0.27
	R_{min}		0.42	0.37	0.38	0.33
NH_4Cl	n_{ij}	1	0.36	7.29	6.76	4.16
		3	0.85	6.72	6.32	3.84
		6	1.48	5.96	5.69	3.42
	R_{max}		0.33	0.29	0.33	0.27
	R_{min}		0.42	0.37	0.39	0.33
NH_4Br	n_{ij}	1	0.37	7.22	6.83	4.14
		3	0.88	6.56	6.35	3.80
		6	1.55	5.67	5.64	3.38
	R_{max}		0.49	0.45	0.32	0.27
	R_{min}		0.69	0.62	0.38	0.33
$(CH_3)_4NCl$	n_{ij}	1	0.95	28.67	6.76	4.07
		3	2.22	26.49	6.32	3.60
		5	3.20	24.08	5.90	3.18
	R_{max}		0.49	0.45	0.33	0.27
	R_{min}		0.69	0.62	0.39	0.33
$(CH_3)_4NBr$	n_{ij}	1	0.91	28.48	6.83	4.06
		3	2.20	26.05	6.34	3.57
		5	3.15	23.54	5.89	3.14
	R_{max}		0.35	0.31	0.32	0.27
	R_{min}		0.44	0.40	0.38	0.33
CH_3NH_3Cl	n_{ij}	1	0.29	7.48	6.77	4.12
		3	0.65	7.01	6.38	3.72

		5	0.97	6.58	6.00	3.34
	R_{max}		0.27	0.22	0.35	0.27
	R_{min}		0.40	0.31	0.43	0.33
CH_3COONa	n_{ij}	1	0.16	5.36	8.20	4.25
		3	0.45	4.94	8.12	4.05
		5	0.75	4.51	7.90	3.84

The simulated and experimental excess coordination numbers, N_{ij}, are shown in Figure 3.10 for salt solution and in Figure 3.11 for amino acids aqueous solution as a function of the molality. Figure 3.12 and Figure 3.13 show the simulated and experimental activity derivatives, a_{cc}, as a function of molality. The KBFF models quantitatively reproduce the experimental data, although the simulated values are not as accurate for tetramethyl ammonium chloride $(CH_3)_4NCl$ and tetramethyl ammonium bromide $(CH_3)_4NBr$ systems. The water-water excess coordination number, N_{ww}, represented by green lines and symbols on Figure 3.10 and Figure 3.11, remains relatively constant at approximately -0.95 and is independent of the molality of the various solutes, except for $(CH_3)_4NCl$ and $(CH_3)_4NBr$, suggesting that ion aggregation is small at the molecular level. Figures 3.12 and Figure 3.13 show that the solute-solute excess coordination numbers (black lines) also do not vary significantly, at least compared to the variation of the solute-water excess coordination numbers (red lines). This indicates that the solute-water interactions again determine the solution behavior. This suggests that in diluted solution, the small activity derivative for the solutes such as $(CH_3)_4NCl$, $(CH_3)_4NBr$, and betaine which include the $(CH_3)_4N$ group, shown in Figure 3.12 and Figure 3.13, is due to the presence of more hydrophobic groups, compared to that of the other solutes. In concentrated solution, the simulation data of activity derivatives are slightly overestimated compared to other solution composition for $(CH_3)_4NCl$, $(CH_3)_4NBr$, and betaine which include the $(CH_3)_4N$ group. Figure

3.9 and Figure 3.10 also indicate that CH_3NH_3Cl and *β*-alanine solutions display smaller solute-water interaction in dilute solutions due to the additional hydrophobicity, compared to NH_4Cl, NH_4Br, glycine and *dl*-alanine, and a larger solute-water interaction in dilute solutions compared to $(CH_3)_4NCl$, $(CH_3)_4NBr$, and betaine, which is as expected. For CH_3COONa solutions, simulation results reproduce the experimental data well. Figure 3.10 show that the N_{ij}s of CH_3COONa display similar trends to NH_4Cl and NH_4Br.

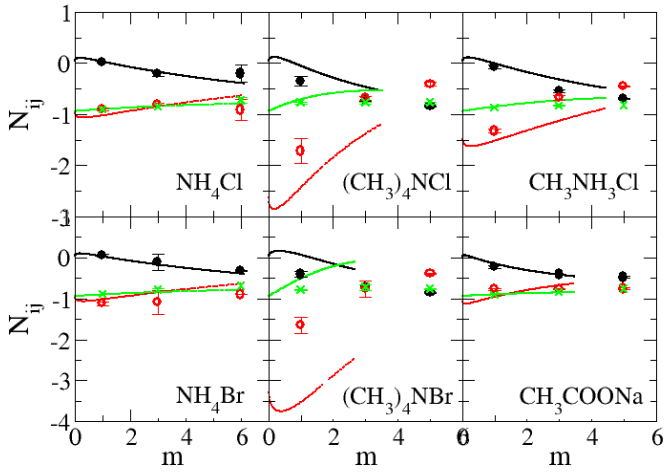


Figure 3.10 Excess coordination numbers as a function of salt molality. The N_{cc} (black lines), N_{cw} (red dot lines), and N_{ww} (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density.[35-37] The N_{cc} (black ●), N_{cw} (red ○), and N_{ww} (green x) are obtained from simulations.

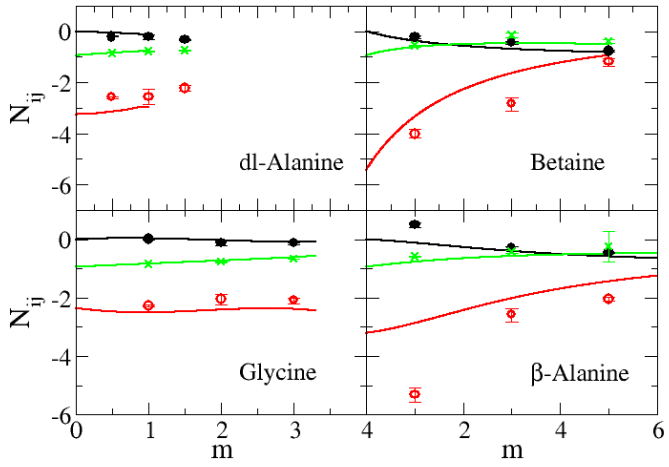


Figure 3.11 Excess coordination numbers as a function of solute molality. The N_{cc} (black lines), N_{cw} (red dot lines), and N_{ww} (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density.[38-39] The N_{cc} (black ●), N_{cw} (red ○), and N_{ww} (green x) are obtained from simulations.

In Figure 3.12 and Figure 3.13 the simulated activity derivatives a_{cc} as a function of molality are compared to the experimental values. The KBFF model reproduced the correct increase in a_{cc} with concentration as displayed by the experimental data.

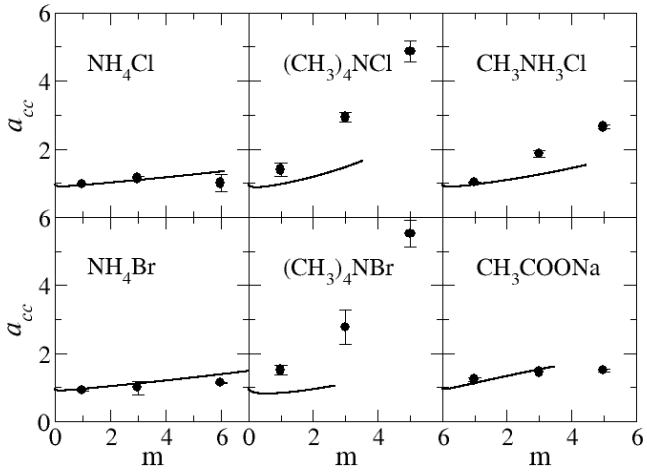


Figure 3.12 Activity derivatives as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient experimental data,[35, 37] and dots correspond to the KBFF model.

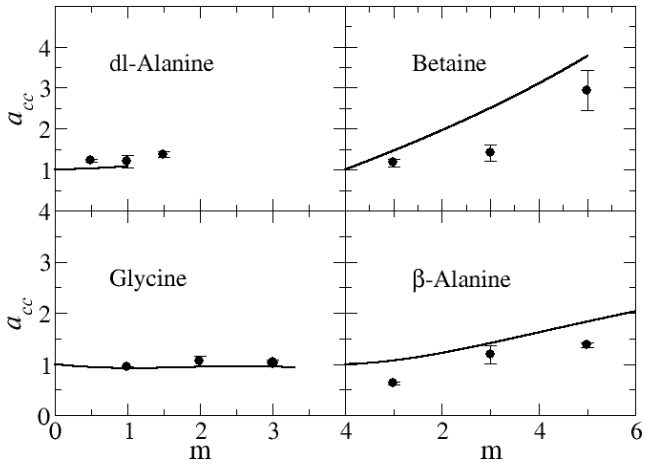


Figure 3.13 Activity derivatives as a function of solute molality. Lines are obtained from a KB analysis using experimental activity coefficient experimental data,[38-39] and dots correspond to the KBFF model.

Figure 3.14 and Figure 3.15 show the partial molar volumes of water and of the salts. The partial molar volume of the salts increases monotonically, and that of water decreases monotonically as the salt concentration increases. Also, as the size of the salts increases the partial molar volume of the salt concentration increases. The KBFF models reproduce the experimental data quantitatively apart from the case of $(CH_3)_4NBr$, presumably due to an overestimation of the cation/anion attraction.

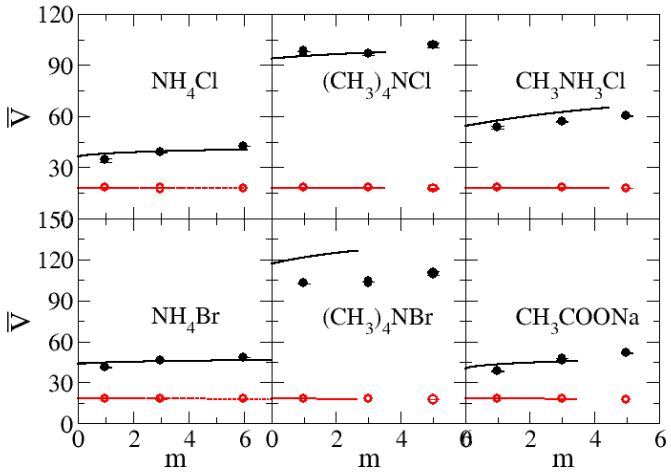


Figure 3.14 Partial molar volumes as a function of salt molality. Lines are obtained from a KB analysis using experimental activity coefficient and experimental density data,[35, 40] and dots correspond to the KBFF model. The black lines represent the partial molar volume of salts and the red dotted lines indicate partial molar volume of water. The dots (●) represent partial molar volume of salts and the dots (red ○) indicate partial molar volume of water.

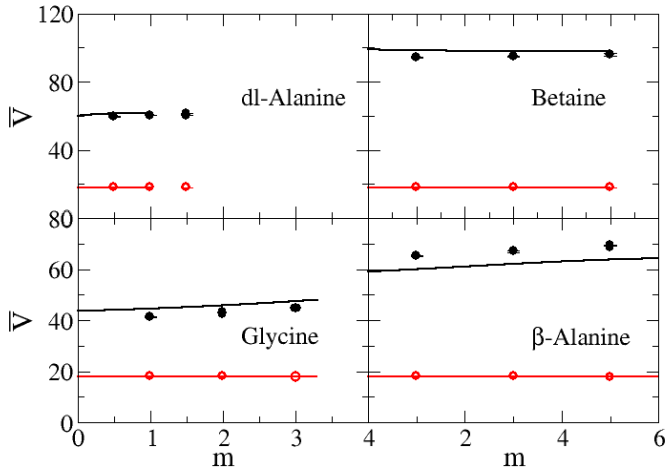


Figure 3.15 Partial molar volumes as a function of solute molality. Lines are obtained from a KB analysis using experimental activity coefficient and experimental density data,[35, 40] and dots correspond to the KBFF model. The black lines represent the partial molar volume of solutes and the red dotted lines indicate partial molar volume of water. The dots (●) represent partial molar volume of solutes and the dots (red ○) indicate partial molar volume of water.

The self-diffusion constants, calculated using the mean square fluctuation approach,[33] are displayed in Figure 3.16 and Figure3.17 as a function of salt or solute molality. They all exhibit an essentially linear decrease with molality. The KBFF appears to reproduce the experimental data very well.

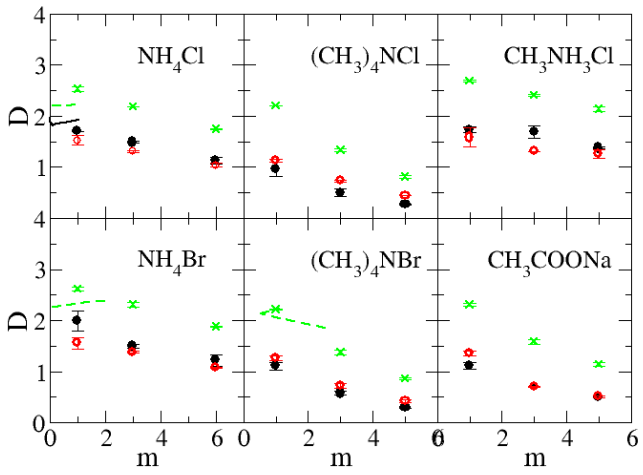


Figure 3.16 Diffusion constants as a function of salt molality. The D_c (black lines) is obtained from experimental diffusion constant data[41] and the D_+ (black ●), D_- (red ○), and D_w (green x) are obtained from simulations.

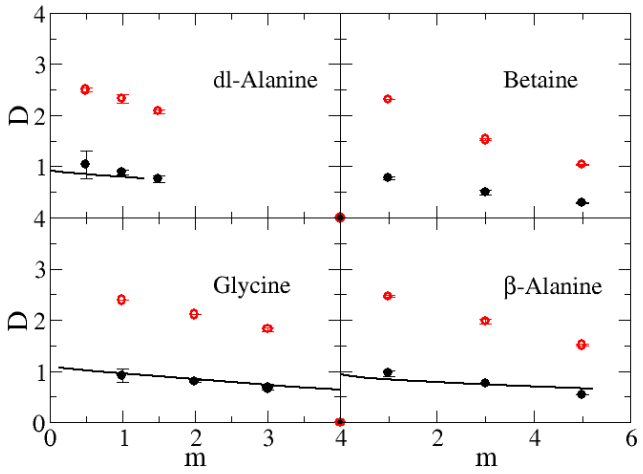


Figure 3.17 Diffusion constants as a function of solute molality. The D_c (black lines) is obtained from experimental diffusion constant data[41] and the D_c (black ●) and D_w (red ○) are obtained from simulations.

The relative permittivity of the salts and amino acids, calculated from the dipole moment fluctuations,[34] are displayed in Figure 3.18 and Figure 3.19. They all exhibit an essentially linear variation. The relative permittivities for salt solutions decrease as a function of molality, but for amino acid aqueous solutions they increase as a function of molality. Unfortunately, we could not find experimental data for the salt and betaine aqueous solution. The KBFF model reproduces the experimental data well.

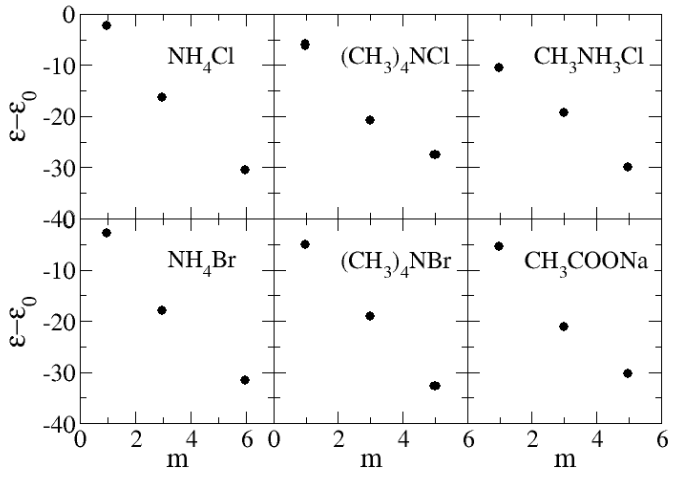


Figure 3.18 Simulated relative permittivities as a function of salt molality.

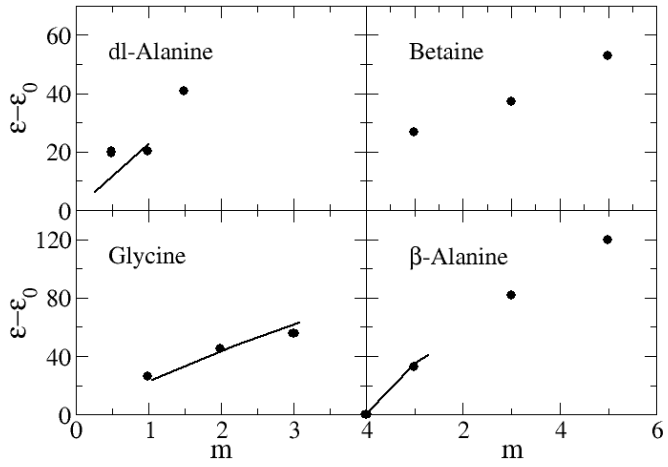


Figure 3.19 Relative permittivities as a function of solute molality. Lines are obtained from experimental dielectric constant data,[42-43] and dots obtained from simulations.

Conclusions

The KBFF models developed here have reproduced the majority of the experimentally derived Kirkwood-Buff integrals for amino acid solutions and small molecules which mimic amino acid fragments in water. The force field parameters for amino acids have been developed from parameters of small molecules which mimic each amino acid fragments. In other words, fragment additivity appears to be observed.

References

1. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, J Am Chem Soc **117** (19), 5179-5197 (1995).
2. P. Cieplak, W. D. Cornell, C. Bayly and P. A. Kollman, J Comput Chem **16** (11), 1357-1377 (1995).
3. A. K. Rappe and W. A. Goddard, J Phys Chem-Us **95** (8), 3358-3363 (1991).
4. S. W. Rick, S. J. Stuart and B. J. Berne, J Chem Phys **101** (7), 6141-6156 (1994).
5. T. A. Halgren and W. Damm, Curr Opin Struc Biol **11** (2), 236-242 (2001).
6. A. Laio, J. VandeVondele and U. Rothlisberger, J Phys Chem B **106** (29), 7300-7307 (2002).
7. O. Kitao and T. Ogawa, Mol Phys **101** (1-2), 3-17 (2003).
8. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, J Am Chem Soc **114** (25), 10024-10035 (1992).
9. Z. T. Xu, H. H. Luo and D. P. Tieleman, J Comput Chem **28** (3), 689-697 (2007).
10. J. Chang, A. M. Lenhoff and S. I. Sandler, J Phys Chem B **111** (8), 2098-2106 (2007).
11. S. Weerasinghe and P. E. Smith, J Chem Phys **118** (13), 5901-5910 (2003).
12. N. Bentenitis, N. R. Cox and P. E. Smith, J Phys Chem B **113** (36), 12306-12315 (2009).
13. V. Pierce, M. Kang, M. Aburi, S. Weerasinghe and P. E. Smith, Cell Biochem Biophys **50** (1), 1-22 (2008).
14. F. Chen and P. E. Smith, J Phys Chem B **112** (30), 8975-8984 (2008).
15. M. Kang and P. E. Smith, Fluid Phase Equilibr **256** (1-2), 14-19 (2007).
16. M. Kang and P. E. Smith, J Comput Chem **27** (13), 1477-1485 (2006).

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

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

19. P. E. Smith, J Phys Chem B **108** (48), 18716-18724 (2004).

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

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

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

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

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

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

26. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, J Comput Chem **26** (16), 1701-1718 (2005).

27. E. Lindahl, B. Hess and D. van der Spoel, J Mol Model **7** (8), 306-317 (2001).

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

29. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, J Comput Phys **23** (3), 327-341 (1977).

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

31. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, J Chem Phys **81** (8), 3684-3690 (1984).

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

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

34. P. E. Smith and W. F. Vangunsteren, J Chem Phys **100** (1), 577-585 (1994).

35. R. A. Robinson and R. H. Stokes, *Electrolyte solutions*, 2 ed. (London Butter Worths, 1959).

36. B. E. Conway, R. E. Verrall and Desnoyer.Je, T Faraday Soc **62** (526P), 2738-& (1966).

37. S. Lindenbaum and G. E. Boyd, J Phys Chem-Us **68** (4), 911-& (1964).

38. E. N. Tsurko, R. Neueder and W. Kunz, J Solution Chem **36** (5), 651-672 (2007).

39. F. Otero-Ravina, R. Romero, A. Rodriguez-Martinez, F. Gude, A. I. Diaz, S. Pita, J. R. Gonzalez-Juanatey, F. Valdes and D. Sanchez-Guisande, Nefrologia **26** (2), 253-260 (2006).

40. O. Sohnel and P. Novotny, *Densities of aqueous solutions of inorganic substances*. (Amsterdam - Oxford - New York - Tokyo, 1985).

41. N. Mousseau and P. Derreumaux, Accounts Chem Res **38** (11), 885-891 (2005).

42. H. Fricke and A. Parts, J Phys Chem-Us **42**, 1171-1185 (1938).

43. J. Wyman and T. L. McMeekin, Journal of Ammerican Chemical Society **55**, 908-914 (1933).

CHAPTER 4 - Kirkwood-Buff Theory of Molecular and Protein Association, Aggregation and Cellular Crowding*

An analysis of the effect of a cosolvent on the association of a solute in solution using the Kirkwood-Buff theory of solutions is presented. The approach builds on the previous results of Ben-Naim by extending the range of applicability to include any number of components at finite concentrations in both closed and semi-open systems. The derived expressions, which are exact, provide a foundation for the analysis and rationalization of cosolvent effects on molecular and biomolecular equilibria including protein association, aggregation, and cellular crowding. A slightly different view of cellular crowding is subsequently obtained. In particular, it is observed that the addition of large cosolvents still favors the associated form even when traditional excluded volume effects are absent.

Introduction

Protein association, protein unfolding, protein aggregation, and cellular crowding are known to affect the normal function of cellular systems.[1-7] In many cases, the resulting small changes in normal protein-protein intra and intermolecular interactions are thought to lead to a variety of human diseases.[8,9] Consequently, it is important to study these processes, at both the thermodynamic and atomic levels, in an effort to understand and eventually manipulate the behavior of such systems. One way to affect and help understand peptide and protein association is through the use of additives, or cosolvents.[10] A general theory describing these types of effects

* "Reprinted with permission from Moon Bae Gee and Paul E. Smith, Journal of Chemical Physics, 131, 165101-165110, 2009. Copyright 2009, American Institute of Physics."

which relates their thermodynamic properties to the interactions between species in solution is therefore desired. This is a major aim of the present work.

The basic thermodynamic effects of additives on chemical equilibria were outlined some time ago using the concepts of binding and linkage.[11-13] Binding polynomials are then typically used to help illustrate the resulting thermodynamic relationships. While binding polynomials can be used to provide an accurate representation of real experimental data, the physical interpretation of the binding constants requires some care for systems involving only weakly binding cosolvents.[14] This is immediately apparent when attempting to define corresponding binding sites, occupation numbers, and equilibrium constants using coordinate data provided by computer simulations. The transient nature of the interactions between the cosolvent and biomolecule render such a comparison of the experimental and simulation data essentially impossible. Other thermodynamic approaches have been outlined,[15-21] but it remains difficult to relate these to realistic distributions between the various species in solution.

The most common approach used to understand the effects of an additive on protein folding, association, and aggregation has involved scaled particle theory (SPT).[6,19,22,23] Here, the change in free energy for association, denaturation, etc, is determined from the change in the free energy for insertion of the initial and final states into a system of hard spheres as a function of the additive volume fraction. These calculations generally suggest that the native and any associated states are favored on increasing the additive concentration. These are exactly the trends observed experimentally,[6] and for simulations using simple excluded volume crowders.[3,24] As SPT is dominated by repulsive excluded volume effects, the above trends are usually interpreted in terms of a decrease in the free volume, or increase in the excluded volume, within the solution. Unfortunately, it is difficult to extend SPT to include attractive interactions which undoubtedly

occur between proteins in real systems.[6] Hence, other models are still required which complement existing approaches.

The application of the Kirkwood-Buff (KB) theory of solutions to biological systems has recently attracted some interest.[14] KB theory provides a rigorous link between intermolecular distributions in solution and the thermodynamic characteristics of that solution.[25,26] In particular, it has been used to understand the preferential interactions of cosolvents with proteins and small molecules,[27-34] to study changes in the hydration of solutes,[35-38] to investigate a variety of solution properties,[39-41] and for hard sphere models of cellular crowding.[42] Previously, Ben-Naim has used KB theory to understand the effects of cosolvents on the association equilibrium of a solute in solution.[43] The current approach is based on the previous work of Ben-Naim, and is similar to the approach of Hall for studying micelles.[44] Here, we extend the Ben-Naim approach to describe the effects of a cosolvent on the equilibrium constant for association to include semi-open and not just closed systems, and for systems with any number of components all of which can appear at finite concentrations. To achieve this we adopt a recent approach which avoids the usual matrices used for closed systems.[45] The resulting expressions are then used to illustrate some of the differences between open and closed systems, and between real and ideal solutions.

Kirkwood-Buff Theory

Kirkwood-Buff theory is an exact theory of solutions.[46] The principle quantities of interest are the Kirkwood-Buff integrals (KBIs) defined by,

$$G_{ij} = G_{ji} = 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1]\, r^2 dr \qquad (4.1)$$

where g_{ij} is the radial distribution function (rdf) between the center of mass of species i and j. These integrals can be used to relate thermodynamic properties of a solution to the molecular

distributions in that solution. This is achieved by starting with the Grand Canonical ensemble expression,[25]

$$RT\left(\frac{\partial \ln \rho_i}{\partial \mu_j}\right)_{T,\mu_{k\neq j}} = \delta_{ij} + N_{ij} \tag{4.2}$$

where R is the gas constant, T is the absolute temperature, $\rho_i = N_i/V$ is the number density of species *i*, μ is the chemical potential, and δ_{ij} is the Kroenecker delta function. The $N_{ij} = \rho_j G_{ij}$ values describe the change in the number of *j* particles on introduction of a central *i* particle, from the number of *j* particles found in an equivalent volume of bulk solution at the same composition. More details concerning the application of KB theory to understand a variety of solution properties can be found elsewhere.[14,26,39]

The major aim of this work is to develop expressions for cosolvent effects on biomolecular equilibria in both closed and semi-open ensembles. Traditionally, KB theory starts from Equation 4.2 and uses a series of thermodynamic transformations to provide expressions for similar derivatives in closed ensembles.[25,46] A general matrix formulation is available for chemical potential derivatives in closed systems – the most common system of interest.[46] However, we will avoid this approach as the corresponding expressions are difficult to manipulate for large numbers of components. Furthermore, it is also difficult to transform from the closed ensemble results back to expressions valid for semi-open systems.[29]

Chemical Equilibria

Let us consider a formal n_c component system containing a primary solvent (1), a biomolecule (2), and a series of cosolvents (3, 4, . . .). The cosolvents can be small molecules such as urea, a salt, a proton, or larger molecules such as different proteins. We will refer to

species 2 as a biomolecule, but it can easily refer to any associating species. Furthermore, the biomolecule can exist in two forms, one being a monomer (M) and the other an aggregate (A) of any shape formed from *n* monomers. Any counterions associated with the biomolecule will be assumed to have no effect on the equilibrium, and to be different from any of the cosolvents if the latter are salts (no Donnan effect). Hence, we have the equilibrium process,

$$nM \rightarrow A \tag{4.3}$$

which can be described by an equilibrium constant K. Formally, the equilibrium constant involves the activities of the various species. In the majority of thermodynamic approaches it is then assumed that the activities of the biomolecule species can be replaced by their concentrations as the biomolecules usually appear at low concentrations, and concentrations are relatively easy to determine experimentally. Here, we do not assume ideal behavior of any species, but simply define our equilibrium constant in terms of concentrations to match the usual experimental representation of the data. Consequently, the equilibrium constant can be defined in several ways using a variety of concentration scales. We chose,

$$K = \frac{\rho_A}{\rho_M^n} \tag{4.4}$$

in terms of number densities or molarities. Other choices will be discussed later. The number of each form of the biomolecule present in solution are related by,

$$N_M + nN_A = N_2 \qquad dN_M + ndN_A = dN_2 \tag{4.5}$$

and,

$$f_M = \frac{N_M}{N_2} \qquad f_A = \frac{nN_A}{N_2} \qquad f_M + f_A = 1 \tag{4.6}$$

where f_i is the fraction of either A or M at equilibrium. The material equilibrium condition indicates that,

$$\mu_A = n\mu_M \qquad\qquad d\mu_A = nd\mu_M = nd\mu_2 \tag{4.7}$$

as long as we remain at equilibrium. From Equation 4.4 a general change in the equilibrium constant can be written as,

$$d\ln K = d\ln \rho_A - nd\ln \rho_M \tag{4.8}$$

We note that all the above expressions can be applied to any chemical equilibrium that follows Equation 4.3 in any thermodynamically reasonable ensemble. Our main aim is to express the changes in the number densities of A and M in terms of the KB integrals and thereby provide a simple physical picture of the effect of a cosolvent on the equilibrium constant.

General Kirkwood-Buff Theory of Chemical Equilibria

Our system corresponds to a pseudo n_c+1 component system with thermodynamic constraints between two of the components (M and A). Before proceeding we note that the application of KB theory to this type of problem involves some subtle issues. First, N_2 is a true independent thermodynamic variable. However, N_M and N_A are not as they are related through Equation 4.5. Hence, we will avoid taking derivatives with respect to the chemical potentials or concentrations of M *and* A, although derivatives involving either one (M *or* A) can be used in place of the chemical potential or concentration of 2. The application of KB theory implies that the KB integrals used here for semi-open or closed systems correspond to an equivalent system at the same composition but open to all species (including M and A). This issue has been discussed by Ben-Naim.[26] One can apply KB theory to understand such a system as long as one does not treat M and A as independent thermodynamic variables.[47]

The traditional approach to this type of problem starts with the matrix formulation of KB theory for closed systems. However, the evaluation of the matrix determinants for a large number

of components where all components are present at finite concentrations is rather cumbersome. In addition, transforming back to an open or semi-open system from the closed system results is also difficult. Hence, we use a different approach which avoids the direct evaluation of any matrices for small n_c values, thereby greatly simplifying the problem.[45] Let us consider the number density of each species to be functions of T and all the chemical potentials. The differential of the number densities at constant T is then provided by,

$$d\ln\rho_i = \sum_j \left(\frac{\partial \ln \rho_i}{\partial \mu_j} \right)_{T,\mu_{k\neq j}} d\mu_j \tag{4.9}$$

for any *i*, and where the summation is over all *j* components. The partial derivatives can be expressed in terms of KBIs through Equation 4.2 to give,

$$RTd\ln\rho_i = \sum_j (\delta_{ij} + N_{ij}) d\mu_j \tag{4.10}$$

for any constant T ensemble. To our knowledge the above equation was first derived by Hall,[48] but using a different route. The general Gibbs-Duhem (GD) relation at constant temperature can be written,

$$dP = \sum_j \rho_j d\mu_j \tag{4.11}$$

where P is the pressure. Equations 4.10 and 4.11 can be applied to any number of components in any ensemble with T constant.

Hence, for our n_c+1 component system of species 1, M, A, 3, 4, . . ., one can use Equation 4.7 to eliminate $d\mu_A$ from the relationships provided in Equation 4.10 for *i* = A and M to give,

$$RTd\ln\rho_A = (n + nN_{AA} + N_{AM}) d\mu_M + \sum_{j\neq A,M}^{n_c+1} N_{Aj} d\mu_j$$

$$RTd\ln\rho_M = (1 + N_{MM} + nN_{MA})d\mu_M + \sum_{j\neq A,M}^{n_c+1} N_{Mj}d\mu_j \tag{4.12}$$

One could have focused on species A instead of M, but the results will be the same. However, we cannot retain *both* $d\mu_M$ and $d\mu_A$ terms for the remainder of the analysis as this implies they are thermodynamically independent. To generate a relationship for changes in the equilibrium constant we use Equation 4.8 and the relationships in Equation 4.12 to provide,

$$RTd\ln K = (nN_{AA} + N_{AM} - nN_{MM} - n^2N_{MA})d\mu_M + \sum_{j\neq A,M}^{n_c+1}(N_{Aj} - nN_{Mj})d\mu_j \tag{4.13}$$

However, the above expressions can be simplified further.

A series of relationships between KBIs involving the biomolecule can be established. These can be written as,

$$N_{i2} = N_{iM} + nN_{iA}$$

$$N_{2i} = f_M N_{Mi} + f_A N_{Ai} \tag{4.14}$$

for $i \neq$ A, M, or 2 and,

$$N_{M2} = 1 + N_{MM} + nN_{MA}$$

$$N_{A2} = n + nN_{AA} + N_{AM}$$

$$1 + N_{22} = f_M(1 + N_{MM} + nN_{MA}) + f_A(n + nN_{AA} + N_{AM}) = f_M N_{M2} + f_A N_{A2} \tag{4.15}$$

The above expressions were obtained from the general fluctuation formula in the Grand Canonical (μVT) ensemble,

$$\delta_{ij} + N_{ij} = \frac{< N_i N_j > - < N_i >< N_j >}{< N_i >} \tag{4.16}$$

by suitable substitutions of $N_2 = N_M + nN_A$. We note that the above relationships do not assume N_2 is constant. They merely reflect a change of index for the M, A, and 2 species as illustrated in

Figure 1. Their meaning will be discussed later. Application of the above relationships to Equations 4.10 – 4.13 provides,

$$RTd\ln K = \sum_{j}^{n_c}(N_{Aj} - nN_{Mj})d\mu_j \tag{4.17}$$

$$RTd\ln \rho_i = \sum_{j}^{n_c}(\delta_{ij} + N_{ij})d\mu_j \tag{4.18}$$

$$RTd\ln m_i = \sum_{j}^{n_c}(\delta_{ij} + N_{ij} - N_{1j} - \delta_{1j})d\mu_j \tag{4.19}$$

$$dP = \sum_{j}^{n_c}\rho_j d\mu_j \tag{4.20}$$

where d ln m_i = d ln ρ_i – d ln ρ_1, and m_i is the dimensionless molality (ρ_i/ρ_1). The above equations represent the basic expressions required for this study. The summations over the formal n_c components only involve indices 1, 2, 3, etc, and not M or A. They can be applied to any constant T ensemble and are valid for any concentration of solvent, biomolecule, and cosolvents. We note that Equations 4.18-4.20 could have been written directly for n_c component systems. However, Equation 4.17 is not so obvious.

Results

We will apply Equations 4.17-4.20 to a variety of thermodynamic ensembles to develop expressions for the effect of a cosolvent on the chemical equilibrium in terms of KB integrals. During the following analysis several combinations of KBIs appear repeatedly. Hence, in an effort to simplify the results we will define the following,

$$P^i_{jk} = N_{ij} - \frac{\rho_j}{\rho_k}N_{ik} = \rho_j(G_{ij} - G_{ik}) = -\frac{\rho_j}{\rho_k}P^i_{kj} \tag{4.21}$$

which should be read as the preference or affinity of *j* over *k* for species *i*. For example, P_{31}^{A} quantifies the preference of 3 over 1 for the biomolecule in form A. This notation will only be used when it applies to specific differences between affinities of the various components for the two biomolecule forms. A simple physical interpretation of the sign associated with the above expression is that when $P_{jk}^{i} > 0$ the local ratio of *j* to *k* molecules around a central *i* molecule is larger than the bulk ratio of *j* to *k* molecules, and vice versa. We will also make use of the notation,

$$N_{ij}^{+} = N_{ij} + m_j(1 + N_{11} - N_{i1} - N_{j1}) \tag{4.22}$$

originally introduced by Hall,[48] to help simplify many of the results.

General Expressions for any Number of Components in a Closed System

Our initial focus will be on fully closed systems at constant pressure. Eliminating $d\mu_1$ from Equation 4.17 using Equation 4.20 and then taking derivatives with respect to one of the cosolvent molalities one obtains the expression,

$$\left(\frac{\partial \ln K}{\partial \ln m_j}\right)_{T,P,m_{k\neq j}} = \sum_{i>1}(P_{i1}^{A} - nP_{i1}^{M})\mu_{ij} \tag{4.23}$$

or alternatively,

$$RT\left(\frac{\partial \ln K}{\partial \mu_j}\right)_{T,P,m_{k\neq j}} = \left(\frac{\partial \ln K}{\partial \ln a_j}\right)_{T,P,m_{k\neq j}} = \mu_{jj}^{-1}\sum_{i>1}(P_{i1}^{A} - nP_{i1}^{M})\mu_{ij} \tag{4.24}$$

where the sum involves species 2 (not M and A), a_j is the activity of *j*, and we have defined,

$$\mu_{ij} = \frac{1}{RT}\left(\frac{\partial \mu_i}{\partial \ln m_j}\right)_{T,P,m_{k\neq j}} = \frac{1}{RT}\left(\frac{\partial \mu_i}{\partial \ln N_j}\right)_{T,P,N_{k\neq j}} \tag{4.25}$$

The molality derivative has been chosen here as there is a general recursion relationship for these derivatives in any closed multicomponent solution.[45] Reference to the stability requirements for solutions indicates that $\mu_{ii} \geq 0$ and $\mu_{ij} < 0$.[49] Consequently, on increasing the concentration of *j* the association process is favored when $P_{i1}^{A} - nP_{i1}^{M}$ is positive for $i = j$ and negative for $i \neq j$. A change of concentration scales can be performed using the following thermodynamic relationships,

$$\left(\frac{\partial \ln \rho_i}{\partial \ln m_i}\right)_{T,P,m_{j\neq i}} = 1 - \phi_i \qquad \left(\frac{\partial \ln x_i}{\partial \ln m_i}\right)_{T,P,m_{j\neq i}} = 1 - x_i \qquad (4.26)$$

where φ_i is the volume fraction and x_i the mole fraction. It should be noticed these derivatives are positive and so a change in cosolvent concentration variable does not affect the sign of the previous effects.

In a closed system the affinity of all species for the M and A forms contribute to the overall effect. These contributions are expressed relative to species 1 (usually taken as the primary solvent). This is a direct consequence of our choice to eliminate $d\mu_1$ from Equation 4.17 using the GD equation, and not due to the use of molality based derivatives. Clearly, one could eliminate any $d\mu_i$ to obtain a set of equivalent expressions. Finally, one could obtain an expression for a closed system at constant volume (T, $\rho_{j\neq i}$) from Equations 4.17-4.20. However, we have not pursued this further as the resulting expressions are rather complicated and this ensemble is not relevant for most biological systems, although it is the ensemble adopted in SPT.

Constant T, P, and m_2 Ensemble

The most common situation involves a closed system with $n_c = 3$ at constant T and P, where one is interested in the effect of a single cosolvent (3) on the biomolecular equilibrium when the

biomolecule (2) and primary solvent (1) concentrations are constant. To develop expressions for the effect of a cosolvent on the equilibrium in this situation we eliminate $d\mu_1$ from Equations 4.17-4.19 using Equation 4.20. Then, noting that m_2 is constant we have from Equation 4.19 with $i = 2$ that,

$$0 = (1 + N_{22}^{+})d\mu_2 + N_{23}^{+}d\mu_3 \tag{4.27}$$

This can then be used to eliminate $d\mu_2$ from Equation 4.17 to give,

$$RT\left(\frac{\partial \ln K}{\partial \mu_3}\right)_{T,P,m_2} = P_{31}^{A} - nP_{31}^{M} - \frac{(P_{21}^{A} - nP_{21}^{M})N_{23}^{+}}{1 + N_{22}^{+}} \tag{4.28}$$

To relate this to the cosolvent concentration one can eliminate $d\mu_1$ from Equation 4.18 with $i = 3$ using Equation 4.20, and then combine with Equation 4.27 to provide,

$$RT\left(\frac{\partial \ln \rho_3}{\partial \mu_3}\right)_{T,P,m_2} = 1 + N_{33} - N_{13} - \frac{(N_{32} - m_2 N_{31})N_{23}^{+}}{1 + N_{22}^{+}} \tag{4.29}$$

Alternatively, if one is interested in the effect of cosolvent molality one can eliminate $d\mu_1$ from Equation 4.19 using Equation 4.20 with $i = 3$, and then combine with Equation 4.27 to generate,

$$RT\left(\frac{\partial \ln m_3}{\partial \mu_3}\right)_{T,P,m_2} = 1 + N_{33}^{+} - \frac{N_{32}^{+}N_{23}^{+}}{1 + N_{22}^{+}} \tag{4.30}$$

which can then be combined with Equation 4.28 if desired. Equation 4.30 is in agreement with previous results.[45] We note that the final terms in Equations 4.28-4.30 all disappear as the biomolecule concentration tends to zero, a common situation, and are then consistent with previous published expressions.[27,29,47]

In some cases the cosolvent concentration might be constant and the biomolecule concentration may vary. The correct expressions for this situation can be obtained from Equations 4.28-4.30 via a simple $2 \leftrightarrow 3$ index change, or by the same approach that led to

Equations 4.28-4.30 but where m_3 is constant. In principle, Equation 4.18 can be used to obtain expressions for d ln x_i and thereby chemical potential derivatives with respect to mole fractions. These expressions are rather cumbersome and, as mole fractions are rarely the concentration scale of choice for biological systems, we have not pursued this further here.

General Expressions for any Number of Components in Semi-Open Systems

If the system corresponds to that of a real cell, or some approximation to a real cell, then it may be open to one or more components. Furthermore, the addition of the cosolvent can occur with volume or pressure held constant. Let us consider a system which contains a set of species at fixed concentrations (ρ), together with a set of species with fixed chemical potentials (μ). Taking derivatives of Equation 4.17 with respect to one of the cosolvent molarities with volume fixed one obtains the expression,

$$\left(\frac{\partial \ln K}{\partial \ln \rho_j}\right)_{T,\rho_{k\neq j},\mu} = \sum_i^{n_c'} (N_{Ai} - nN_{Mi})\mu_{ij}' \tag{4.31}$$

where the sum is over the restricted set of components for which the system is closed (n_c'). Alternatively, with pressure constant one finds,

$$\left(\frac{\partial \ln K}{\partial \ln \rho_j}\right)_{T,P,N_{k\neq j},\mu} = \sum_i^{n_c'} (N_{Ai} - nN_{Mi})\mu_{ij}'' \tag{4.32}$$

The corresponding chemical potential derivatives are defined,

$$\mu_{ij}' = \frac{1}{RT}\left(\frac{\partial \mu_i}{\partial \ln \rho_j}\right)_{T,\rho_{k\neq j},\mu} = \frac{1}{RT}\left(\frac{\partial \mu_i}{\partial \ln N_j}\right)_{T,V,N_{k\neq j},\mu}$$

$$\mu_{ij}'' = \frac{1}{RT}\left(\frac{\partial \mu_i}{\partial \ln \rho_j}\right)_{T,P,N_{k\neq j},\mu} \tag{4.33}$$

Expressions for the required derivatives can be obtained after solving the set of simultaneous equations provided by Equations 4.18-4.20. This is illustrated in the following two sections.

Constant T, μ_1, and ρ_2 Ensemble

A common osmotic system involves a closed system with n_c = 3 at constant T, V and μ_1, where one is interested in the effect of a single cosolvent (3) on a biomolecular equilibrium when the biomolecule concentration is constant and the system is closed to the cosolvent. To develop expressions for the effect of a cosolvent on the equilibrium in this situation we note that ρ_2 is constant and hence from Equation 4.18 with i = 2 one has,

$$0 = (1 + N_{22})d\mu_2 + N_{23}d\mu_3 \tag{4.34}$$

This can then be used to eliminate $d\mu_2$ from Equation 4.17 to give,

$$RT\left(\frac{\partial \ln K}{\partial \mu_3}\right)_{T,\mu_1,\rho_2} = N_{A3} - nN_{M3} - \frac{(N_{A2} - nN_{M2})N_{23}}{1 + N_{22}} \tag{4.35}$$

To relate this to the cosolvent concentration one can eliminate $d\mu_2$ from Equation 4.18 with i = 3, and then use Equation 4.34 to provide,

$$RT\left(\frac{\partial \ln \rho_3}{\partial \mu_3}\right)_{T,\mu_1,\rho_2} = 1 + N_{33} - \frac{N_{32}N_{23}}{1 + N_{22}} \tag{4.36}$$

Alternatively, if one is interested in the effect of cosolvent molality one can eliminate $d\mu_2$ from Equation 4.19 with i = 3, and then use Equation 4.34 to generate,

$$RT\left(\frac{\partial \ln m_3}{\partial \mu_3}\right)_{T,\mu_1,\rho_2} = 1 + N_{33} - N_{13} - \frac{(N_{32} - N_{12})N_{23}}{1 + N_{22}} \tag{4.37}$$

which can then be combined with Equation 4.35 if desired. We note that the final terms in Equations 4.35-4.37 all disappear as the biomolecule concentration tends to zero.

Constant T, P, μ_1, and N_2 Ensemble

A second common osmotic system involves a closed system with n_c = 3 at constant T, P and μ_1, where one is interested in the effect of a single cosolvent (3) on a biomolecular equilibrium when the number of biomolecules is constant, and the system is also closed with respect to the cosolvent. To develop expressions for the effect of a cosolvent on the equilibrium in this situation we note that P is constant and hence from Equation 4.20 one has,

$$0 = \rho_2 d\mu_2 + \rho_3 d\mu_3 \tag{4.38}$$

This can then be used to eliminate $d\mu_2$ from Equation 4.17 to give,

$$RT\left(\frac{\partial \ln K}{\partial \mu_3}\right)_{T,P,\mu_1,N_2} = N_{A3} - nN_{M3} - \frac{\rho_3}{\rho_2}(N_{A2} - nN_{M2}) \tag{4.39}$$

To relate this to the cosolvent concentration one can eliminate $d\mu_2$ from Equation 4.18 with i = 3, and then use Equation 4.38 to provide,

$$RT\left(\frac{\partial \ln \rho_3}{\partial \mu_3}\right)_{T,P,\mu_1,N_2} = 1 + N_{33} - N_{23} \tag{4.40}$$

Alternatively, if one is interested in the effect of cosolvent molality one can eliminate $d\mu_2$ from Equation 4.19 with i = 3, and then use Equation 4.38 to generate,

$$RT\left(\frac{\partial \ln m_3}{\partial \mu_3}\right)_{T,P,\mu_1,N_2} = 1 + N_{33} - N_{13} - N_{23} + m_3 N_{21} \tag{4.41}$$

which can then be combined with Equation 4.39 if desired.

Symmetric Ideal Solutions

Ideal behavior occurs when the right hand side of Equations 4.29, 4.30, 4.36, 4.37, etc, is unity. Clearly, the corresponding relationships between the KBIs depend on the concentration scale and the ensemble. Symmetric ideal (SI) solutions provide a useful reference point for understanding the properties of real solution mixtures in closed ensembles at constant temperature and pressure.[26] SI solutions are defined by the fact that the mole fraction scale activity coefficients are unity for all components over all compositions, i.e. $d\mu_i = RT\, d \ln x_i$. In terms of the KBIs this requirement is fulfilled when $\Delta G_{ij} = G_{ii} + G_{jj} - 2\, G_{ij} = 0$ for all *i,j* pairs.[26] Recently, we provided a general expression for the KB integrals in SI solutions of any number of components,[50]

$$G_{ij} = RT\kappa_T - V_i - V_j + S_{n_c} \qquad S_{n_c} = \sum_{k=1}^{n_c} \rho_k V_k^2 \tag{4.42}$$

where the sum is over all n_c components of the mixture, κ_T is the isothermal compressibility, and V_i is the molar volume of pure component *i* at the same T and P. For our pseudo n_c+1 component system one can write,

$$S_{n_c+1} = \phi_2 < V_2 > + \sum_{k \neq M,A}^{n_c+1} \rho_k V_k^2$$

$$< V_2 > = f_M V_M + f_A V_A = V_2 (f_M + n f_A) \tag{4.43}$$

$$\phi_2 = \rho_M V_M + \rho_A V_A = (\rho_M + n\rho_A) V_2 = \rho_2 V_2$$

In developing the above expressions we have made the very reasonable assumption that,

$$V_A = nV_M = nV_2 \tag{4.44}$$

For instance, estimated differences in volume between native and denatured proteins are typically small compared to their total volumes.[27,51,52] The value of $<V_2>$ is the average molar

volume of the biomolecule at that particular composition and varies between $V_M = V_2$ for $f_M = 1$ and $V_A = nV_2$ for $f_A = 1$.

Using Equation 4.42 one can develop expressions for the various combinations of KB integrals that appeared in the previous sections. Hence, one finds,

$$N_{ij} - N_{kj} = \rho_j (V_k - V_i)$$

$$N_{ij} - m_j N_{i1} = \rho_j (V_1 - V_j) \tag{4.45}$$

$$N_{ij}^{+} = m_j$$

for any *i* and *j* = 1, 2, etc. For KB integrals involving specific forms of the biomolecule we have,

$$N_{Ai} - nN_{Mi} = (n-1)\rho_i (V_i - RT\kappa_T - S_{n_c+1})$$

$$P_{ij}^{A} - nP_{ij}^{M} = (n-1)\rho_i (V_i - V_j) \tag{4.46}$$

for any *i* and *j* = 1, 2, etc after using Equation 4.44. In both cases the mean value of the molar volume $<V_2>$ should be used when *i* or *j* = 2.

For symmetric ideal solutions one finds that $\mu_{ij} = \delta_{ij} - x_j$ and therefore insertion of the SI expressions into Equation 4.23 and performing the summation leads to a general result for SI solutions,

$$\left(\frac{\partial \ln K}{\partial \ln m_j} \right)_{T,P,m_{k\neq j}}^{SI} = (n-1)\rho_j (V_j - V_m) = (n-1)(\phi_j - x_j) \tag{4.47}$$

valid for any number of components. Here, $V_m = V/(N_1 + N_2 + \cdots)$ is the molar volume, or the average volume per molecule of the solution, and $\varphi_i = \rho_i V_i$ is the volume fraction. Hence, an increase in the concentration of the biomolecule or an additive has no affect on the equilibrium in SI solutions when *n* = 1, while they increase the equilibrium constant for *n* > 1 when their molar volume is larger than the average volume of the solution components. This appears to follow the

result expected for a simple excluded volume effect where the excluded volume is smaller for the aggregate compared to an equivalent number of monomers. However, this is incorrect as we have not assumed anything concerning the character (size or shape) of either the M or A species. The real origin of the result is discussed later. SI solution results for other concentration scales are given by,

$$\left(\frac{\partial \ln K}{\partial \ln \rho_j}\right)^{SI}_{T,P,m_{k\neq j}} = (n-1)\rho_j \frac{V_j - V_m}{1-\phi_j}$$

$$RT\left(\frac{\partial \ln K}{\partial \mu_j}\right)^{SI}_{T,P,m_{k\neq j}} = \left(\frac{\partial \ln K}{\partial \ln x_j}\right)^{SI}_{T,P,m_{k\neq j}} = (n-1)\rho_j \frac{V_j - V_m}{1-x_j} \tag{4.48}$$

where we have used the previous relationships between the concentration derivatives.

Alternative Definitions of the Equilibrium Constant

Our choice of an equilibrium constant utilizing number densities was motivated by the pseudo chemical potential (pcp) approach pioneered by Ben-Naim.[46] The pcp (μ*) is related to the total chemical potential by the equation,

$$\mu_i = \mu_i^* + RT \ln(\Lambda_i^3 \rho_i) \tag{4.49}$$

where Λ is the thermal deBroglie wavelength. The equilibrium condition then provides,

$$RTd \ln K = -(d\mu_A^* - nd\mu_M^*) \tag{4.50}$$

The pcp approach helps to simplify the resulting expressions for the same reasons as outlined previously.[26] Other definitions of the equilibrium constant are possible, but typically generate extra terms which complicate the analysis. This can be seen from the following expressions which relate various definitions of the equilibrium constant to the one used here. For instance,

$$d\ln K_m = d\ln K + (n-1)d\ln \rho_1 \qquad K_m = \left(\frac{m_A}{m_M^n}\right) \tag{4.51a}$$

$$d\ln K_x = d\ln K + (n-1)d\ln \rho \qquad K_x = \left(\frac{x_A}{x_M^n}\right) \tag{4.51b}$$

$$d\ln K_f = d\ln K + (n-1)d\ln \rho_2 \qquad K_f = \left(\frac{f_A}{f_M^n}\right) \tag{4.51c}$$

$$d\ln K_\rho = d\ln K + (n-1)d\ln \rho_M \qquad K_\rho = \left(\frac{\rho_A}{\rho_M}\right) \tag{4.51d}$$

where ρ is the total number density. The final terms in Equations 4.51a-d represent additional contributions which depend on the ensemble. They are properties of the solution mixture itself, and not directly related to the affinity of any of the species to the different forms of the biomolecule (see later). Also, we note that defining an equilibrium constant by $K' = n\rho_A/\rho_M^n$ does not affect any of the results presented here as $d \ln K' = d \ln K$. The final expression (Equation 4.51d) does not include the factor of *n* for the equilibrium constant. Hence, this does not correctly reflect the stoichiometry of the association reaction. However, it may relate more meaningfully to the experimental data which are typically indirect measures of biomolecular concentrations obtained through spectroscopic or activity measurements, or for solutions where a unique value of *n* might not be known. Clearly, different results are obtained with different definitions of the equilibrium constant, and therefore comparisons with experimental data should be performed with care.

Approximate Free Energy Curves

KB theory provides expressions for derivatives of the equilibrium constant (or free energy) for a particular process in terms of the intermolecular distributions observed in solution. To obtain changes in the equilibrium constant one has to then integrate. Unfortunately, the general dependence of the N_{ij}'s on composition is unknown. However, one can make some reasonable approximations and investigate the results. First, most studies involve the solute (2) at low concentrations. Second, if the G_{ij}'s are assumed to be relatively constant then one can obtain approximate free energy curves. Therefore, for two of the main ternary systems (Equations 4.28 and 4.29, and Equations 4.31 and 4.32) one finds,

$$-\beta\Delta\Delta G = \ln\left(\frac{K}{K_0}\right)_{T,P,m_2\to 0} = \frac{P_{31}^{A} - nP_{31}^{M}}{N_{33} - N_{13}}\ln(1 + N_{33} - N_{13}) \tag{4.52}$$

$$-\beta\Delta\Delta G' = \ln\left(\frac{K}{K_0}\right)_{T,\mu_1,\rho_2\to 0} = \frac{N_{A3} - nN_{M3}}{N_{33}}\ln(1 + N_{33}) \tag{4.53}$$

which may be compared to the corresponding exact SI result,

$$-\beta\Delta\Delta G^{SI} = \ln\left(\frac{K}{K_0}\right)^{SI}_{T,P,m_2\to 0} = -(n-1)\ln[1 + \rho_3(V_1 - V_3)] \tag{4.54}$$

where $\Delta\Delta G = \Delta G(\rho_3) - \Delta G(0)$, $\beta = 1/RT$, and K_0 is the equilibrium constant in the absence of cosolvent. A similarity to denaturation binding models is clearly apparent.[53-55] For small ρ_3 the right hand side reduces to $P_{31}^{A} - nP_{31}^{M}$ and $N_{A3} - nN_{M3}$ for the closed and open systems, respectively. Hence, using these approximations the cosolvent effect is predicted to be linear in cosolvent molarity for low cosolvent concentrations – a result observed experimentally.[56] We

note that the infinitely dilute biomolecule limit applied to Equation 4.38 results in $d\mu_3 = 0$ for the T, P, μ_1 ensemble.

There are additional relationships relating the various KBIs which can be related to the properties of solution mixtures. For instance, taking derivatives of Equation 4.17 with respect to P and keeping T and all N_i constant provides,

$$(1-n)RT\kappa_T = \sum_i^{n_c} (N_{Ai} - nN_{Mi})\overline{V_i} \tag{4.55}$$

where $\overline{V_i}$ is the partial molar volume (pmv). The above expression assumes that N_A and N_M are reasonably independent of pressure. Equation 4.55 can be developed further for the case of $n_c = 3$ and an infinitely dilute biomolecule, by assuming that the left hand side of Equation 4.55 is small and may be neglected. This is true for $n = 1$ and should be reasonable for $n < 10$ for all but small cosolvent concentrations. In this case one can relate the preference of 3 over 1 for the biomolecule to just the affinity of 3 for both forms,

$$P_{31}^A - nP_{31}^M \approx \frac{N_{A3} - nN_{M3}}{\phi_1} \tag{4.56}$$

This is a generalization of a previous result.[27,57] The relationship exists because, under these approximations, an increase in the local density of one species around the biomolecule should be accompanied by some degree of depletion of another species in the same region. Using Equation 4.56 in Equation 4.52 and comparing with Equation 4.53 one can predict the relative effects of the addition of a cosolvent in both open and closed systems. The result is,

$$\phi_1 \Delta\Delta G \approx \Delta\Delta G' \tag{4.57}$$

for low biomolecule and cosolvent concentrations. This also assumes that the KBIs are reasonably independent of the osmotic pressure, i.e. $G_{ij}(T,N,P) \approx G_{ij}(T,N,P+\Pi)$. Therefore, the

cosolvent effect should be larger in magnitude in a closed system compared to a similar corresponding semi-open system.

Relationship to the Ben-Naim Result for Closed Systems

Previously, Ben-Naim studied the effects of cosolvents on molecular association using KB theory.[43] In particular, the effect of a cosolvent on an equilibrium involving an infinitely dilute solute expressed in terms of the quantity $\left(\partial N_A / \partial N_3\right)_{T,P,m_2 \to 0}$ in closed systems. The results presented here are different from the Ben-Naim expression for several reasons. First, we have used an equilibrium constant in terms of numbers densities instead of just N_A. Second, the cosolvent concentration has been expressed in terms of molarity or molality rather than N_3. This does not change the general conclusions obtained from the respective expressions, although it does change the expressions themselves. To illustrate further we will transform our result (Equations 3.28 and 3.30) into the corresponding Ben-Naim expression. To do this we note that one can define an equilibrium constant using the molecule numbers (K_N). This is related to the equilibrium constant used here by,

$$d\ln K = d\ln\left(\frac{N_A}{N_M^n}\right) - d\ln\left(\frac{V}{V^n}\right) = d\ln K_N + (n-1)d\ln V \tag{4.58}$$

Taking the appropriate derivative in the required ensemble one obtains,

$$\left(\frac{\partial \ln K}{\partial \ln N_3}\right)_{T,P,m_2} = \left(\frac{\partial \ln K_N}{\partial \ln N_3}\right)_{T,P,m_2} + (n-1)\rho_3 \overline{V_3} \tag{4.59}$$

Finally, the relationship between changes in the equilibrium constant and changes in N_A can be obtained from Equation 4.5 with N_2 constant,

$$d\ln K_N = \frac{N_M + n^2 N_A}{N_A N_M} dN_A \tag{4.60}$$

Combining these relationships with Equations 4.28 and 4.30 one obtains the Ben-Naim result in our notation,

$$\left(\frac{\partial N_A}{\partial N_3}\right)_{T,P,m_2\to 0} = \frac{N_A N_M}{N_3(N_M + n^2 N_A)} \frac{P_{31}^A - nP_{31}^M - (n-1)m_3(1+N_{11}-N_{31})}{1+N_{33}^+} \tag{4.61}$$

where we have used the KB expression for the pmv of 3 in a binary solution of 1 and 3.[46] Clearly, the expression presented here is simpler and easier to interpret. This is a direct consequence of investigating the changes in an equilibrium constant defined using number densities, i.e. the pseudo chemical potential approach.

Discussion

In the above sections we have provided an analysis, using KB theory, of the effects of increasing biomolecule and cosolvent concentrations on molecular association in solution. KBIs involving species 2, M, and A appear in these expressions, even though they are different representations of the same species. This was done deliberately. The use of N_{22} in the above expressions serves to indicate that this corresponds to a KBI between all forms of the biomolecule where one does not distinguish between monomer and aggregate. Alternatively, the use of N_{M3} (or N_{A3}) indicates a correlation between M and 3 (or A and 3) which is specific for one particular form of the biomolecule. In addition, the use of N_{M2} (or N_{A2}) signifies a correlation between the monomer (or aggregate) form and any other form of the biomolecule, both monomer and aggregate. Hence, we have nonspecific effects involving the KBIs for species 2, together with specific effects involving the KBIs for species M or A.

The resulting expressions for closed systems involve terms of the form $P_{i1}^A - nP_{i1}^M$. These terms quantify the excess (or depletion) of *i* over 1 in the vicinity of an aggregate over the same

excess (or depletion) of *i* in the vicinity of *n* monomers. Therefore, if a species *i* displays a higher affinity for the aggregate than *n* individual monomers, then $P_{i1}^{A} - nP_{i1}^{M} > 0$ and an increase in the concentration of *i* leads to an increase in equilibrium constant and a shift in the equilibrium towards the associated form. The cosolvent urea serves as a reasonable example. Urea is well known to denature proteins.[56] Hence, when $n = 1$ the preference of urea (over water) for the denatured state must be larger, on average, than that for the native state. Assuming this is related to the corresponding increase in exposed surface area upon denaturation,[58] this suggests that urea should stabilize the monomer over the aggregate as an aggregate typically possesses less surface area than *n* monomers. Hence, urea is a denaturant of proteins but should help prevent protein aggregation. This behavior has been observed experimentally.[59,60] The opposite effects will be observed for classic osmolytes.

Some general trends are observed in the results. From Equation 4.17 it is clear that the effect on the equilibrium does not involve any KBIs between the biomolecule and any species that is kept at a constant chemical potential. Any species at constant N will contribute to changes in the equilibrium as quantified through the KBIs. Hence, all the species contribute in closed systems (see Equation 4.23). The main reason for this is quite simple. When a cosolvent displaces a solvent molecule in the vicinity of the biomolecule both the addition of cosolvent and the removal of a solvent (or other) molecule from the biomolecule affect the Gibbs free energy of the system. On the other hand, when a cosolvent displaces a solvent molecule in a system open to the solvent, only the addition of the cosolvent affects the free energy of the system. A change in concentration scale ($\mu \rightarrow m \rightarrow \rho$) typically results in a scaling effect (Equation 4.26), but this does not alter the sign of the effects contributing to the change in equilibrium constant.

The expression provided in Equation 4.47 for closed SI solutions indicates that the association equilibrium is increased by the addition of additives with a larger molar volume than the average molecular volume of the solution at that composition. The same expression can be obtained from Equation 4.51b noting that d ln $K_x = 0$ for SI solutions (see Equation 4.7). Hence, we have two contributions to changes in the equilibrium constant for SI solutions. First, on addition of a cosolvent the volume of the solution increases. This affects the number densities of M and A, and directly increases the equilibrium constant when $n > 1$. Second, the addition of another particle to the system changes the total number of species present and therefore the mole fraction of each species. This causes an increase in the monomer form to maintain the equilibrium condition. Finally, we note that dilution, i.e. the addition of solvent, will favor dissociation as long as the solvent has a low molar volume. These effects dominate in the absence of a particular affinity between any of the pairs of species in solution. There is no effect on a simple denaturation equilibrium in SI solutions.

It is well known that cellular crowding, as described by the addition of rather large cosolvents, will tend to favor association.[6] The usual explanation for this observation is that aggregates display less surface area, and therefore less excluded volume, compared to a collection of their monomers. Excluded volume effects also favor the native state over the denatured state for the $n = 1$ case. In contrast, the SI result discussed above indicates that any excluded volume contributions are balanced by other (favorable) contributions so that $\Delta G_{ij} = 0$ for all *i,j* pairs. In a hard sphere fluid model, as adopted in SPT, the SI condition does not hold and excluded volume contributions are dominant. These also favor the associated or native forms. However, we emphasize that even for ideal (mole fraction scale) solutions the addition of a large

cosolvent favors association when the equilibrium constant is defined using molarities, and this effect is not related to any change in excluded volume.

The results predicted by Equation 4.54 for closed ternary SI solutions are illustrated in Figure 4.2. A variety of cosolvents are considered with different molar volumes relative to the primary solvent. The effects are linear in volume fraction at low cosolvent concentrations but display deviation from linear behavior at higher concentrations. All the results scale with n-1 and are independent of the size or shape of the monomer or aggregate, although Equation 4.44 was used during the derivation. The magnitude of the effects obtained from Equation 4.54 for closed SI solutions is similar to that predicted by SPT depending on the biomolecule shapes used (see Figure 3 from reference 6, for instance). Clearly, we have two solution models which display similar results in qualitative agreement with experiment. Both are approximations to real solution conditions of biological interest. The SI approach represents a thermodynamic model, while SPT presents a physical model for solutions. It is unclear at present which, if any, is more reasonable.

The analysis of cosolvent effects on a chemical equilibrium using KB theory is exact. The KBIs quantify the local composition of the solution surrounding the various species corresponding to a four component system for the case of a single cosolvent. There is no problem extracting all the required KBIs from computer simulations as long as one can define the aggregate in a consistent manner. From a thermodynamic point of view, however, this is formally a three component system. Consequently, it is difficult to extract the individual KBIs from an analysis of the thermodynamic data alone. For instance, a general n_c component system has $n_c(n_c+1)/2$ unique G_{ij} integrals. To determine the integrals from experimental data using the KB inversion approach requires 1 compressibility, n_c-1 independent pmvs, and therefore $n_c(n_c-1)/2$ independent chemical potential derivatives (μ_{ij}) as a function of composition. However, for

our pseudo four component system one can use the relationships in Equation 4.28-4.30 as long as one knows K and *n* for a particular system as a function of composition. Hence, one still has only six (n_c = 3) unique KBIs as many of the ten (n_c = 4) KBIs are related through Equations 4.14 and 4.15 and therefore are not unique.

Conclusions

In this study we have applied KB theory to investigate the effects of cosolvents on molecular equilibria in solution. This is an extension of the previous approach of Ben-Naim to include any number of components at finite concentrations in both closed and semi-open systems. No assumption concerning the character, specifically the size and/or shape, of the molecules has been made. Hence, the results are totally general and can be applied to a variety of processes such as protein denaturation, protein aggregation, and cellular crowding, where the cosolvent can be either small (H^+, urea) or large (another protein). In addition, the extension to include semi-open systems also allows one to consider systems under typical biological (cellular) conditions. Here, the effect of a cosolvent appears to be reduced compared to closed systems.

The results obtained here depend on the definition of the equilibrium constant. We advocate the use of molarities in accord with the pseudo chemical potential approach of Ben-Naim. In this case, even SI solutions, where significant excluded volume effects are absent, indicate an increase in association on the addition of a cosolvent crowder. This does not mean that excluded volume effects are not important in these systems. Merely that at least some of the typical effect can be explained without invoking excluded volume. Significant deviations from SI behavior will be observed when the cosolvent displays a preference for either form of the biomolecule.

A particular advantage of the KB approach is that one has exact expressions for the cosolvent effects in terms of KBIs which are directly related to molecular distributions. Hence, one can develop a series of models, each providing different approximations to the real KBIs, and thereby determine the resulting thermodynamic effects of various approaches. We are currently using this type of approach to compare in more detail the KB based results for hard sphere crowders to those obtained from the traditional SPT approach.

References

1. P. van den Berg, E. W. Chung, C. V. Robinson, P. L. Mateo, and C. M. Dobson, EMBO Journal **18**, 4794 (1999).
2. H. X. Zhou and K. A. Dill, Biochemistry **40**, 11289 (2001).
3. M. S. Cheung, D. Klimov, and D. Thirumalai, Proceedings of the National Academy of Sciences of the United States of America **102**, 4753 (2005).
4. R. J. Ellis and A. P. Minton, Biological Chemistry **387**, 485 (2006).
5. K. Richter, M. Nessling, and P. Lichter, Biochimica et Biophysica Acta-Molecular Cell Research **1783**, 2100 (2008).
6. H. X. Zhou, G. N. Rivas, and A. P. Minton, Annual Review of Biophysics **37**, 375 (2008).
7. L. A. Munishkina, A. Ahmad, A. L. Fink, and V. N. Uversky, Biochemistry **47**, 8993 (2008).
8. C. M. Dobson, Nature **426**, 884 (2003).
9. C. A. Ross and M. A. Poirier, Nature Medicine **10**, S10-S17 (2004).
10. S. N. Timasheff, Advances in Protein Chemistry **51**, 355 (1998).
11. J. Wyman and S. J. Gill, *Binding and Linkage*, (University Science Books, Mill Valley, California, 1990).
12. C. Tanford, Advances in Protein Chemistry **24**, 1 (1970).
13. J. A. Schellman, Biopolymers **17**, 1305 (1978).
14. V. Pierce, M. Kang, M. Aburi, S. Weerasinghe, and P. E. Smith, Cell Biochemistry and Biophysics **50**, 1 (2008).
15. T. L. Hill, Journal of Chemical Physics **34**, 1974 (1961).

16. C. Tanford, Proceedings of the National Academy of Sciences of the United States of America **71**, 1811 (1974).
17. F. H. Stillinger and A. Ben-Naim, Journal of Chemical Physics **74**, 2510 (1981).
18. A. P. Minton, Molecular and Cellular Biochemistry **55**, 119 (1983).
19. A. P. Minton, G. C. Colclasure, and J. C. Parker, Proceedings of the National Academy of Sciences of the United States of America **89**, 10504 (1992).
20. A. P. Minton, Biophysical Journal **63**, 1090 (1992).
21. V. A. Parsegian, R. P. Rand, and D. C. Rau, Proceedings of the National Academy of Sciences of the United States of America **97**, 3987 (2000).
22. S. B. Zimmerman and A. P. Minton, Annual Review of Biophysics and Biomolecular Structure **22**, 27 (1993).
23. H. X. Zhou, Proteins **72**, 1109 (2008).
24. M. Wojciehowski and M. Cieplak, Biosystems **94**, 248 (2008).
25. J. G. Kirkwood and F. P. Buff, Journal of Chemical Physics **19**, 774 (1951).
26. A. Ben-Naim, *Molecular Theory of Solutions*, (Oxford University Press, New York, 2006).
27. M. Aburi and P. E. Smith, Journal of Physical Chemistry B **108**, 7382 (2004).
28. M. Kang and P. E. Smith, Fluid Phase Equilibria **256**, 14 (2007).
29. P. E. Smith, Biophysical Journal **91**, 849 (2006).
30. S. Shimizu and C. L. Boon, Journal of Chemical Physics **121**, 9147 (2004).
31. S. Shimizu, W. M. McLaren, and N. Matubayasi, Journal of Chemical Physics **124**, 234905 (2006).
32. I. L. Shulgin and E. Ruckenstein, Journal of Chemical Physics **123**, 054909 (2005).

33. D. Trzesniak, N. F. A. Van Der Vegt, and W. F. van Gunsteren, Molecular Physics **105**, 33 (2007).

34. J. Rosgen, B. M. Pettitt, and D. W. Bolen, Protein Science **16**, 733 (2007).

35. P. E. Smith, Journal of Physical Chemistry B **103**, 525 (1999).

36. R. Chitra and P. E. Smith, Journal of Physical Chemistry B **105**, 11513 (2001).

37. S. Shimizu, Proceedings of the National Academy of Sciences of the United States of America **101**, 1195 (2004).

38. S. Shimizu and N. Matubayasi, Chemical Physics Letters **420**, 518 (2006).

39. E. Matteoli and G. A. Mansoori, *Fluctuation Theory of Mixtures*, (Taylor & Francis, New York, 1990).

40. Y. Marcus, Monatshefte Fur Chemie **132**, 1387 (2001).

41. A. Ben-Naim, A. M. Navarro, and J. M. Leal, Physical Chemistry Chemical Physics **10**, 2451 (2008).

42. J. Roesgen, *Macromolecular Crowding and Solvation: Direct and Indirect Impact on Protein Reactions*, J. W. Shriver (Humana Press, 2009), Vol. 490, Chap. 9, pp.195-225.

43. A. Ben-Naim, Journal of Chemical Physics **63**, 2064 (1975).

44. D. G. Hall, *Thermodynamics of Micellar Solutions*, E. Wyn-Jones and J. Gormally (Elsevier, Amsterdam, 1983), Vol. 26, Chap. 2, pp.7-69.

45. M. Kang and P. E. Smith, Journal of Chemical Physics **128**, 244511 (2008).

46. A. Ben-Naim, *Statistical Thermodynamics for Chemists and Biochemists*, (Plenum Press, New York, 1992).

47. P. E. Smith, Journal of Physical Chemistry B **108**, 18716 (2004).

48. D. G. Hall, Transactions of the Faraday Society **67**, 2516 (1971).

49. I. Prigogine and R. Defay, *Chemical Thermodynamics*, (Longmans, Green & Co, London, 1954).

50. P. E. Smith, Journal of Chemical Physics **129**, 124509 (2008).

51. C. A. Royer, Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology **1595**, 201 (2002).

52. P. E. Smith, Biophysical Chemistry **113**, 299 (2005).

53. J. A. Schellman, Biophysical Chemistry **96**, 91 (2002).

54. K. C. Aune and C. Tanford, Biochemistry **8**, 4586 (1969).

55. G. I. Makhatadze, Journal of Physical Chemistry B **103**, 4781 (1999).

56. R. F. Greene Jr and C. N. Pace, Journal of Biological Chemistry **249**, 5388 (1974).

57. S. Shimizu, Journal of Chemical Physics **120**, 4989 (2004).

58. J. K. Myers, C. N. Pace, and J. M. Scholtz, Protein Science **4**, 2138 (1995).

59. D. K. Klimov, J. E. Straub, and D. Thirumalai, Proceedings of the National Academy of Sciences of the United States of America **101**, 14760 (2004).

60. N. J. Nosworthy and A. Ginsburg, Protein Science **6**, 2617 (1997).

Figure Captions

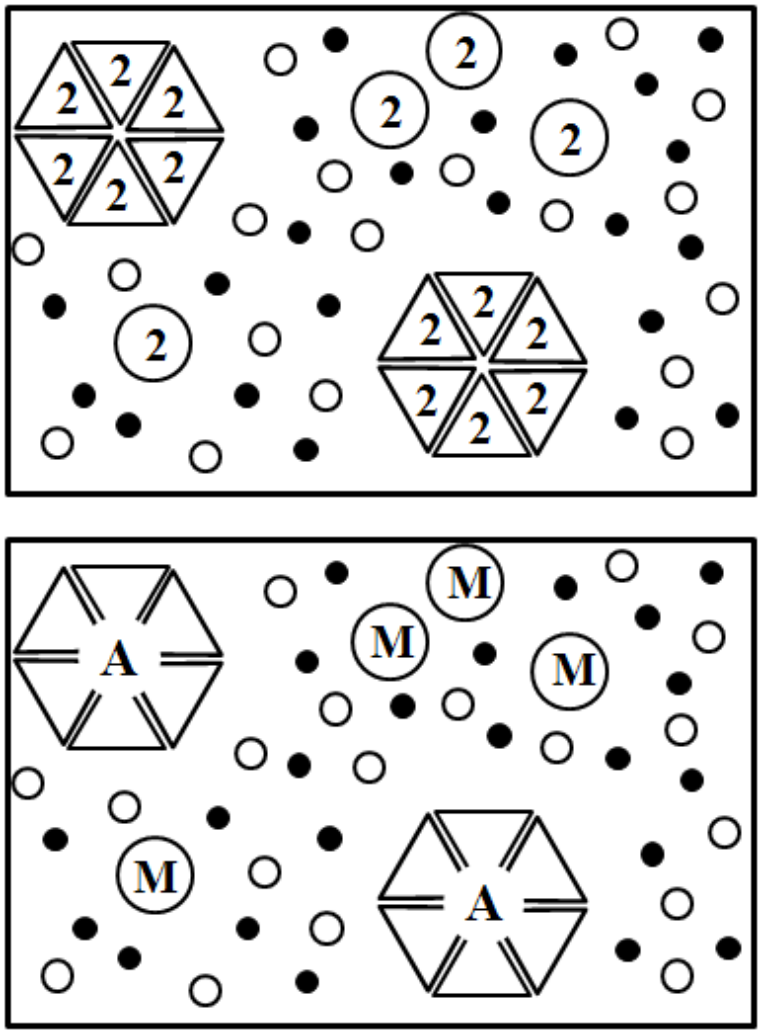
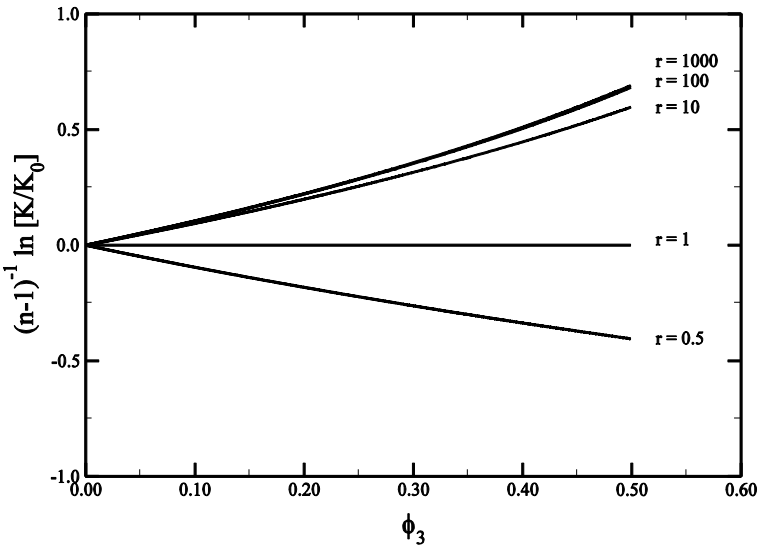
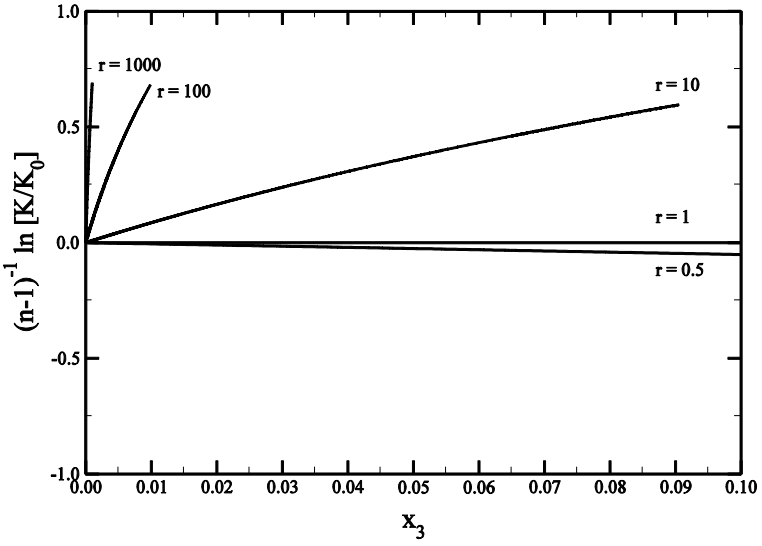


Figure 4.1 The two representations of the same system used in this study. The system contains a solvent (1, shaded spheres), a solute (2), and a cosolvent (3, open spheres). In this case the solute can exist in two forms - one being the monomer (M) and the other being an aggregate (A) of n = 6 monomers. The monomer can adopt different shapes in the associated and free forms.


1.0
0.5
0.0
-0.5
-1.0
0.00
0.10
0.20
0.30
0.40
0.50
0.60
r = 1000
r = 100
r = 10
r = 1
r = 0.5
(n-1)^-1 ln [K/K0]
ϕ3


1.0
0.5
0.0
-0.5
-1.0
0.00
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.10
r = 1000
r = 100
r = 10
r = 1
r = 0.5
(n-1)^-1 ln [K/K0]
x3

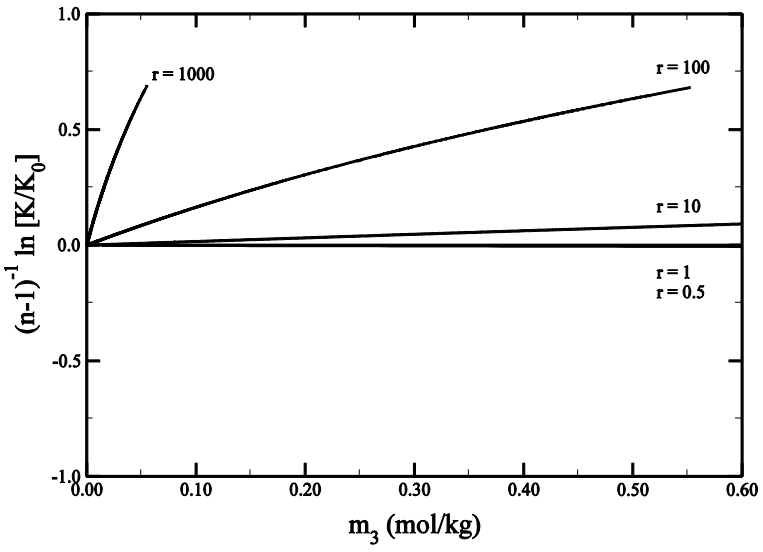


Figure 4.2 SI results for the effects of a single crowder (3) on the association equilibrium (nM → A) of an infinitely dilute solute (2) in a primary solvent (1) in a closed system. The data was obtained using Equation 4.54 for different molar volume ratios (r) of the crowder and solvent such that $V_3 = r\ V_1$. The results are plotted as a function of cosolvent volume fraction ($\varphi_3 = \rho_3 V_3$), cosolvent mole fraction (x_3), and cosolvent molality (m_3) with water as the solvent. In this case $x_3 = \varphi_3/[\varphi_3+r(1-\varphi_3)]$ and $m_3 = 1000\varphi_3/r(1-\varphi_3)/18.015$. For reference, the molar volume of pure water is $V_1 = 18\ cm^3/mol$ and so r = 1000 would correspond to a 25 kDa protein of approximately 225 residues. All curves are truncated at a cosolvent volume fraction of 0.5.

CHAPTER 5 - Molecular Dynamics Simulations of Peptide Aggregation in Closed and Semi-Open Systems

In an effort to develop an understanding of peptide aggregation at the atomic level we have performed simulations of polyglycine ($(gly)_n$) peptides using our recently developed force fields. Experimentally, the association of glycine polypeptides increases with n. Our force fields reproduce this behavior, and the reasons behind this trend are investigated. In addition, we also simulate these systems in a semi-open ensemble, designed to mimic cellular environments typically open to water, using a simple direct approach. The differences between the two ensembles are investigated and compared with our recent theoretical descriptions of aggregating systems using Kirkwood-Buff theory.

Introduction

Peptide aggregation is one of the most interesting issues arising in several pharmaceutical and scientific areas. Many diseases including Alzheimer's, Huntington's, Amyotrophic Lateral Sclerosis (ALS), and prion diseases,[1-2] are related to peptide aggregation which is affected by among other things, the salt concentration or type, temperature, and pH of the cellular environment.[3-8] Potentially, understanding the connections between peptide aggregation and the aggregation environment could help to elucidate the dominant factors which lead to aggregation. Therefore, intensive experimental, theoretical, and computational research has been tried to describe changes in solution distributions, including peptide aggregation, at the atomic level.[9]

Unfortunately, it is difficult to describe diseases at atomic level using typical experimental approaches because the cellular environment is too complicated to be assigned. Furthermore, from the point of view of the atomistic computer simulation, peptide aggregation requires

prohibitive amounts of computer time. Therefore, several approaches focus on studies of small peptide fragments,[3, 10] which hopefully represent the behavior of large proteins, because the small peptide fragments are computationally easy to handle.

Recently, N-methylacetamide (NMA) has been proposed as an attractive model for the investigation of peptide aggregation.[11-13] This is because NMA is one of the simplest molecules containing the peptide linkage and has the advantage of being small enough to be accessible by both current experimental and computational techniques. According to the these studies, it is found that as the NMA concentration is increased the water molecules become increasingly isolated into short linear clusters, while the NMA molecules show significant self-association even at the lowest concentrations explored.[11] Alternatively, Smith and coworkers have used another approach, based on KB integrals, to help quantify the NMA self interaction in dilute solution.[14] In relatively dilute solutions their analysis suggests that the molecules are highly solvated and there are no apparent strong hydrogen bonds between NMA molecules.

Another nice candidate model system may be the small polyglycine peptides, which are zwitterionic in aqueous solution, giving rise to hydrophilic functional groups in the CO_2^- and NH_3^+ groups. In addition, the trend of self-association for the peptide backbone can be investigated as the peptide bonding should increase in diglycine and triglycine which have one or two peptide bonds, respectively. Recently, the aggregation of Gly-*l*-Ala dipeptide molecules in water was investigated by using molecular dynamics simulation.[15] This work suggested that the zwitterionic N-termial and C-termial are responsible for driving peptide association in aqueous solution.

Aggregation is sensitive to environmental conditions. Typical cellular environments involve a system open to water. In an effort to develop the use of computer simulations for semi-open

systems, as a model for cellular systems, several simulations of osmotic pressure have been performed. The simulations have not only been performed to test the force fields used in computer simulations, but they also provide a route to study a wide range of conditions. Murad and coworkers has introduced a novel technique for studying fluids in confined geometries, such as semi-permeable membranes, based on the conventional microcanonical molecular dynamics methods.[16-18] Rowley and coworkers have also proposed osmotic molecular dynamics (OMD) for obtaining chemical potential of Lennard-Jones (LJ) fluids and their mixtures.[19-22] Recently, Luo and Roux described a simple method to compute the osmotic pressure directly from molecular dynamics (MD) simulation of concentrated aqueous solutions by introducing an idealized semi-permeable membrane.[23] This approach was then used for developing force fields of Na^+, K^+, and Cl^- ions. The approach used semi-permeable walls to mimic cell membranes which are open to water molecules, followed by a calculation of the pressure on the walls. The pressure is the osmotic pressure. Even though previous work has reproduced the obtained experimental osmotic pressure well for many systems, the simulations involving more complicated solutions including amino acids and polypeptides have not yet been explored. In the previous chapter we presented a theory for the effects of a cosolvent, pH, or osmotic pressure on a biomolecular equilibrium. Here, we start to explore these types of effects by simulation.

In a previous chapter (chapter 3) we present the force field parameters for the polyglycines, as unprotected N-terminal and C-terminal zwitterionic peptides, which are applicable over the whole concentration range. In order to elucidate the local structure of aqueous polyglycine peptide solutions, the relationships between solvent and solute structure and the relationship to peptide aggregation is investigated using Kirkwood-Buff theory.[24] In addition, in order to mimic real cellular environments we employ a membrane mimic composed of simple frozen particles

and attempt to reproduce the experimentally observed osmotic pressure changes for these simple peptide systems.

Methods

Preferential Interactions

In order to quantify peptide aggregation we need to develop a method to analyze the experimental data. The preferential interaction plays an important role in quantifying solution distributions. Smith and coworkers have presented an analysis of preferential interactions based on Kirkwood-Buff integrals.[25-28] The preferential interaction is defined by

$$P_{jk}^{i} = N_{ij} - \frac{\rho_j}{\rho_k} N_{ik} = \rho_j (G_{ij} - G_{ik}) = -\frac{\rho_j}{\rho_k} P_{kj}^{i}. \tag{5.1}$$

This should be read as the preference or affinity of *j* over *k* for species *i*. This notation will only be used when it applies to specific differences between affinities of the various components for a biomolecule found in different forms (native or denatured, monomer or aggregate, etc). A simple physical interpretation of the sign associated with the above expression is that when $P_{jk}^{i} > 0$ the local ratio of *j* to *k* molecules around a central *i* molecule is larger than the bulk ratio of *j* to *k* molecules. In binary systems, the preferential interaction is defined by

$$P_{cw}^{c} = N_{cc} - \frac{\rho_c}{\rho_w} N_{cw} = \rho_c (G_{cc} - G_{cw}) = -\frac{\rho_c}{\rho_w} P_{wc}^{c}. \tag{5.2}$$

where water molecules and cosolvent are described by *w* and *c*, respectively. In both of the above equations G_{ij} is the KB integral between species *i* and *j*. Excess coordination numbers are then

defined as $N_{ij} = \rho_j G_{ij}$.

Chemical Equilibrium

In a binary system, there will be a primary solvent (*w*) and then a cosolvent (*c*) such as glycine, diglycine, or triglycine. At the same time, the polyglycine molecules can be described in terms of aggregates. Hence, we have an equilibrium process which can be described by an equilibrium constant *K*.

$$nM \rightarrow A,\ K = \frac{\rho_A}{\rho_M^n} \tag{5.3}$$

The equilibrium constant can be defined in several ways using a variety of concentration scales. We chose *K* in terms of number densities or molarities. This equilibrium can be affected by a variety of factors, as outlined in chapter 4, all of which can be described using KB integrals.

Closed System

Our initial focus will be on fully closed systems at constant pressure. One obtains the following expression for changes in K,

$$\begin{aligned} \left(\frac{\partial \ln K}{\partial \ln m_c}\right)_{T,p} &= (P_{cw}^A - nP_{cw}^M)\mu_{cc} \\ \mu_{cc} &= \frac{1}{RT}\left(\frac{\partial \mu_c}{\partial \ln m_c}\right)_{T,p} = \frac{1}{RT}\left(\frac{\partial \mu_c}{\partial \ln N_c}\right)_{T,p} \end{aligned} \tag{5.4}$$

due to the addition of more peptide (*c*). Reference to the stability requirements for solutions indicates that $\mu_{cc} \geq 0$.[29] Consequently, on increasing the concentration of the peptide the association process is favored when $P_{cw}^A - nP_{cw}^M$ is positive. In a closed system the affinity of all

species (*c* and *w*) for the M and A forms contribute to the overall effect. These contributions are expressed relative to the primary solvent (*w*).

Semi-open System

If the system corresponds to that of a real cell, or some approximation to a real cell, then it may be open to water molecules. Considering a change in the concentration of a aggregating peptide (*c*) in a system open to water we find,[28]

$$\left(\frac{\partial \ln K}{\partial \ln \rho_c}\right)_{T,\mu_w} = \left(N_{Ac} - nN_{Mc}\right)\mu_{cc}^{'} \tag{5.4}$$

where the corresponding chemical potential derivative is defined,

$$\mu_{cc}^{'} = \frac{1}{RT}\left(\frac{\partial \mu_c}{\partial \ln \rho_c}\right)_{T,\mu_w} \tag{5.5}$$

The different expressions obtained for open and closed systems (Equations 5.3 and 5.4) suggest that the changes in aggregation behavior may also be different. Here, we would like to investigate this issue.

Kirkwood-Buff force fields

In this work we use our previously developed force fields described in chapter 3 as described in Figure 5.1. For instance, diglycine is composed of three different group; the N-terminal, C-terminal, and peptide bonds. The force field for each part has been developed separately. The parameters for the N-terminus were based on several amine salts; NH_4Cl, NH_4Br, and CH_3NH_3Cl. The C-terminal parameters were obtained by using a model for CH_3COONa. For the peptide bond, Kang and Smith have already published the required parameters for amides.[14]

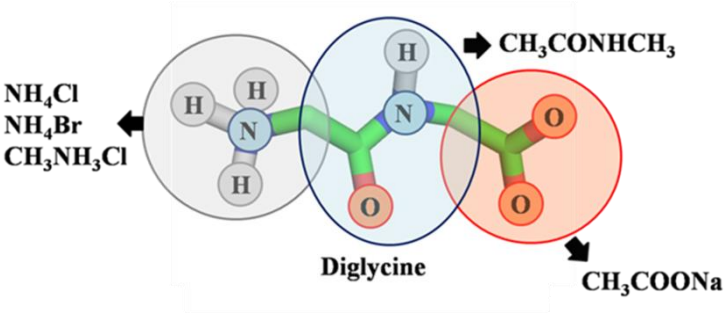


Figure 5.1 The strategy for the development of force field parameters. NH_4Cl, NH_4Br, and CH_3NH_3Cl are model for the N-terminus, and CH_3COONa corresponds to the C-terminus. The $CH_3CONHCH_3$ molecule describes the peptide group which has been developed previously by Smith and Kang[14].

Table 5.1 shows the Lennard-Jones parameters and charge distributions used in our simulations using the approach outlined in Figure 5.1.

Table 5.1 Nonbonded force field parameters for the KBFF model

	Atom	σ (nm)	ε (kJ/mol)	q (e)
N-terminal	N	0.3370	0.5620	+0.5
	H	0.8672	0.3748	0.0
	CH2	0.4170	0.3770	+0.5
Amide	C	0.3360	0.3300	+0.62
	O	0.3100	0.5600	-0.62
	N	0.3110	0.500	-0.7
	H	0.1580	0.0880	+0.36
	CH2	0.4170	0.3770	+0.34
C-terminal	C	0.3360	0.3300	+1.0
	O	0.3500	0.6047	-1.0
Walls	LJB	0.3	0.01	0.0

Table 5.2, Table 5.3, and Table 5.4 presents the bonded parameters used for the glycine, diglycine, and triglycine aqueous solutions by the KBFF model, respectively. The parameters for the bonds, angles, and improper dihedrals have been obtained from the GROMOS force fields.[30-31]

Table 5.2 Bonded parameters for aqueous glycine solution according to the KBFF models. Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
H-N	0.1	Constraint	
N-CH_2	0.1468		
CH_2-C	0.1520		
C-O	0.1250		
Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
H-N-H	109.5	334.72	
H-N-CH_2	109.5	376.56	
N-CH_2-C	114.1	502.10	
CH_2-C-O	117.0	502.08	
O-C-O	126.0	502.08	
Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
H-N-CH_2-C	0.0	4.0002	3
N-CH_2-C-O	0.0	1.0002	6
Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
C-CH_2-O-O	0.0	167.36	

Table 5.3 Bonded parameters for aqueous diglycine solution according to the KBFF models. Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})	
H-N	0.1000		
N-CH_2	0.1468		
CH_2-C	1.5300		
C=O	1.2300		
C-N	1.3300	Constraint	
N-H	1.0000		
N-CH_2	1.4700		
CH_2-C	0.1520		
C-O	0.1250		
Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})	
H-N-H	109.5	334.72	
H-N-CH_2	109.5	376.56	
N-CH_2-C	109.5	376.56	
CH_2-C=O	121.0	502.10	
O=C-N	124.0	502.10	
C-N-H	123.4	292.90	
C-N-CH_2	122.0	502.10	
H-N-CH_2	115.0	376.60	
N-CH_2-C	109.5	376.56	
CH_2-C-O	117.0	502.08	
O-C-O	126.0	502.08	
Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
H-N-CH_2-C	0.0	3.77	3
N-CH_2-C-N	0.0	1.00	6
CH_2-C-N-CH_2	180	33.5	2

C-N-CH_2-C	180	1.00	6
N-CH_2-C-O	0.0	1.00	6
Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
C-CH_2-N-O	0.0	167.36	
N-C-CH_2-H	0.0	167.36	
C-CH_2-O-O	0.0	167.36	

Table 5.4 Bonded parameters for aqueous triglycine solution according to the KBFF models. Potential functions are: angles, V_θ=1/2 $k_\theta(\theta - \theta_0)^2$; dihedrals, $V_\varphi = k_\varphi$ [1 + cos(nφ - δ)]; and impropers, V_ω=1/2 $k_\omega(\omega - \omega_0)^2$.

Bonds	b_o (nm)	k_b (kJ mol^{-1} nm^{-2})
H-N	0.1000	Constraint
N-CH_2	0.1468	
CH_2-C	1.5300	
C=O	1.2300	
C-N	1.3300	
N-H	1.0000	
N-CH_2	1.4700	
CH_2-C	1.5300	
C=O	1.2300	
C-N	1.3300	
N-H	1.0000	
N-CH_2	1.4700	
CH_2-C	0.1520	
C-O	0.1250	
Angles	θ_o (deg)	k_θ (kJ mol^{-1} rad^{-2})
H-N-H	109.5	334.72
H-N-CH_2	109.5	376.56
N-CH_2-C	109.5	376.56
CH_2-C=O	121.0	502.10

O=C-N	124.0	502.10	
C-N-H	123.4	292.90	
C-N-CH_2	122.0	502.10	
H-N-CH_2	115.0	376.60	
N-CH_2-C	109.5	376.56	
CH_2-C=O	121.0	502.10	
O=C-N	124.0	502.10	
C-N-H	123.4	292.90	
C-N-CH_2	122.0	502.10	
H-N-CH_2	115.0	376.60	
N-CH_2-C	109.5	376.56	
CH_2-C-O	117.0	502.08	
O-C-O	126.0	502.08	
Proper Dihedrals	ϕ_s (deg)	k_ϕ (kJ mol^{-1})	multiplicity
H-N-CH_2-C	0.0	3.77	3
N-CH_2-C-N	0.0	1.00	6
CH_2-C-N-CH_2	180	33.5	2
C-N-CH_2-C	180	1.00	6
N-CH_2-C-N	0.0	1.00	6
CH_2-C-N-CH_2	180	33.5	2
C-N-CH_2-C	180	1.00	6
N-CH_2-C-O	0.0	1.00	6
Improper Dihedrals	ω_0 (deg)	k_ω (kJ mol^{-1} rad^{-2})	
C-CH_2-N-O	0.0	167.36	
N-C-CH_2-H	0.0	167.36	
C-CH_2-N-O	0.0	167.36	
N-C-CH_2-H	0.0	167.36	
C-CH_2-O-O	0.0	167.36	

Semi-open System Model

In order to study a model of a semi-permeable membrane, which acts as a mimic of a real cellular environment, we employ a wall of simple frozen particles which is permeable to water only. Then we attempt to reproduce the experimentally observed osmotic pressure changes in this system.

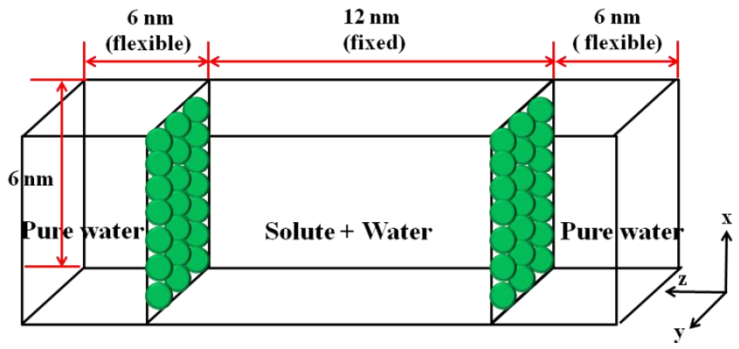


Figure 5.2 The simulation box used for the semi-open systems. The green balls indicate the semipermeable membranes for the water molecule which mimic a cell environment. Each wall has 400 particles.

The green wall "particles" carry no charge and are frozen in a fixed position. They do not interact with water molecules. However, they interact with solute molecules, but only through a simple repulsive LJ potential. Hence, water molecules move freely in and out of the central region, while the peptide molecules are constrained to remain within the central region.

Molecular dynamics simulations

All molecular dynamic simulations of polyglycine aqueous solutions were performed using the with the SPC/E water model[32] in the isothermal isobaric (*NpT*) ensemble at 300 K and 1 atm as implemented in the GROMACS program (v3.3.1).[30-31] A time-step of 2 fs was used and the geometry of the water molecules was constrained using SETTLE.[33] All bonds were constrained using LINCS.[34] A twin range cut-off of 0.8 nm/1.5 nm was employed with a nonbonded pair list update of every 10 steps. The weak coupling technique was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively.[35] In order to evaluate electrostatic interactions, the particle mesh Ewald technique (PME) was used.[36] The initial cubic boxes of different solutions have been generated by adding water molecules and polyglycine molecules until the required concentration was obtained. Configurations were saved every 0.1 ps for analysis.

Results

A summary of the set of simulations and some preliminary results are summarized in Table 5.5 for the closed systems.

Table 5.5 Summary of the MD simulations of polyglycine aqueous solution in closed systems. All simulations were performed at 300 K and 1atm in the *NpT* ensemble. Symbols are N_w, number of water molecules; N_s, number of solutes; V, average simulation volume; m_s, solute molality; C_s, solute molarity; ρ, mass density; E_{pot}, average total potential energy per molecule ($N_s + N_w$); and T_{sim}, total simulation time.

	N_s	N_w	m_s (mol/Kg)	V (nm^3)	C_s (mol/l)	ρ (g/cm^3)	E_{pot} (kJ/mol)	T_{sim} (ns)
GLY	90	16520	1.0	215.809	0.95	124	-49.04	11
	124	6898	1.0	215.809	0.95	124	-59.42	11
	238	6611	2.0	215.209	1.84	238	-65.45	11
	343	6341	3.0	214.840	2.65	343	-71.32	11
GLYGLY	90	16520	0.3	507.799	0.29	1.005	-49.04	11
	290	16040	1.0	518.764	0.93	1.048	-59.49	11
	420	15560	1.5	521.235	1.43	1.070	-65.70	11
GLYGLYGLY	90	16520	0.3	513.630	0.29	1.017	-51.62	11

The corresponding simulation results for the semi-open systems are summarized in Table 5.6. Because semi-permeable membranes allow the water molecules to freely move through the membrane the molality of the central region between the membranes is not constant. Therefore, the solute molarities have been used as a measure of concentration, instead of molalities, when determining the osmotic pressure of polyglycine aqueous solution in the semi-open systems.

Table 5.6 Summary of the MD simulations of polyglycine aqueous solution in semi-open system. All simulations were performed at 300 K and 1 atm in the $\mu_w pT$ ensemble. Symbols are N_w, number of water molecules; N_s, number of solutes; V, average simulation volume of the box between membranes; m_s, approximate solute molality of the box between membranes; C_s, solute molarity of the box between membranes; and T_{sim}, total simulation time.

	N_s	N_w	m_s (mol/Kg)	V (nm^3)	C_s (mol/l)	π (atm)	T_{sim} (ns)
NaCl	500	28314	2.0	454	1.83	72	6
	960	27184	4.0	454	3.52	150	6
	1380	26051	6.0	454	5.05	226	6
GLY	76	28648	0.3	454	0.28	7.9	11
	248	28236	1.0	454	0.91	19.6	11
	476	27062	2.0	454	1.84	52.2	11
	686	25956	3.0	454	2.65	60.2	11
GLYGLY	76	28510	0.3	454	0.28	6.2	11
	240	27796	1.0	454	0.88	17.6	11
	384	27324	1.5	454	1.27	27.3	11
GLYGLYGLY	76	28364	0.3	454	0.28	7.3	11

The radial distribution functions (rdfs) obtained from the 0.3 m simulations in closed systems are shown on Figure 5.3 for the polyglycines with n = 1 - 3. The glycine to glycine rdf displayed a large first and a significant second peak but the diglycine to diglycine and triglycine to triglycine rdfs present broad peaks, which do not distinguish between the first peak and the second peak. The g_{cc} rdfs indicate that the interaction between glycine and glycine is stronger than those between di or triglycine and di or triglycine, because the distance between the center of mass and center of mass for the glycine is the closest and the first peak is the highest.

However, the g_{cw} rdfs indicate the interaction between glycine and water is stronger than those between di or triglycine and water, because the distance between the center of mass and center of mass for the glycine is the closest and the first peak is the highest. Therefore, glycine appears to have the largest interactions for both. This indicates that the larger polyglycine molecules do not allow other solute molecules to approach closely. All rdfs approach unity beyond 1.5 nm.

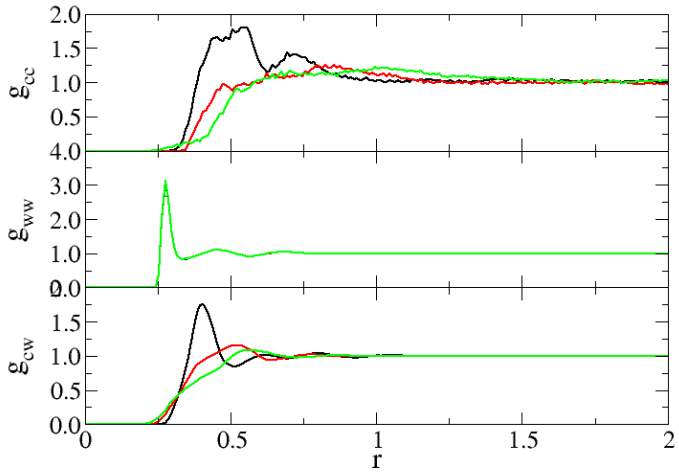


Figure 5.3 Radial distribution functions from 0.3 m solutions obtained from the closed glycine (black lines), diglycine (red lines), triglycine (green lines) simulations.

The simulated and experimental excess coordination numbers, N_{ij}, are shown in Figure 5.4 for the polyglycine as a function of molality as obtained for the closed systems. The KBFF model quantitatively reproduces the experimental data. The water-water excess coordination numbers, N_{ww}, are represented by green lines and symbols on Figure 5.4 and increase with solute concentration and solute size. Figure 5.4 also shows that the solute-solute excess coordination numbers (black lines) also do not vary significantly from solute to solute, at least compared to the variation of the solute-water excess coordination numbers (red lines). This indicates that

polyglycine molecules exclude water molecules from their vicinity area and this effect increases with n.

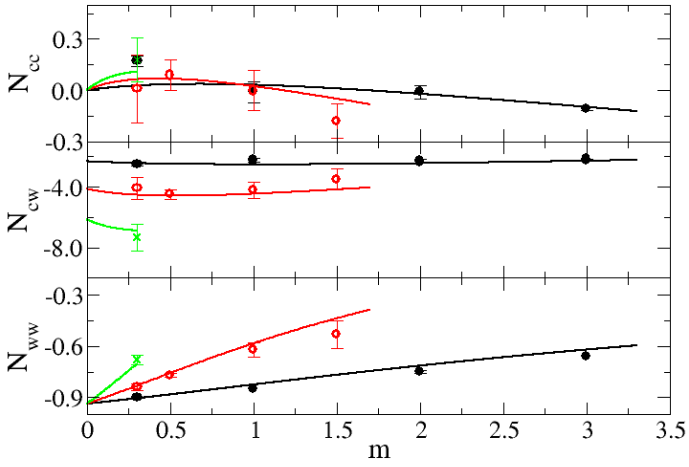


Figure 5.4 Excess coordination numbers as a function of molar salt concentration: The glycine (black lines), diglycine (red dot lines), and triglycine (green dash lines) are obtained from a KB analysis using experimental activity coefficient[37] and density. The glycine (black ●), diglycine (red ○), and triglycine (green x) are obtained from simulations.

In Figure 5.5, the simulated activity derivatives a_{cc} as a function of molality are compared to the experimental values. The KBFF model reproduced the correct increase in a_{cc} with concentration as displayed by the experimental data. The values of a_{cc} play an important role for describing the accurate force field because KBIs are more sensitive to activity coefficients than other physical properties such as the compressibility and the density.[38] Hence, the most accurate simulations require parameters that yield accurate activity derivatives.

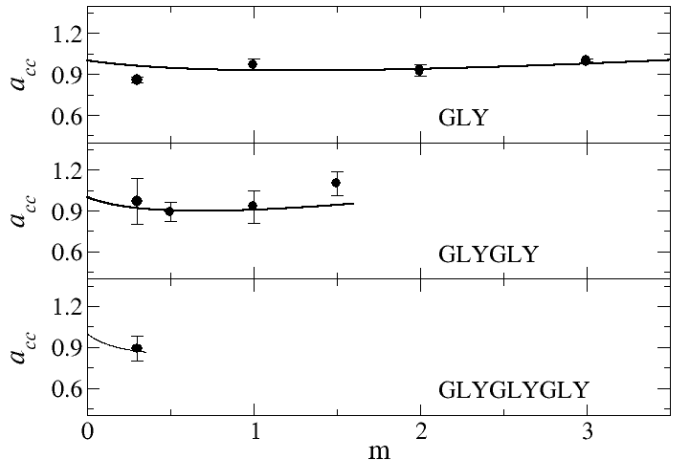


Figure 5.5 Activity derivatives as a function of molality for closed systems. Lines are obtained from a KB analysis using experimental activity coefficient experimental data[37] and dots correspond to the KBFF model

Figure 5.6 shows the partial molar volumes of water and of the polyglycines. The partial molar volume of the solute increases, while that of water is almost constant, as the concentration increases. Also, as the size of the glycine peptide increases the partial molar volume of the peptide increases, while the partial molar volumes of water in three different systems are essentially constant. The KBFF reproduces the experimental data quantitatively.

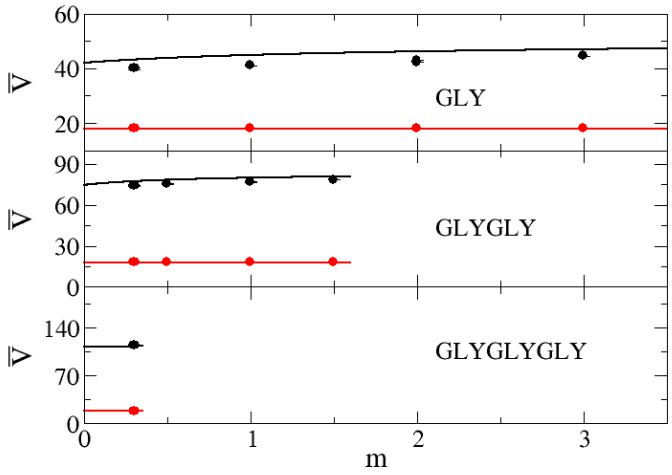


Figure 5.6 Partial molar volumes as a function of molality. Lines are obtained from a KB analysis using experimental activity coefficient[37] and experimental density data,[39] and dots correspond to the KBFF model. The black lines represent the partial molar volume of salts and the red dotted lines indicate partial molar volume of water. The dots (●) represent partial molar volume of salts and the dots (red ○) indicate partial molar volume of water.

In order to quantify the polyglycine self interaction in dilute solutions, the preferential interaction based on the KB integrals can be used. The quantity G_{cc}-G_{cw} describes the preferential interaction of polyglycine with other polyglycine molecules as function of composition.[25-27] A positive preferential interaction indicates that the ratio of polyglycine to water in the vicinity of a polyglycine molecule exceeds the ratio of polyglycine to water in the bulk solution, while a negative number indicates that the ratio of polyglycine to water in the bulk exceeds the ratio of polyglycine to water in the vicinity of polyglycine. The preferential interaction at infinite dilution (G_{cc}^{∞}-G_{cw}^{∞}) is potentially very informative and describes the preferential interaction between two polyglycine molecules in pure water. Experimentally, the preferential interaction of polyglycine increase as n increase. The values of G_{cw} for all polyglycine molecules observed in

the simulations reproduce the experimental trend of peptide increase well. Unfortunately, the simulated value of G_{cc} is overestimated for glycine molecules, underestimated for the diglycine molecules, and yet correct for the triglycine molecules compared to the experimental data. Because the low concentration (0.3 m) simulations involve a low number of solutes, there is not enough statistics to obtain reliable simulation data for G_{cc} (as indicated by the large error bars in Figure 5.7). This results in the simulated preferential interactions ($G_{cc} - G_{cw}$) of polyglycine molecules having a slight mismatch with the experimental data.

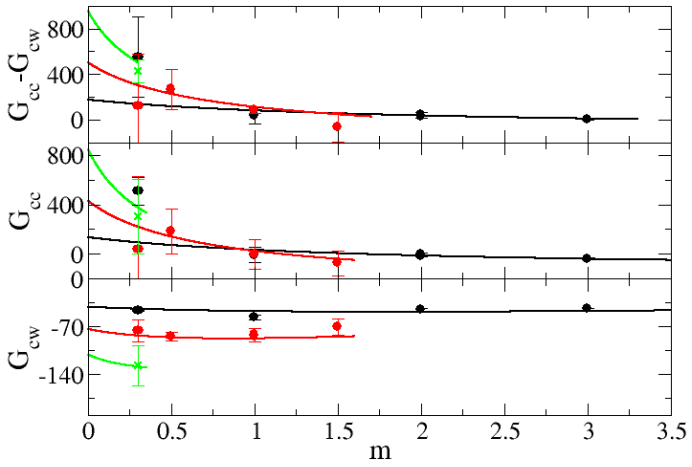


Figure 5.7 Preferential interaction (G_{cc} - G_{cw}) and Kirkwood-Buff integrals (G_{cc} and G_{cw}) as a function of molar salt concentration in closed system. The glycine (black lines), diglycine (red dot lines), and triglycine (green dash lines) are obtained from a KB analysis using experimental activity coefficient and density.[37, 39] The glycine (black ●), diglycine (red ○), and triglycine (green x) are obtained from simulations.

The radial distribution functions (rdfs) obtained from the 0.3 m simulations for the polyglycine peptides are shown on Figure 5.8 for closed systems, and in Figure 5.9 for the semi-open systems. The labels NT, OT, and OW correspond to the nitrogen atoms of the N-terminus, the oxygen atoms of the C-terminus, and the oxygen atoms of water molecules, respectively. The rdf (NT to OT) in glycine displays a large first and a significant second peak, but the rdf (NT to OT) in diglycine and triglycine present as broad peaks which are overlapped by the first peak and the second peak in both systems. The results suggest that the interactions between the N-terminal and C-terminal groups increase as n increases.

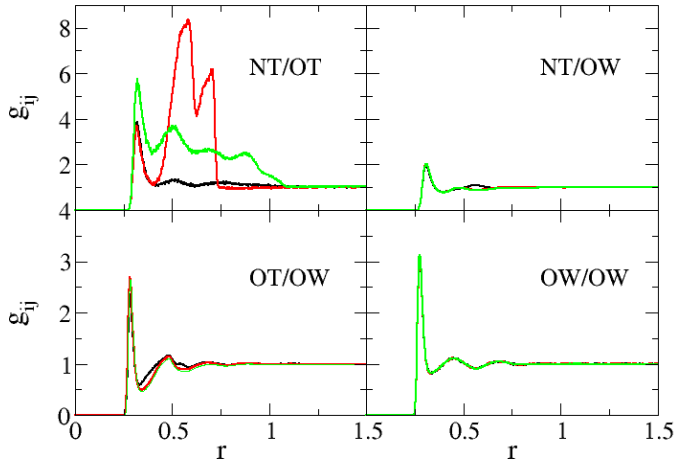


Figure 5.8 Radial distribution functions of 0.3 m solutions obtained from the glycine (black lines), diglycine (red lines), triglycine (green lines) simulations in closed system. Nitrogen of N-terminal, oxygen of C-terminal, oxygen of water is denoted by the symbols NT, OT, and OW, respectively.

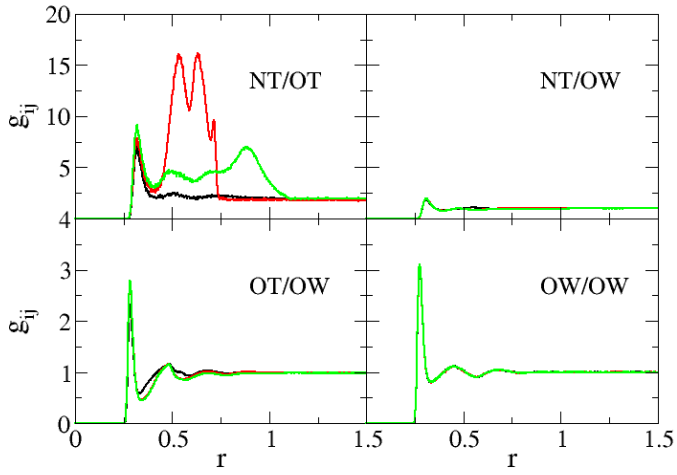


Figure 5.9 Radial distribution functions of 0.3 m solutions obtained from the glycine (black lines), diglycine (red lines), triglycine (green lines) simulations in semi-open system. Nitrogen of N-terminal, oxygen of C-terminal, oxygen of water are denoted by the symbols NT, OT, and OW, respectively.

The first shell coordination numbers, n_{ij}, as well as the distances to the first rdf maximum, R_{max}, and the first rdf minimum, R_{min}, were calculated from the corresponding rdfs as a function of the solution molarity and are presented in Table 5.7 for the closed systems, and in Table 5.8 for the semi-open systems. As n increases the values of R_{min} (+/-) for the polyglycines also increases, while the interactions involved with water molecules display similar values. In addition, the coordination numbers (n_{ij}) of the polyglycines to the given position, R_{min} (+/-) at same composition (0.3m) also increase with n in both open and semi-open systems. This suggests that the self-association of polyglycine increases with n, and that the interaction between the N-terminal and C-terminal groups is the dominant interaction leading to the self-association.

Table 5.7 First shell coordination number (n_{ij}) of polyglycine aqueous solutions in closed systems: R_{max} and R_{min} are the distances (nm) to the first maximum and minimum of the radial distribution functions. N-terminal, C-terminal, and the water oxygen are denoted by the symbols +, -, and o, respectively

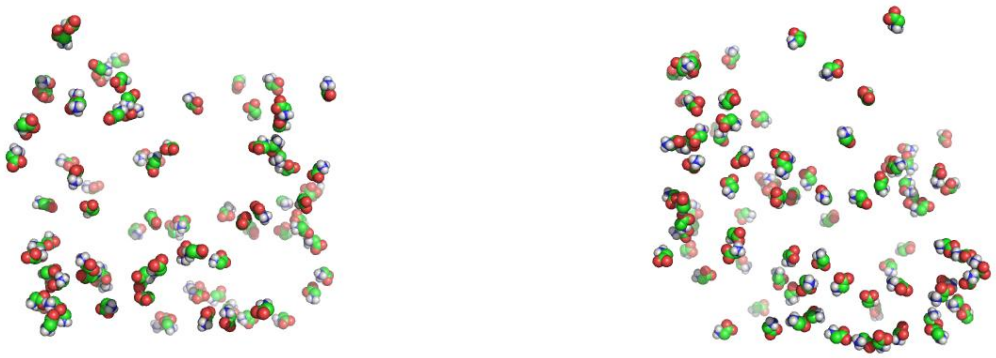
		m_s	+/-	+/o	-/o	o/o
GLY	R_{max}		0.32	0.31	0.28	0.27
	R_{min}		0.42	0.39	0.33	0.33
	n_{ij}	0.3	0.15	6.21	3.05	4.29
		1.0	0.39	6.10	2.97	4.20
		2.0	0.63	5.98	2.89	4.07
		3.0	0.86	5.89	2.83	3.95
GLYGLY	R_{max}		0.32	0.31	0.28	0.27
	R_{min}		0.75	0.39	0.33	0.33
	n_{ij}	0.3	2.69	6.34	3.31	4.27
		1.0	4.04	6.18	3.22	4.14
		1.5	4.79	6.08	3.17	4.05
GLYGLYGLY	R_{max}		0.32	0.31	0.28	0.28
	R_{min}		1.09	0.39	0.34	0.34
	n_{ij}	0.3	4.04	6.25	3.41	4.62

Table 5.8 First shell coordination number (n_{ij}) of polyglycine aqueous solutions in open system. R_{max} and R_{min} are the distances (nm) to the first maximum and minimum of the radial distribution functions. N-terminal, C-terminal, and the water oxygen are denoted by the symbols +, -, and o, respectively.

		m_s	+/-	+/o	-/o	o/o
GLY	R_{max}		0.32	0.31	0.28	0.28
	R_{min}		0.42	0.39	0.34	0.34
	n_{ij}	0.3	0.13	6.13	3.29	4.67
		1.0	0.41	5.90	3.25	4.64
		2.0	0.66	5.80	3.17	4.58
		3.0	0.89	5.89	2.83	3.95
GLYGLY	R_{max}		0.32	0.31	0.28	0.28
	R_{min}		0.75	0.39	0.34	0.34
	n_{ij}	0.3	2.79	6.25	3.60	4.67
		1.0	4.17	6.10	3.48	4.61
		1.5	4.92	6.04	3.42	4.57
GLYGLYGLY	R_{max}		0.32	0.31	0.28	0.28
	R_{min}		1.10	0.39	0.33	0.34
	n_{ij}	0.3	4.24	6.20	3.39	4.66

Figure 5.10 and Figure 5.11 present snapshots of 0.3 m (a) glycine, (b) diglycine, and (c) triglycine aqueous solution in closed and in open systems after 11 ns MD simulation, shown without water molecules for clarity. Both figures indicate that the self-association of polyglycine in water increases as n increases. These figures appear to agree with the experimental data.

XZ plane YZ plane



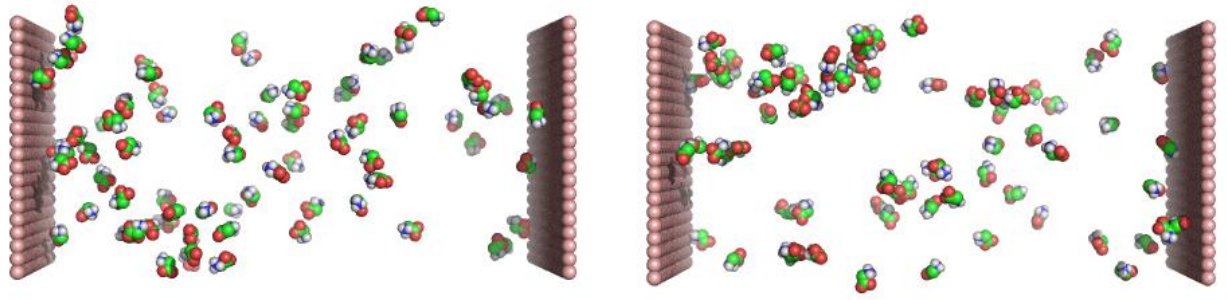
(a)

(b)

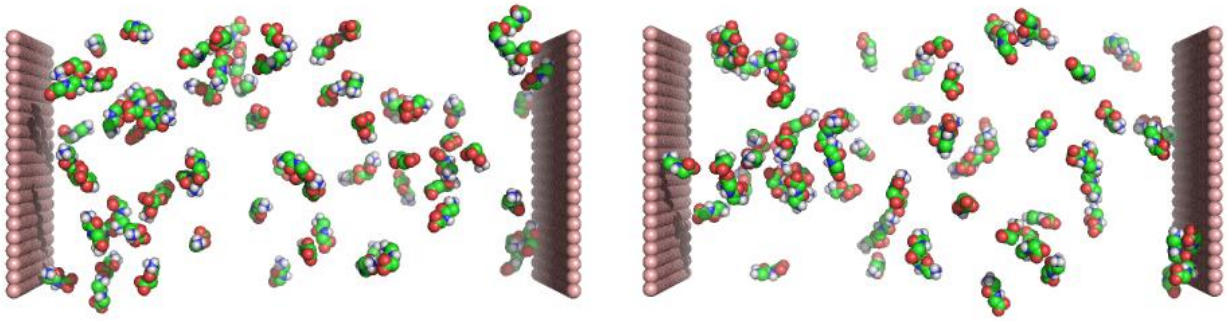
(c)

Figure 5.10 Snapshots of 0.3 m (a) glycine, (b) diglycine, and (c) triglycine aqueous solution in closed system after 11 ns MD simulation without water molecules for clarity: White, blue, red, and green balls indicate hydrogen, nitrogen, oxygen, and carbon atom, respectively.

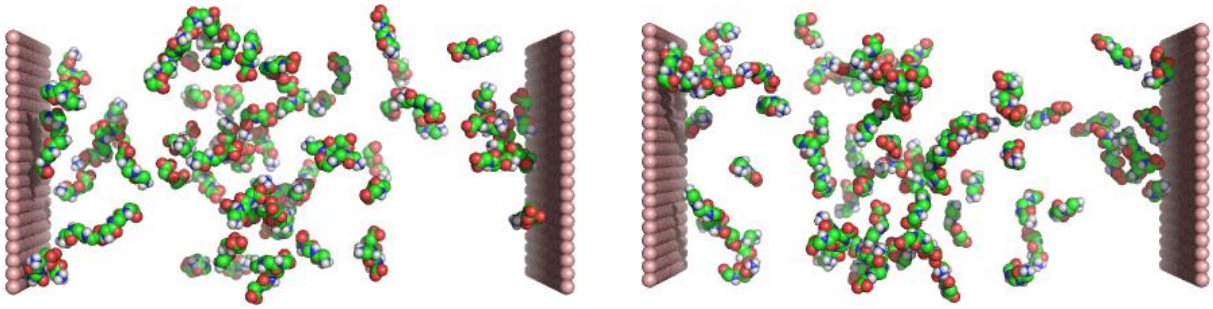
XZ plane YZ plane



(a)



(b)



(c)

Figure 5.11 Snapshots of 0.3 m (a) glycine, (b) diglycine, and (c) triglycine aqueous solution in open system after 11 ns MD simulation without water molecules for clarity: White, blue, red, and green balls indicate hydrogen, nitrogen, oxygen, and carbon atom, respectively.

In order to obtain the pressure profile through the semi-open systems we employed the technique used by Alejandre and Tildesley.[40] Figure 5.12 represents our initial studies of NaCl aqueous solutions used as a test case. The semi-permeable membrane approach composed of simple frozen particles reproduces the experimentally observed osmotic pressure changes as indicated below.

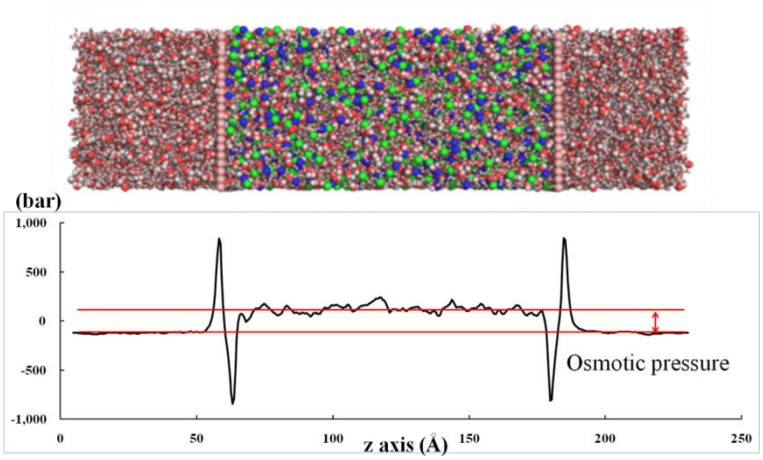


Figure 5.12 Snapshots of 6 m NaCl aqueous solution in closed system after 6 ns MD simulation (top). Water molecules are described by red balls (oxygen) and white balls (hydrogen). Blue, green, and pink balls indicate Cl^-, Na^+, frozen particle in the walls, respectively. The pressure profile of 6 m NaCl aqueous solutions along z-axis (bottom). Black line indicates the pressure profile and red lines describe the difference between the pressures of inside walls and of outside walls.

Figure 5.13 represents the osmotic pressures for the NaCl aqueous solution as the function of molarity. The osmotic pressures of KBFF models reproduce the experimental value well.

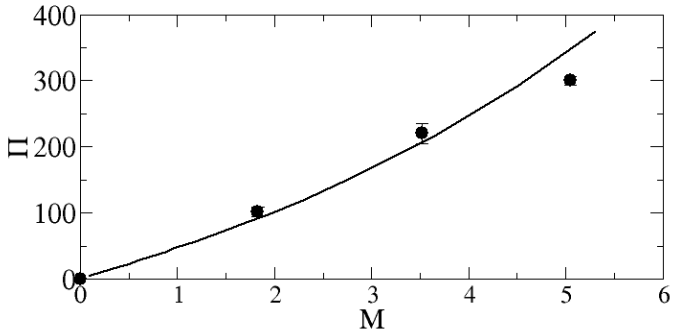


Figure 5.13 Osmotic pressure (bar) of NaCl aqueous solutions as function of salt molarity in semi-open systems. Lines are obtained from a experimental data,[41] dots correspond to the KBFF model.

For the polyglycine aqueous solutions in semi-open systems we used a different method to obtain the osmotic pressures, which is much simpler to implement. In this approach, the osmotic pressures for the polyglycines in water have been obtained by determining the sum of the forces acting on the walls.[16-18] Figure 5.14 represents the osmotic pressures for a series of polyglycine peptides in water as a function of peptide molarity. As the concentration increased the osmotic pressure of the solutions increased. The osmotic pressures exhibited by the KBFF models reproduce the experimental values very well.

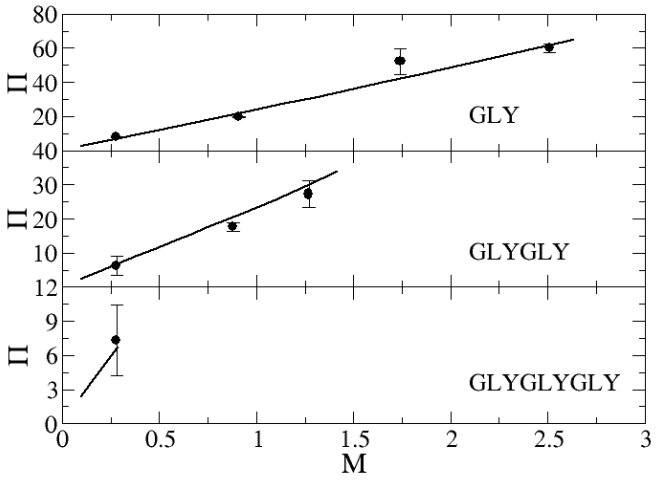


Figure 5.14 Osmotic pressure (bar) of polyglycine aqueous solutions as function of molarity. Lines are obtained from a experimental data,[42] dots correspond to the KBFF model.

Conclusions

A Kirkwood-Buff (KB) analysis of the experimental data indicates that self-association of glycine polypeptides increases with n. A Kirkwood-Buff (KB) model of polyglycine reproduced experimental KBIs well as n increase. Our studies of NaCl aqueous solutions using a semi-permeable membrane composed of simple frozen particles reproduced the experimentally observed osmotic pressure changes, while the studies of polyglycine aqueous solutions using the same approach also reproduced the experimentally observed osmotic pressure changes. This not only provides an approach to compare and contrast peptide aggregation in open and closed systems, but may also help to elucidate the dominant factors for the self-association of polyglycines.

References

1. J. D. Batchelor, A. Olteanu, A. Tripathy and G. J. Pielak, J Am Chem Soc **126** (7), 1958-1961 (2004).
2. R. Truant, R. S. Atwal, C. Desmond, L. Munsie and T. Tran, Febs J **275** (17), 4252-4262 (2008).
3. G. Colombo, P. Soto and E. Gazit, Trends Biotechnol **25** (5), 211-218 (2007).
4. A. Wang, A. D. Robertson and D. W. Bolen, Biochemistry-Us **34** (46), 15096-15104 (1995).
5. J. F. Carpenter, B. S. Kendrick, B. S. Chang, M. C. Manning and T. W. Randolph, Method Enzymol **309**, 236-255 (1999).
6. A. L. Fink, Fold Des **3** (1), R9-R23 (1998).
7. Y. S. Kim, S. P. Cape, E. Chi, R. Raffen, P. Wilkins-Stevens, F. J. Stevens, M. C. Manning, T. W. Randolph, A. Solomon and J. F. Carpenter, J Biol Chem **276** (2), 1626-1633 (2001).
8. Y. F. Maa and C. C. Hsu, Int J Pharm **140** (2), 155-168 (1996).
9. V. S. Pande, A. Y. Grosberg and T. Tanaka, Rev Mod Phys **72** (1), 259-314 (2000).
10. J. R. Kim, T. J. Gibson and R. M. Murphy, Biotechnol Progr **22** (2), 605-608 (2006).
11. S. K. Allison, S. P. Bates, J. Crain and G. J. Martyna, J Phys Chem B **110** (42), 21319-21326 (2006).
12. Y. A. Mantz, H. Gerard, R. Iftimie and G. J. Martyna, J Phys Chem B **110** (27), 13523-13538 (2006).
13. T. W. Whitfield, G. J. Martyna, S. Allison, S. P. Bates, H. Vass and J. Crain, J Phys Chem B **110** (8), 3624-3637 (2006).

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

15. S. E. McLain, A. K. Soper, I. Daidone, J. C. Smith and A. Watts, Angew Chem Int Edit **47** (47), 9059-9062 (2008).

16. S. Murad, Adsorption **2** (1), 95-101 (1996).

17. S. Murad, J. G. Powles and B. Holtz, Mol Phys **86** (6), 1473-1483 (1995).

18. S. Murad and J. G. Powles, J Chem Phys **99** (9), 7271-7272 (1993).

19. M. Henrichsen and R. L. Rowley, Fluid Phase Equilibr **137** (1-2), 75-85 (1997).

20. R. L. Rowley, T. D. Shupe and M. W. Schuck, Fluid Phase Equilibr **104**, 159-171 (1995).

21. R. L. Rowley, M. W. Schuck and J. C. Perry, Mol Phys **86** (1), 125-137 (1995).

22. R. L. Rowley, T. D. Shupe and M. W. Schuck, Mol Phys **82** (5), 841-855 (1994).

23. Y. Luo and B. Roux, J Phys Chem Lett **1** (1), 183-189 (2010).

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

25. M. Kang and P. E. Smith, Fluid Phase Equilibr **256** (1-2), 14-19 (2007).

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

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

28. M. B. Gee and P. E. Smith, J Chem Phys **131** (16), - (2009).

29. I. Prigogine and R. Defay, (Wiley, New York, 1954), pp. xxxii, 543 p.

30. D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, J Comput Chem **26** (16), 1701-1718 (2005).

31. E. Lindahl, B. Hess and D. van der Spoel, J Mol Model **7** (8), 306-317 (2001).

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

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

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

35. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, J Chem Phys **81** (8), 3684-3690 (1984).

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

37. M. K. Khoshkbarchi and J. H. Vera, Ind Eng Chem Res **35** (8), 2735-2742 (1996).

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

39. Q. Yuan, Z. F. Li and B. H. Wang, J Chem Thermodyn **38** (1), 20-33 (2006).

40. J. Alejandre, J. W. Emsley, D. J. Tildesley and P. Carlson, J Chem Phys **101** (8), 7027-7036 (1994).

41. R. A. Robinson and R. H. Stokes, *Electrolyte solutions*, 2 ed. (London Butter Worths, 1959).

42. E. N. Tsurko, R. Neueder and W. Kunz, J Solution Chem **36** (5), 651-672 (2007).