

MOLECULAR DYNAMICS SIMULATIONS AND THEORY OF INTERMOLECULAR
INTERACTIONS IN SOLUTIONS

by

MYUNGSHIM KANG

B.S., Seoul National University, Korea, 1999
M.S., Seoul National University, Korea, 2001

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 Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Abstract

In the study of biological systems, molecular dynamics (MD) simulations have played an important role in providing atomic details for phenomena of interest. The force field used in MD simulations is a critical factor determining the quality of the simulations. Recently, Kirkwood-Buff (KB) theory has been applied to study preferential interactions and to develop a new force field. KB theory provides a path from quantities determined from simulation data to the corresponding thermodynamic data. Here we combine KB theory and molecular simulations to study a variety of intermolecular interactions in solution. First, recent results concerning the formulation and evaluation of preferential interactions in biological systems in terms of KB integrals are presented. In particular, experimental and simulated preferential interactions of a cosolvent with a biomolecule in the presence of water are described. Second, a force field for the computer simulation of aqueous solutions of amides is presented. The force field is designed to reproduce the experimentally observed density and KB integrals for N-methylacetamide (NMA), allowing for an accurate description of the NMA activity. Other properties such as the translational diffusion constant and heat of mixing are also well reproduced. The force field is then extended to include N,N'-dimethylacetamide and acetamide with good success. The models presented here provide a basis for an accurate force field for peptides and proteins. Comparison between the developed KB force fields (KBFF) and existing force fields is performed for amide and glycine and proves that the KBFF approach is competitive. Also, explicit expressions are developed for the chemical potential derivatives, partial molar volumes, and isothermal compressibility of solution mixtures involving four components at finite concentrations using the KB theory of solutions. A general recursion relationship is also provided which can be used to generate the chemical potential derivatives for higher component solutions. Finally, a pairwise

preferential interaction model (PPIM), described by KB integrals is developed to quantify and characterize the interactions between functional groups observed in peptides.

MOLECULAR DYNAMICS SIMULATIONS AND THEORY OF INTERMOLECULAR
INTERACTIONS IN SOLUTIONS

by

MYUNGSHIM KANG

B.S., Seoul National University, 1999
M.S., Seoul National University, 2001

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Science

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2009

Approved by:

Major Professor
Dr. Paul E. Smith

Copyright

MYUNGSHIM KANG

2009

Abstract

In the study of biological systems, molecular dynamics (MD) simulations have played an important role in providing atomic details for phenomena of interest. The force field used in MD simulations is a critical factor determining the quality of the simulations. Recently, Kirkwood-Buff (KB) theory has been applied to study preferential interactions and to develop a new force field. KB theory provides a path from quantities determined from simulation data to the corresponding thermodynamic data. Here we combine KB theory and molecular simulations to study a variety of intermolecular interactions in solution. First, recent results concerning the formulation and evaluation of preferential interactions in biological systems in terms of KB integrals are presented. In particular, experimental and simulated preferential interactions of a cosolvent with a biomolecule in the presence of water are described. Second, a force field for the computer simulation of aqueous solutions of amides is presented. The force field is designed to reproduce the experimentally observed density and KB integrals for N-methylacetamide (NMA), allowing for an accurate description of the NMA activity. Other properties such as the translational diffusion constant and heat of mixing are also well reproduced. The force field is then extended to include N,N'-dimethylacetamide and acetamide with good success. The models presented here provide a basis for an accurate force field for peptides and proteins. Comparison between the developed KB force fields (KBFF) and existing force fields is performed for amide and glycine and proves that the KBFF approach is competitive. Also, explicit expressions are developed for the chemical potential derivatives, partial molar volumes, and isothermal compressibility of solution mixtures involving four components at finite concentrations using the KB theory of solutions. A general recursion relationship is also provided which can be used to generate the chemical potential derivatives for higher component solutions. Finally, a pairwise

preferential interaction model (PPIM), described by KB integrals is developed to quantify and characterize the interactions between functional groups observed in peptides.

Table of Contents

List of Figures	xii
List of Tables	xvii
Acknowledgements.....	xix
Dedication.....	xx
CHAPTER 1 - Introduction	1
1.1 General Introduction.....	1
1.2 Molecular Simulation.....	2
1.3 Force Fields for the Simulation of Biological Systems	4
1.4 Kirkwood-Buff Theory	7
1.5 Kirkwood-Buff Derived Force Field	12
1.6 Summary.....	14
1.7 References.....	16
CHAPTER 2 - Preferential Interaction Parameters in Biological Systems by Kirkwood–Buff Theory and Computer Simulation	23
Abstract.....	23
2.1 Introduction.....	24
2.2 Background and Theory.....	25
2.3 Methods.....	28
2.4 Results.....	30
2.5 Conclusions.....	36
2.6 References.....	38

CHAPTER 3 - A Kirkwood-Buff Derived Force Field for Amides	41
Abstract	41
3.1 Introduction	42
3.2 Methods	43
3.3 Results	49
3.4 Conclusions	64
3.5 References	65
Supplementary Data for Chapter 3	70
CHAPTER 4 - A Comparison of Force Fields for Amides and Glycine	72
Abstract	72
4.1 Introduction	73
4.2 Methods	74
4.3 Results	75
4.4 Conclusion	84
4.5 References	86
CHAPTER 5 - Kirkwood-Buff Theory of Four and Higher Component Mixtures	89
Abstract	89
5.1 Introduction	90
5.2 Theory	91
5.2.1 General Approach	92
5.2.2 Four Component Mixtures	99
5.2.3 Three Component Mixtures	101
5.2.4 Two Component Mixtures	102

5.2.5	Open and Semi-Open Systems.....	103
5.2.6	A General Recursion Relationship for the Chemical Potential Derivatives	104
5.2.7	Five Component Systems	105
5.3	Discussion.....	106
5.4	Conclusions.....	108
5.5	References.....	109
CHAPTER 6 - Pairwise Preferential Interaction Model.....		111
Abstract.....		111
6.1	Introduction.....	112
6.2	Pairwise Preferential Interaction Model (PPIM) Approach.....	113
6.2.1	Thermodynamics of Solutions and KB Theory	113
6.2.2	Preferential Interactions	114
6.2.3	Experimental Data	115
6.2.4	Decomposition Approach	119
6.2.5	Evaluation of Group Contributions.....	120
6.3	Discussion.....	122
6.4	Conclusion	124
6.5	References.....	125
CHAPTER 7 - Summary and Future Work		127
Appendix A - Personal Publications and Copies of Permission Letter from the Publisher.....		129
A.1	Personal Publication List	129
A.2	Copies of the Permission Letter from the Publisher	131
A.2.1	ELSEVIER LICENSE	131

A.2.2 WILEY PERIODICALS, Inc., A WILEY COMPANY	136
A.2.3 AMERICAN INSTITUTE OF PHYSICS LICENSE	138

List of Figures

- Figure 1.1 The role of Kirkwood-Buff theory 9
- Figure 1.2 Radial distribution function (rdf). The rdf displays the local solution structures, including solvation shells. As the distance between species i and j , r_{ij} , gets larger, the rdf goes to unity, meaning the distribution becomes similar to the bulk distribution. 10
- Figure 1.3 An example of KB integral G_{ij} as a function of integration distance r (nm) between species i and j . This KB integral corresponds to the rdf displayed in Figure 1.2. 11
- Figure 1.4 An example of excess coordination number N_{II} for the entire range of the composition. The sign and of N_{ij} indicates the feature of intermolecular interactions between species i and j : positive N_{ij} indicates attractive interactions between i and j and negative N_{ij} represent repulsive interactions. The graph displays the N_{II} , and therefore the self interactions between species I 's change as the composition of solution mixture changes. . 12
- Figure 2.1 The structure of native Lysozyme in 8M urea at pH 7. The cartoon model displays Lysozyme colored by chain, and the green stick model indicates urea. The water molecules are not displayed for clarity. 31
- Figure 2.2 The simulated distance dependent PI of 8M urea with native Lysozyme. As a function of urea force field at pH 7 according to Equations 2.10 and 2.11 (top). As a function of urea force field and pH using Equation 2.11 (bottom). The experimentally observed PIs are 16 at pH 7 and -10 at pH 2. 32
- Figure 2.3 The urea (top) and water (bottom) rdfs as a function of the closest distance (r) to any protein atom for the two different urea force fields. The simulated data correspond to 8M urea and a protein at pH 7. 33

Figure 2.4 The time history of the 8M urea PI corresponding to the KBFF model of urea and a protein at pH 7. The value of Γ_{23} was determined for $R = R_c = 1.0$ nm using Equation 2.11. 34

Figure 3.1 Center of mass based rdfs for $x_2 = 0.1$ solutions of NMA, DMA, and ACT..... 48

Figure 3.2 Excess coordination numbers (N_{ij}) as a function of composition for NMA solutions. The lines corresponding to the KBFF and CHARMM force fields were obtained after determining the simulated values of f_{22} and then obtaining the fitting constants for Equation 3.2 via a fit of the simulated data to Equation 3.4. 50

Figure 3.3 The snapshot of the simulated aqueous NMA solution at $x_2=0.2$ with KBFF (left) and CHARMM (right). While the simulation with the KBFF model displayed random distribution, the simulation with CHARMM displayed self aggregation of NMA molecules. 51

Figure 3.4 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for NMA solutions. The crosses were obtained directly from the KB integrals. The dashed lines were obtained by fitting the simulated excess volumes to Equation 3.2 and then applying Equation 3.3. 52

Figure 3.5 Activity derivative (f_{22}) and excess Gibbs energy (G) and enthalpy of mixing (H) for NMA solutions. The solid lines correspond to the experimental data (H at 308K), the crosses and dashed lines to the KBFF model, and the diamonds and dotted lines to the CHARMM model. 54

Figure 3.6 Excess coordination numbers (N_{ij}) as a function of composition for DMA solutions. 56

Figure 3.7 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for DMA solutions. The crosses were obtained directly from the KB integrals. The dashed lines were obtained by fitting the simulated excess volumes to Equation 3.2..... 57

Figure 3.8 Excess coordination numbers (N_{ij}) as a function of composition for ACT solutions. The black, red and green lines correspond to the experimental data at 298K for N_{22} , N_{11} , and N_{21} , respectively. The crosses, diamonds and circles corresponding to the respective data from the KBFF model obtained at 313K. 58

Figure 3.9 Potential of mean force profiles obtained from the center of mass rdfs in NMA solutions (top). Probability distribution for the angle (θ) between the N-H and C=O dipoles of hydrogen bonded ($r_{HO} < 0.275$ nm) NMA molecules (bottom). 62

Figure 4.1 Center of mass based rdfs for $x_2 = 0.1$ solutions of NMA (2) in water (1). As the rdf for NMA to NMA, g_{22} , increased, the rdf for NMA to water, g_{21} , decreased. Notice that g_{22} provided by GROMOS45a3 didn't approach unity within the studied distance, indicating high self aggregation of solutes. 76

Figure 4.2 Excess coordination numbers (N_{ij}) as a function of composition for NMA (2) in water (1) solutions. The GROMOS values were not displayed for clarity: N_{22} 's at $x_2=0.1$ and 0.2 were 52 ± 5 and 47 ± 2 , respectively. N_{21} 's at $x_2=0.1$ and 0.2 were -209 ± 18 and 189 ± 9 , respectively. And N_{11} 's at $x_2=0.1$, 0.2 and 0.5 were 92 ± 8 , 188 ± 8 , and 193 ± 24 respectively. 78

Figure 4.3 The snapshot of the simulated aqueous NMA solutions at $x_2=0.1$ with KBFF (left) and GROMOS45a3 (right). As the large positive N_{22} indicated in Figure 4.2, high degree of aggregation is observed in the system produced by GROMOS45a3..... 79

Figure 4.4 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for NMA(2) and water(1) solutions. Underestimation of density of pure water may be the cause of underestimation of density as well as the partial molar volume of water in the whole composition of mixture. 80

Figure 4.5 Enthalpy of mixing (H_m^E) for NMA solutions. The experimental data is from 308K and the simulations were performed at 313K. A large and positive enthalpy of mixing indicates high self aggregation..... 81

Figure 4.6 Center of mass based rdfs for $m_2=1.5\text{mol}/\text{kg}$ solutions of glycine (2) in water (1). The rdfs for water to water, g_{11} , are similar in all force field. The deviation between the force fields is larger in g_{22} 82

Figure 4.7 Excess coordination numbers (N_{ij}) as a function of composition for glycine (2) in water (1) solutions. The N_{ij} 's for OPLS are not displayed for clarity: N_{22} , N_{21} , and N_{11} were 14 ± 3 , -34 ± 6 , and 1 ± 1 at $m_2 = 1.5 \text{ mol}/\text{kg}$, 25 ± 2 , -62 ± 5 , and 7 ± 1 at $m_2 = 2.95 \text{ mol}/\text{kg}$, respectively. 83

Figure 4.8 The snapshots of the simulated glycine/water systems at $m_2=2.95 \text{ mol}/\text{kg}$ with KBFF(left) and OPLS(right). Overestimation of self aggregation is observed in the system with OPLS..... 84

Figure 4.9 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for glycine (2) and water (1) solutions. The density was slightly overestimated by CHARMM, and was slightly underestimated by OPLS and GROMOS45a3. KBFF demonstrated the best agreement with the experimental data in partial molar volumes of glycine and water. 85

Figure 6.1 Experimental data at 298K and 1atm for the molal activity coefficients (γ_2), solution density (ρ in g/cm^3) and the resulting preferential interactions (P_{22}^∞ in cm^3/mol) for a series

of Gly_n peptides as a function of peptide molality (m_2). The symbols indicate the raw experimental data while the solid lines are the corresponding fits after using Equation 6.4 and the parameters presented in Table 6.1. Gly (X), Gly₂ (O) and Gly₃ (*). All PIs are positive indicating a tendency for self association which decreases as the solute concentration increases. 118

List of Tables

Table 2.1 Simulated properties of 8M urea in water. Data taken from reference 15.....	30
Table 3.1 Bonded parameters for KBFF. Potential functions are: angles, $V_{\theta}=1/2 k_{\theta}(\theta-\theta_0)^2$; dihedrals, $V_{\varphi}=k_{\varphi}[1+\cos(n\varphi-\delta)]$; and impropers, $V_{\omega}=1/2 k_{\omega}(\omega-\omega_0)^2$. Force constants are in kJ/mol/rad, angles in degrees, and distances in nm.	44
Table 3.2 Nonbonded parameters for NMA, ACT, DMA, and water. LJ combination rules: $\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}}$ and $\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$	46
Table 3.3 Summary of the MD simulations of amide solutions. Symbols are: N_i , number of i molecules; m_2 , amide molality; V , average simulation volume; C , molarity; E_{pot} , average potential energy per molecule, and T_{sim} , total simulation time. All simulations were performed at $T=313K$ and $P=1atm$	47
Table 3.4 Properties of the pure liquids. Experimental data: density from refs 26-28; diffusion constants from refs 57,58; dielectric constants from refs 59,60; predicted compressibilities from refs 60,61; and thermal expansion coefficient from ref 60. Intramolecular contributions to E_{pot} were 0 and -107.9 kJ/mol for the KBFF and CHARMM models, respectively. Average molecular dipole moments were 4.6, 5.5, and 4.7 D for NMA, DMA, and ACT, respectively.	59
Table 3.5 Properties of NMA and water mixtures ($x_2=0.5$). Experimental data were taken from refs 25,26,47, and 62.....	60
Table 3.6 Atom based first shell coordination numbers.	61
Table 3.7 Center of mass based first shell coordination numbers. id; infinite dilution (one solute molecule).....	70
Table 3.8 Fitting constants for Equation 3.2.....	71

Table 4.1 Summary of the MD Simulations of NMA and Glycine Solutions.....	75
Table 6.1 Fitting parameters for Equation 6.4. Errors in the final PIs are estimated as <10%. Several parameters adopted values close to zero during the fitting procedure. In this case, the fit was repeated with these values set at zero. All the experimental peptide data refer to racemic mixture taken from references ¹⁵⁻¹⁷	117
Table 6.2 Decomposition of the molecule PIs into group contributions. Solute abbreviations are given in Table 6.1. Group abbreviations are peptide (P), charge (Q) and methyl (M). The QQ group interaction includes all combinations between two zwitterions (i.e. two +/-, one +/, one -/- interaction). The PQ interaction is a combination of the P+ and P- interactions (similarly for MQ).....	121

Acknowledgements

I would like to thank my research advisor Dr. Paul E. Smith for his guidance, instruction, and support throughout my graduate study. His enthusiasm for chemistry will guide me in my career as a good model for a scientist. I can't forget what he said when I just got back from lunch one day: "I couldn't help it." He assigned a complicated matrix of Kirkwood-Buff theory to me just before the lunch time, but he was so eager to solve it that he couldn't wait for lunch time without solving it by himself!

To members of Dr. Smith's group, past and present, who I have spent most of time with in the lab. I have had the pleasure of working with them. I appreciate that the chemistry is not the only thing I've learned in the lab during those years. Special thanks to Veronica Pierce, who has been such a good friend, both inside and outside of the lab.

I appreciate the collaboration with Dr. Stefan Bossmann and Dr. Viktor Chikan for the MspA project. And I owe thanks to Dr. Christer Aakerøy for giving the enlarged crystal coordinates of 4-pyridinecarboxamide.

Thanks to all of my Ph.D committee members Dr. Stefan Bossmann, Dr. Christer Aakerøy, Dr. Om Prakash, and Dr. David A Gustafson for their valuable time and input.

I am grateful to the faculty members, staffs, and fellow students in the chemistry department for their help and camaraderie during my study in Kansas State University.

Most of all, I am grateful for my family for all their love, support and guidance not just over the past five and-a-half years but throughout my life. They have been, and will be, there for me always. Although they are in the other side of the earth with 15 hours apart, it is them that have kept me going through all the difficulties and not being lost on the way. Thank you!

Dedication

To my mother

CHAPTER 1 - Introduction

1.1 General Introduction

Most proteins form organized three-dimensional structures under physiological conditions which are essential to their function. Both intra- and intermolecular interactions hold proteins in the particular conformation known as the native or folded structure. Basically, the structure depends on the sequence of the protein itself, but can also be affected by the environmental factors surrounding it. Various factors have been known to be involved with such an effect on intra- and intermolecular interactions; including changes in temperature, pH, and cosolvent *in vitro*, and mutations due to genetic defects and aging *in vivo*. It has been known that when the balance of those interactions is disturbed, proteins can become partially folded, misfolded, or denatured. Consequently, they may not act as they are supposed to, but lose their biological functions and/or aggregate with each other in many cases. For example, aggregation of proteins is related to aging and severe aging-related diseases such as Alzheimer's, Huntington's, Parkinson's, Amyotrophic lateral sclerosis (ALS), and prion diseases¹⁻³. A better understanding concerning these interactions between proteins will hopefully help us treat and even prevent those diseases. In addition, specific insight to the interaction between a protein of interest and its ligand can assist in the design of better ligands with specific desired properties.

A variety of techniques have been developed to probe the interactions which perturb structures and have revealed interesting properties of the interactions. One of the early approaches was to add some additional molecules to a system containing biomolecules. It has been observed that some of them can affect the properties of biomolecules in one way or the other. They are usually referred to as cosolvents since water is the primary solvent in the

biological system. Cosolvents affect the degree of aggregation and equilibrium between the folded state and the unfolded state of a protein by interfering with intramolecular interactions within a protein itself, as well as intermolecular interactions between the protein and water. The ones that disturb the native state of a protein to give non native structures are referred to as protein denaturants, or chaotropes, whereas others that stabilize the native structure of a protein are known as protein stabilizers, or kosmotropes. For example, urea, guanidinium chloride, and lithium perchlorate are chaotropic agents. Most organic solvents are also denaturing. Among the common stabilizers are porcine gelatin, recombinant human gelatins (rhGs), sucrose, lysine, arginine, Tween20, polyethylene glycol (PEG), and propylene glycol (PG)⁴⁻⁶.

The application of external mechanical forces represents another approach to study the interactions in biological systems. To investigate the binding properties of biomolecules, as well as their response to external mechanical manipulations, many experimental techniques have been developed. These include biomembrane force probe, atomic force microscopy (AFM), optical tweezers, and surface force apparatus experiments⁷⁻¹⁰. Despite of all these efforts, our understanding of these interactions in most biological systems is still not sufficient to quantitatively describe them due to their complexity and the technical limitations of current methods in space and time scale. Hopefully, computer simulations can provide us with valuable insights concerning these effects at the atomic level.

1.2 Molecular Simulation

It has been more than a century since pioneers, especially Ludwig Boltzmann (1844-1906) and Josiah Willard Gibbs (1839-1903), made a founding contribution to the concepts of statistical mechanics, which represents the link between the microscopic molecular details and

interactions of constituent particles, and the macroscopic and bulk properties observable in the laboratory¹¹. However, it was not until several decades ago that accessible numerical computers with reasonable cost became available for chemists and physicists to apply Boltzmann's principle to even the simplest systems using a number of limitations and assumptions^{12,13}. Along with the rapid advance in computer technology in the last few decades, many other limitations for computer simulations have been overcome. Computer simulation can provide details of many phenomena of interest at the atomic level, but there are still limitations in time scale and system size due to high computational cost for biological systems. Simplifying approximations, such as classical mechanics to describe molecular interactions via empirical potentials, are introduced in order to study most systems of interest. Although some approximations are still required, these approaches have provided numerous valuable insights into a variety of systems of biological interest.

Computer simulations yield exact results for problems in statistical mechanics, within the given assumptions and conditions, and therefore have been used as evaluation tools for theories and models for many interesting systems developed during the history of statistical mechanics. Computer simulation provides a direct route between models and theoretical predictions, and also between models and experimental results. Using computer simulations of the microscopic details of a system, the corresponding macroscopic properties of the system can be calculated and be compared with those of experiments. Computer simulations have also proven to be useful when studying a system under the extreme conditions of temperature and pressure. For example, it is difficult to perform experiment on a high-temperature plasma, a shock wave, a nuclear reactor or a planetary core, while it is feasible to simulate those systems in computer¹².

In molecular simulations, a variety of computational methods are used to relate the microscopic properties of individual molecules and their intermolecular and interatomic interactions with macroscopic bulk properties of the system. Monte-Carlo (MC) simulations¹⁴ and molecular dynamics (MD)^{15,16} simulations are the two major approaches. In both types of molecular simulations, a system should sample a sufficient number of microscopic configurations compatible with the interactions in the system, in addition to any thermodynamic constraints on the system, such as temperature, pressure and density. In the Monte-Carlo approach, 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. The acceptance of a new configuration is dependent on the Boltzmann distribution¹¹. In the molecular dynamics method, the net force on a molecule arising from all the other molecules in the system is evaluated for the initial arrangement, and each molecule in the box is moved using Newton's laws for a short time interval. These steps are repeated to provide with the dynamic properties of the system¹³. Since dynamical properties are also often of interest, molecular dynamics simulations have been used to investigate the biological systems.

1.3 Force Fields for the Simulation of Biological Systems

A force field refers to the set of equations, or potential functions, and their parameters used to describe the potential energy and its derivatives, i.e. the forces acting in a system of particles. A simple general form of a force field consists of covalent and noncovalent contributions to describe the total energy as given by:

$$E_{\text{total}} = E_{\text{bonded}} + E_{\text{nonbonded}} \quad (1.1)$$

with

$$E_{\text{bonded}} = E_{\text{bond}} + E_{\text{angle}} + E_{\text{dihedral}} \quad (1.2)$$

$$E_{\text{nonbonded}} = E_{\text{vanderWaals}} + E_{\text{electrostatic}} \quad (1.3)$$

In a molecular dynamics simulation each particle in the system behaves as described by the force field used in the simulation. Therefore, it is critical that all the parameters for the atomic properties and their interactions in a system are defined properly in order to obtain a correct description of the system. The force field used and sampling achieved are the crucial factors determining the quality of a molecular dynamics simulation.

The potential functions and parameters are typically derived from experiments and quantum calculations. That is why they are called ‘empirical’. Bonded parameters are usually optimized from experimental data such as gas-phase geometries and vibrational spectra, and torsional energy surfaces supplemented with ab initio results. For the optimization of non-bonded parameters, various sources of data can be used, including molecular volumes, experimental heats and free energies of vaporization, compressibility, solvation, density, and dipole moments. In particular, partial charges on the atoms of a molecule have been determined by ab initio calculations of gas phase complexes with a single water molecule in most of the existing force fields developed for biomolecules. Alternatively, gas phase quantum calculations followed by a scaling process are used in order to determine the partial charges on atoms suitable for liquids and solutions¹⁷. This necessarily involves some approximations. When the results of a simulation are interpreted, the approximations and limited conditions for the considered experiments used for their parameterization have to be taken into account. Hence, significant endeavors have been made to refine existing force fields, as well as to create new ones, during the last half of a century with two major concerns: accuracy and efficiency.

One way to improve force fields is to make them more accurate. Although, most of the potential functions and parameters have been well established, there is still room for improvement. In particular, the electrostatic interactions deserve special attention since electrostatic parameters, such as the partial charges on atoms used in most force fields, are parameterized using gas phase quantum mechanical calculations. Hence, polarization was not rigorously considered. To obtain improved accuracy with regard to electrostatic interactions, polarizable force fields have been developed by many research groups, including McKerell, van Gunsteren, Brooks, Borodin, and others¹⁸⁻³³. These typically yield better results, but the simulations are much slower than implicit effective charge approaches due to the iterative nature of the calculations.

Another way to improve force fields is to increase the efficiency of simulations so that larger and longer simulation can be performed with reasonable computational cost or simulation time. High computational cost of computer simulations still limits the time scale and the system size. In biological systems, the main molecules are often proteins, and usually they have big molecular weights. In addition, they are surrounded by water and other species like Na^+ and Cl^- under physiological conditions. Biological systems have a large number of molecules which slow down the simulations. Hence, the time scale of some biological events of interest are beyond that of typical simulations and the goal remains to simulate larger systems for longer times to access interesting features with reasonable computational cost. A wide variety of approaches have been made to improve the efficiency of simulations. For example, implicit solvent and coarse grain force fields have been developed^{25,34-41}. They introduce simplified models to be simulated, which means that some detailed information of a system may be traded

for the improved efficiency. With regard to computational techniques, parallelization of molecular dynamics simulations and applications also yields high efficiency⁴².

Although force fields for biological system have been well established, some indications of problems have occurred. For example, lower solvation than expected from experiment has been observed in simulations with existing force fields⁴³⁻⁴⁶. As a consequence, simulations may produce potentials which stabilize the native or folded state too strongly, leading to improper equilibrium between the native and unfolded states. Also it has been noticed that it is difficult to quantify or rank the binding potentials in ligand-protein docking studies with current force fields⁴⁷, even though the structures of complexes can be relatively well reproduced.

Smith and coworkers have noticed that these errors may originate from the lack of ability of the common force fields to maintain the delicate balance between solute-solute interactions and solute-solvent (solvation) interactions^{46,48-54}. Solution mixtures are of our main interest, especially in biological systems. But in the development of typical force field approaches the properties of pure compounds have been used. And then it is assumed that they would remain the same in solution mixtures. Recently, Kirkwood-Buff (KB) theory has been applied to quantify solute-solute and solute-solvent interactions in solution mixtures in the entire range of composition. KB theory is a useful theoretical tool to evaluate the ability of a force field to represent the correct distribution of molecules in solution.

1.4 Kirkwood-Buff Theory

In 1951 Kirkwood and Buff published an important paper concerned with a general statistical mechanical theory of solutions⁵⁵. Kirkwood-Buff theory relates thermodynamic properties of a solution mixture to the molecular distribution functions. The expressions provided

by KB theory are totally general and valid for any kind of particle in the entire range of compositions (Figure 1.1). The derivation doesn't require any assumption of additivity of the total potential energy, which makes it more universally valid than other theories⁵⁶. Even though it was a milestone as a powerful tool in the solution theory, it was not until Ben-Naim developed the useful procedure of inversion that this exact theory of solutions was applied⁵⁷. He provided inversion methods to abstract information on the affinity between a pair of species in the solution mixture. Since then, many chemists and physicist, including Smith, Marcus, Ruckenstein, Shimizu, Hall, Zielkiewicz, Lepori, and many others, have followed his lead to develop KB theory and apply it to the study of various solution mixtures^{44-46,48-51,53,58-99}.

The distribution of components of a system can be expressed in terms of a set of distribution functions known as radial distribution functions. A radial distribution function (rdf), $g(r)$, provides the probability of finding an atom at a distance r from a central atom, relative to the probability at the same distance expected for a completely random distribution at the same density ρ . A radial distribution function $g(r)$ can be calculated by integration of the configurational distribution function over the position of atoms, and then normalizing it^{12,100}. In a system consisting of N particles the radial distribution function for molecules 1 and 2 can be expressed as¹¹:

$$g(r_{12}) = \frac{\iint \dots \int e^{-\beta V_N} dr_3 dr_4 \dots dr_N}{N^2 \iint \dots \int e^{-\beta V_N} dr_1 dr_2 \dots dr_N} \quad (1.4)$$

where $\beta=1/kT$, and V_N is the N-particle potential energy. This is a simple expression of the Boltzmann distribution for the relative locations of two molecules in the system. Figure 1.2 shows a typical radial distribution function. It displays a series of solvation shells in the vicinity of a given molecule or atom. As the distance r goes larger, the distribution of components of the

solution mixture becomes more similar to a random distribution. The radial distribution function provides insight into the liquid structure, and integrals over $g(r)$'s are also useful to express thermodynamic properties of solution mixture.

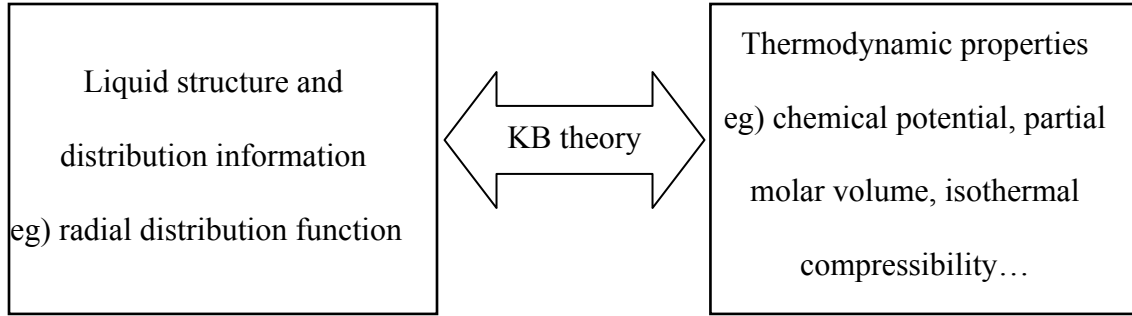


Figure 1.1 The role of Kirkwood-Buff theory

In the KB theory of solutions, thermodynamic properties of a solution mixture can be derived from radial distribution functions, and vice versa. Figure 1.1 illustrates well the role of KB theory as a bridge between these two. KB integrals, G_{ij} , are defined as integrals over radial distribution functions between species i and j ⁵¹:

$$G_{ij} = 4\pi \int_0^{\infty} [g_{ij}^{\mu VT}(r) - 1] r^2 dr \quad (1.5)$$

where $g_{ij}^{\mu VT}(r)$ is a radial distribution functions (rdf) in the μVT ensemble, and r is the corresponding center-of-mass to center-of-mass distance. An excess coordination numbers can be defined as $N_{ij} = \rho_j G_{ij}$, where $\rho_j = N_j/V$ the number density (molar concentration) of species j . A positive N_{ij} indicates an excess of species j in the vicinity of species i over random distribution, while a negative N_{ij} means depletion of species j surrounding species i . In other words, positive N_{ij} can be interpreted as attractive intermolecular interactions between species i and j , and

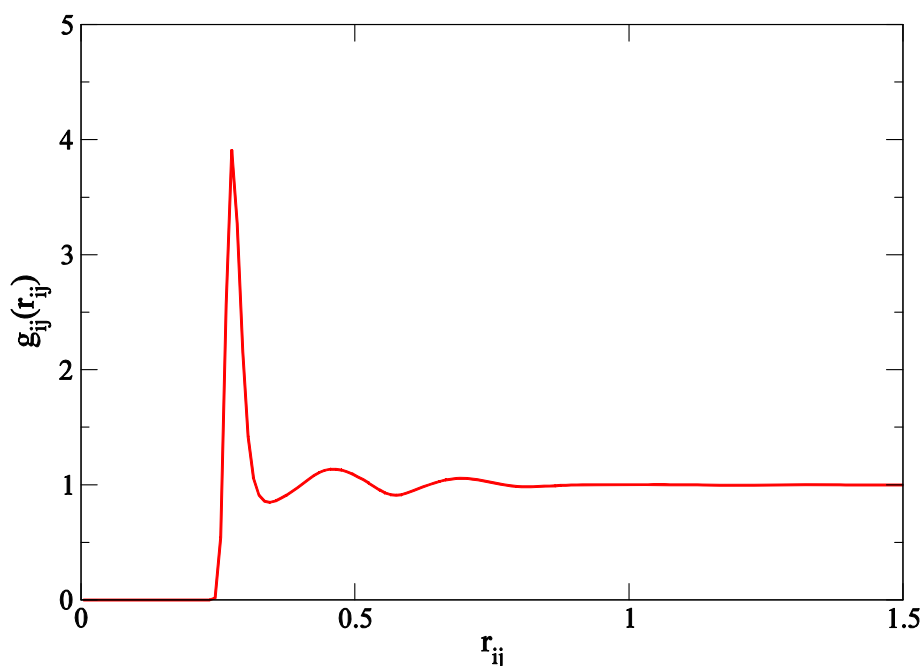


Figure 1.2 Radial distribution function (rdf). The rdf displays the local solution structures, including solvation shells. As the distance between species i and j , r_{ij} , gets larger, the rdf goes to unity, meaning the distribution becomes similar to the bulk distribution.

negative N_{ij} is related to repulsive interactions. Generic examples of KB integrals G_{ij} and excess coordination numbers N_{ij} are illustrated in Figure 1.3 and 1.4.

KB integrals can be determined either by experimental or simulated data. For a binary solution mixtures with water (1) and cosolvent (2) at constant pressure (p) and temperature (T), the chemical potentials (μ_i), partial molar volumes (\bar{V}_i), and isothermal compressibilities (κ_T) can be obtained experimentally. Then the experimental data can be used to determine the three component dependent KB integrals (see chapters 3, 4, and 5 for the detailed equations)¹⁰¹. In a system of a biomolecule (2) and cosolvent (3) with primary solvent of water (1), the preferential

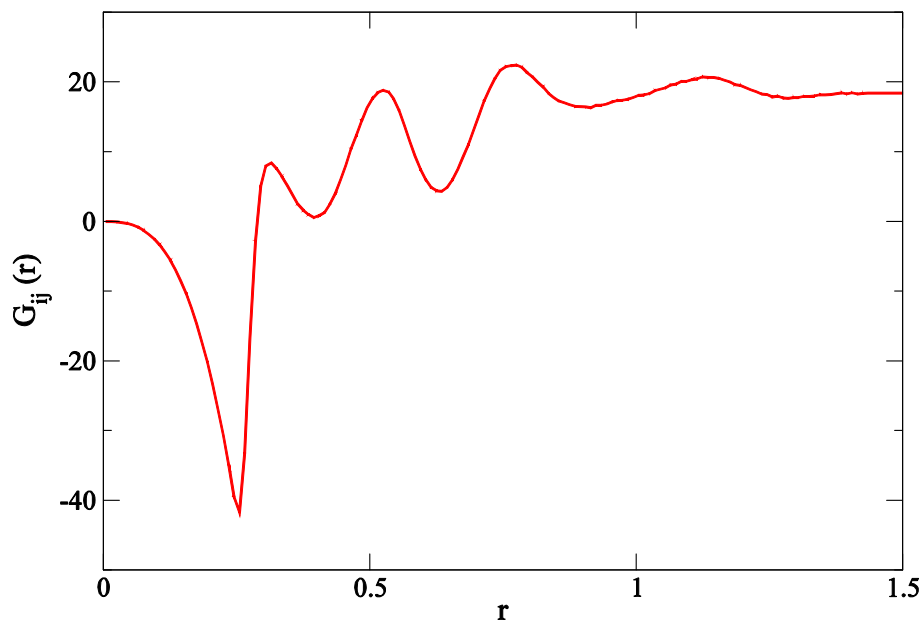


Figure 1.3 An example of KB integral G_{ij} as a function of integration distance r (nm) between species i and j . This KB integral corresponds to the rdf displayed in Figure 1.2.

binding parameters can be obtained from equilibrium dialysis experiments and also expressed using KB integrals⁴⁹.

KB theory has been applied to a number of biomolecular systems, as well as a variety of cosolvent systems. For instance, Rosgen and coworkers have applied KB theory to determine molecular crowding effects on macromolecules and small molecules¹⁰² and to understand structural thermodynamics of protein preferential solvation¹⁰³. Matubayasi and coworkers have used KB theory to analyze the free energy of molecular binding into lipid membranes¹⁰⁴ and to characterize the preferential interactions in bovine serum albumin in the presence of a wide range of salts⁸¹. Shulgin and coworkers have applied KB theory as a theoretical tool to analyze cosolvent contribution to the osmotic secondary virial coefficient in the ternary mixtures containing protein, water and cosolvent¹⁰⁵. Hirata et al. have calculated changes in the partial

molar volumes associated with coil-to-helix transition of peptides using KB theory^{62,106}. Lenhoff and coworkers have applied KB theory to interpret experimental data to understand the concentration dependence of the partial specific volumes of proteins in aqueous solution⁷².

1.5 Kirkwood-Buff Derived Force Field

As simulation data can be evaluated by comparing KB integrals, as well as thermodynamic properties obtained from them, to those from the experimental data, the quality of a force field used in the simulation can be determined. Furthermore, it has been observed that

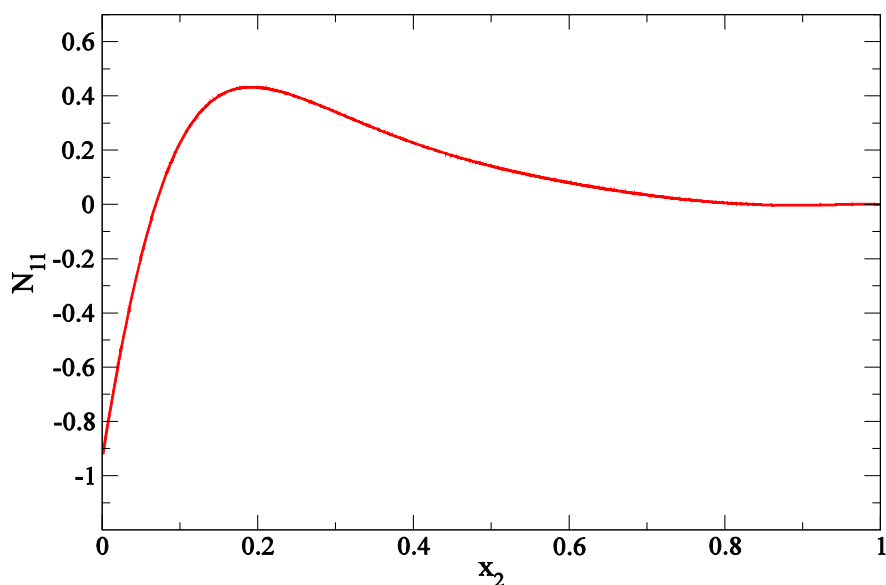


Figure 1.4 An example of excess coordination number N_{II} for the entire range of the composition. The sign and of N_{ij} indicates the feature of intermolecular interactions between species i and j : positive N_{ij} indicates attractive interactions between i and j and negative N_{ij} represent repulsive interactions. The graph displays the N_{II} , and therefore the self interactions between species I 's change as the composition of solution mixture changes.

the KB integrals are far more sensitive to the parameters of a force field than other experimental data. Above all, they are novel indicators for the affinity that quantify interactions between a pair of components in the system, which makes it possible to determine how accurately a particular force field can represent the distribution of molecules in solution mixture. Hence, KB integrals can play a critical role in the parameterization of a new force field. Early tests using the common force fields were bad at reproducing KB integrals, suggesting that an improved force field is needed. By using KB integrals, a force field can be developed to maintain the delicate balance between solute-solute interactions and solute-solvent interactions (solvation).

KB integrals are very sensitive to a force field, in particular to charge distributions. Therefore, in the development of a new force field based on KB theory, Smith and coworkers have focused on charge distributions for atoms. In contrast, bonded parameters are well known from the experimental data. Among the non-bonded parameters, the van der Waals interactions are also well established. But charge distributions on atoms in solution mixtures still need to be improved. As mentioned in section 1.3, in typical force field approaches, the partial charges on atoms of a molecule are determined using ab initio calculations of a gas-phase, complex with water or followed by scaling. In the KB approach, the charges on the atoms are adjusted to best reproduce the density and KB integrals for solution mixtures at several different compositions.

A Kirkwood-Buff derived force field (KBFF) is still a non-polarizable force field, but with the aid of sensitive KB integrals, the most effective charges on atoms can be found. It has been shown to perform better than common non-polarizable force fields with the same computational cost^{44,46,96-98}. With regard to the accuracy, it has been demonstrated that simulation results with KBFF are comparable to those obtained with newly developed

polarizable force field of van Gunsteren and coworkers in selected system including aqueous N-Methylacetamide (NMA) and methanol-water mixture¹⁰⁷.

1.6 Summary

Computer simulations have been useful tools in studying biological systems by providing details at the atomic level. Kirkwood-Buff theory can be used to interpret experimental and computational data and to provide a bridge between them. Here, we use KB theory and computer simulations for a variety of applications.

In chapter 2, the preferential interaction parameters in biological systems are determined by Kirkwood-Buff theory and computer simulation. A system of Lysozyme in 8M urea solution is chosen for study since extensive experimental data are available for this system. The simulated data using two different force fields for urea are compared with experimental data. Comparison between two different urea force fields is performed, which reveals how a new and improved force field can help understanding preferential interactions in biological systems.

KB theory is then used for the parameterization of a new KBFF force field. Chapter 3 involves the detailed development of a force field for amides based on Kirkwood-Buff theory. NMA is chosen as a model for a peptide bond which is essential in biological systems. Also, comparisons with the existing force fields for peptides and proteins are performed to demonstrate the improvement made by our new force field in chapter 4.

KB theory provides accurate and general relations for solution mixtures. But the specific terms to express the relationships between the thermodynamic properties and KB integrals get more complicated as the number of components in a system increases. It is not easy to derive simple and straight-forward relations beyond ternary systems since more components are

involved in matrix operations. Explicit relationships for KB integrals developed for 4 and higher components solution mixtures are discussed in chapter 5.

In addition, KB theory has been applied to develop a preferential interaction model for solution mixtures. In chapter 6, a pairwise preferential interaction model (PPIM), characterized by KB integrals, is developed to quantify and characterize the interactions between functional groups observed in peptides.

1.7 References

1. Truant, R.; Atwal, R. S.; Desmond, C.; Munsie, L.; Tran, T. FEBS Journal 2008, 275, 4252-4262.
2. Batchelor, J. D.; Olteanu, A.; Tripathy, A.; Pielak, G. J. Journal of the American Chemical Society 2004, 126, 1958-1961.
3. Creighton, T. E. Proteins, Structures and Molecular Properties, 2nd ed.; W. H. Freeman and Company: New York, 1996.
4. Derrick, T.; Grillo, A. O.; Vitharana, S. N.; Jones, L.; Rexroad, J.; Shah, A.; Perkins, M.; Spitznagel, T. M.; Middaugh, C. R. Journal of Pharmaceutical Sciences 2007, 96, 761-776.
5. Thyagarajapuram, N.; Olsen, D.; Middaugh, C. R. Journal of Pharmaceutical Sciences 2007, 96, 3304-3315.
6. Arakawa, T.; Dix, D. B.; Chang, B. S. Yakugaku Zasshi 2003, 123, 957-961.
7. Svoboda, K.; Block, S. M. Annual Review of Biophysics and Biomolecular Structure 1994, 23, 247-285.
8. Binnig, G.; Quate, C. F.; Gerber, C. Physical Review Letters 1986, 56, 930-933.
9. Evans, E.; Ritchie, K.; Merkel, R. Biophysical Journal 1995, 68, 2580-2587.
10. Israelachvili, J. N. Intermolecular and surface Forces, Academic Press: London, 1992.
11. Atkins, P. W.; DePaula, J. Physical Chemistry, 7th ed. ed.; W.H. Freeman and Company: New York, 2003.
12. Allen M.P.; Tildesley, D. J. Computer Simulation of Liquids, Oxford University Press: Oxford, 1989.

13. Bopp, P. A.; Buhn, J. B.; Hampe, M. J. *Ullmann's Modeling and Simulation* 2007, 323-340.
14. Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *Journal of Chemical Physics* 1953, 21, 1087-1092.
15. Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* 1957, 27, 1208-1209.
16. Alder, B. J.; Wainwright, T. E. *Journal of Chemical Physics* 1959, 31, 459-466.
17. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* 1998, 102, 3586-3616.
18. Anisimov, V. M.; Vorobyov, I. V.; Lamoureux, G.; Noskov, S.; Roux, B.; MacKerell, A. D. *Biophysical Journal* 2004, 86, 415A.
19. Anisimov, V. M.; Roux, B.; MacKerell, A. D. *Biophysical Journal* 2007, 152A.
20. Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. *Journal of Chemical Theory and Computation* 2005, 1, 153-168.
21. Borodin, O.; Smith, G. D.; Sewell, T. D.; Bedrov, D. *Journal of Physical Chemistry B* 2008, 112, 734-742.
22. Lopes, P. E. M.; Roux, B.; MacKerell, A. D. *Biophysical Journal* 2007, 642A-643A.
23. Vorobyov, I.; Anisimov, V. M.; Greene, S.; Venable, R. M.; Moser, A.; Pastor, R. W.; MacKerell, A. D. *Journal of Chemical Theory and Computation* 2007, 3, 1120-1133.

24. Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. Abstracts of Papers of the American Chemical Society 2005, 230, U1352-U1353.
25. Masella, M.; Borgis, D.; Cuniasse, P. Journal of Computational Chemistry 2008, 29, 1707-1724.
26. Patel, S.; MacKerell, A. D.; Brooks, C. L. Journal of Computational Chemistry 2004, 25, 1504-1514.
27. Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. Journal of Physical Chemistry B 2005, 109, 18988-18999.
28. Xie, W. S.; Pu, J. Z.; MacKerell, A. D.; Gao, J. L. Journal of Chemical Theory and Computation 2007, 3, 1878-1889.
29. Geerke, D. P.; Van Gunsteren, W. F. Molecular Physics 2007, 105, 1861-1881.
30. Geerke, D. P.; Thiel, S.; Thiel, W.; Van Gunsteren, W. F. Journal of Chemical Theory and Computation 2007, 3, 1499-1509.
31. Yu, H. B.; Geerke, D. P.; Liu, H. Y.; van Gunsteren, W. E. Journal of Computational Chemistry 2006, 27, 1494-1504.
32. Yu, H. B.; Van Gunsteren, W. F. Journal of Chemical Physics 2004, 121, 9549-9564.
33. Yu, H. B.; Hansson, T.; Van Gunsteren, W. F. Journal of Chemical Physics 2003, 118, 221-234.
34. Allen, E. C.; Rutledge, G. C. Journal of Chemical Physics 2008, 128, 154115/1-12
35. Anand, P.; Nandel, F. S.; Hansmann, U. H. E. Journal of Chemical Physics 2008, 128, 165102/1-5.
36. Basdevant, N.; Borgis, D.; Ha-Duong, T. Journal of Physical Chemistry B 2007, 111, 9390-9399.

37. Brancato, G.; Rega, N.; Barone, V. *Journal of Chemical Physics* 2008, 128, 144501/1-10.
38. Chen, J.; Brooks, C. L. *Physical Chemistry Chemical Physics* 2008, 10, 471-481.
39. Green, D. F. *Journal of Physical Chemistry B* 2008, 112, 5238-5249.
40. Revalee, J. D.; Laradji, M.; Kumar, P. B. S. *Journal of Chemical Physics* 2008, 128, 035102/1-9.
41. Xie, D.; Zhou, S. Z. *BIT Numerical Mathematics* 2007, 47, 853-871.
42. Chaplot, S. L. *Computational Materials Science* 2006, 37, 146-151.
43. Perera, A.; Sokolic, F. *Journal of Chemical Physics* 2004, 121, 11272-11282.
44. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2003, 107, 3891-3898.
45. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 5901-5910.
46. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 119, 11342-11349.
47. Mattila, K.; Haltia, T. *Proteins-Structure Function and Bioinformatics* 2005, 59, 708-722.
48. Kang, M.; Smith, P. E. *Journal of Computational Chemistry* 2006, 27, 1477-1485.
49. Smith, P. E. *Journal of Physical Chemistry B* 2006, 110, 2862-2868.
50. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2002, 106, 1491-1500.
51. Kang, M.; Smith, P. E. *Fluid Phase Equilibria* 2007, 256, 14-19.
52. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochemistry and Biophysics* 2008, 50, 1-22.
53. Smith, P. E. *Biophysical Journal* 2006, 91, 849-856.
54. Smith, P. E.; Marlow, G. E.; Pettitt, B. M. *Journal of the American Chemical Society* 1993, 115, 7493-7498.
55. Kirkwood, J. G.; Buff, F. P. *Journal of Chemical Physics* 1951, 19, 774-777.
56. Ben Naim, A. *Molecular theory of solutions*, Oxford University Press: New York, 2006.

57. Bennaïm, A. *Journal of Chemical Physics* 1977, 67, 4884-4890.
58. Guha, A.; Mukherjee, D. *Journal of the Indian Chemical Society* 1997, 74, 195-198.
59. Guha, A.; Ghosh, N. K. *Indian Journal of Chemistry Section A-Inorganic Bio-Inorganic Physical Theoretical & Analytical Chemistry* 1998, 37, 97-101.
60. Hall, D. G. *Journal of the Chemical Society-Faraday Transactions* 1991, 87, 3523-3528.
61. Imai, T.; Kinoshita, M.; Hirata, F. *Journal of Chemical Physics* 2000, 112, 9469-9478.
62. Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. *Biopolymers* 2001, 59, 512-519.
63. Imai, T.; Takahiro, T.; Kovalenko, A.; Hirata, F.; Kato, M.; Taniguchi, Y. *Biopolymers* 2005, 79, 97-105.
64. Kang, M.; Smith, P. E. *Journal of Chemical Physics* 2008, 128, 244511/1-8.
65. Lynch, G. C.; Perkyms, J. S.; Pettitt, B. M. *Journal of Computational Physics* 1999, 151, 135-145.
66. Marcus, Y. *Monatshefte fur Chemie* 2001, 132, 1387-1411.
67. Matteoli, E.; Mansoori, G. A. *Journal of Chemical Physics* 1995, 103, 4672-4677.
68. Matteoli, E. *Journal of Molecular Liquids* 1999, 79, 101-121.
69. Nain, A. K. *Journal of Solution Chemistry* 2008, 37, 1541-1559.
70. Pandey, J. D.; Verma, R. *Chemical Physics* 2001, 270, 429-438.
71. Patil, K. J.; Mehta, G. R.; Dhondge, S. S. *Indian Journal of Chemistry Section A-Inorganic Bio-Inorganic Physical Theoretical & Analytical Chemistry* 1994, 33, 1069-1074.
72. Pjura, P. E.; Paulaitis, M. E.; Lenhoff, A. M. *AIChE Journal* 1995, 41, 1005-1009.
73. Rosgen, J.; Pettitt, B. M.; Bolen, D. W. *Protein Science* 2007, 16, 733-743.
74. Rosgen, J. *Osmosensing and Osmosignaling* 2007, 428, 459-486.

75. Ruckenstein, E.; Shulgin, I. *Journal of Physical Chemistry B* 1999, 103, 10266-10271.
76. Ruckenstein, E.; Shulgin, I. *Fluid Phase Equilibria* 2001, 180, 281-297.
77. Schellman, J. A. *Quarterly Reviews of Biophysics* 2005, 38, 351-361.
78. Shimizu, S. *Proceedings of the National Academy of Sciences of the United States of America* 2004, 101, 1195-1199.
79. Shimizu, S.; Smith, D. J. *Journal of Chemical Physics* 2004, 121, 1148-1154.
80. Shimizu, S.; Matubayasi, N. *Chemical Physics Letters* 2006, 420, 518-522.
81. Shimizu, S.; McLaren, W. M.; Matubayasi, N. *Journal of Chemical Physics* 2006, 124, 234905/1-4.
82. Shimizu, S.; Boon, C. L. *Journal of Chemical Physics* 2004, 121, 9147-9155.
83. Shulgin, I.; Ruckenstein, E. *Journal of Physical Chemistry B* 1999, 103, 872-877.
84. Shulgin, I.; Ruckenstein, E. *Polymer* 2003, 44, 901-907.
85. Shulgin, I.; Ruckenstein, E. *Industrial & Engineering Chemistry Research* 2002, 41, 6279-6283.
86. Shulgin, I. L.; Ruckenstein, E. *Journal of Chemical Physics* 2005, 123, 054909.
87. Shulgin, I. L.; Ruckenstein, E. *Journal of Physical Chemistry B* 2007, 111, 3990-3998.
88. Shulgin, I. L.; Ruckenstein, E. *Journal of Physical Chemistry B* 2008, 112, 3005-3012.
89. Shulgin, I. L.; Ruckenstein, E. *Fluid Phase Equilibria* 2007, 260, 126-134.
90. Shulgin, I. L.; Ruckenstein, E. *Journal of Physical Chemistry B* 2006, 110, 12707-12713.
91. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 18716-18724.
92. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 16271-16278.
93. Smith, P. E. *Journal of Chemical Physics* 2008, 129, 124509/1-5.
94. Smith, P. E.; Mazo, R. A. *Journal of Physical Chemistry B* 2008, 112, 7875-7884.

95. Warshavsky, V. B.; Song, X. Y. *Physical Review e* 2008, 77, 051106.
96. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2004, 121, 2180-2186.
97. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2005, 109, 15080-15086.
98. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 10663-10670.
99. Zielenkiewicz, W.; Kulikov, O. V.; Krestov, G. A. *Bulletin of the Polish Academy of Sciences-Chemistry* 1992, 40, 293-305.
100. McDonald, I. R.; Singer, K. *Molecular Physics* 1972, 23, 29-40.
101. Matteoli, E.; Lepori, L. *Journal of Chemical Physics* 1984, 80, 2856-2863.
102. Rosgen J. *Protein Structure, Stability, and Interactions*; Humana Press Inc.: Totowa, NJ, 2009; pp. 195-225.
103. Auton, M.; Bolen, D. W.; Rosgen, J. *Proteins* 2008, 73, 802-813.
104. Matubayasi, N.; Shinoda, W.; Nakahara, M. *Journal of Chemical Physics* 2008, 128, 195107/1-13.
105. Shulgin, I. L.; Ruckenstein, E. *Journal of Physical Chemistry B* 2008, 112, 14665-14671.
106. Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *Journal of Chemical Physics* 2001, 114, 9506-9511.
107. Unpublished data from van Gunsteren, W.F. and Geerke, D.P. 2009.

CHAPTER 2 - Preferential Interaction Parameters in Biological Systems by Kirkwood–Buff Theory and Computer Simulation*

Abstract

Recent results concerning the formulation and evaluation of preferential interactions in biological systems in terms of Kirkwood-Buff (KB) integrals are presented. In particular, experimental and simulated preferential interactions of a cosolvent with a biomolecule in the presence of water are described. It is argued that the preferential interaction parameter defined in a system open to both cosolvent and solvent corresponds to the situation most relevant to the analysis of computer simulation results of cosolvent interactions with proteins and small peptides. Hence, KB theory provides a path from quantities determined from simulation data to the corresponding thermodynamic data.

*Reprinted with permission from "Preferential interaction parameters in biological systems by Kirkwood–Buff theory and computer simulation" by Myungshim Kang and Paul E. Smith, 2007. *Fluid Phase Equilibria*, 256, 14-19. Copyright 2007 by [Elsevier B.V.](#)

2.1 Introduction

Protein denaturation is an important process which remains poorly understood at the atomic level. In principle, computer simulations provide the atomic level detail required for an improved description of cosolvent interactions with proteins. However, the majority of computer simulations of cosolvent effects on peptides and proteins have been rather qualitative in nature¹⁻⁵. In particular, a direct connection between the simulations and experimental thermodynamic data has been noticeably absent. This is primarily a result of the weak binding of many cosolvents to proteins⁶. This presents a conflict between the traditional binding site models used to interpret the experimental data^{7,8}, and the inability to locate binding sites and assign binding constants from the simulation data.

More recently, it has become possible to study cosolvent effects in a quantitative manner by the use of Kirkwood-Buff (KB) theory⁹⁻¹³. The use of KB theory is particularly well suited for the analysis of experimental data as it involves no approximations, and for the analysis of simulation data as it only requires the determination of radial distribution functions (rdfs), or coordination numbers, which are easily obtained from simulations. Our previous studies have involved using KB theory to improve the force fields required for computer simulation¹⁴⁻¹⁹, relating simulation data on cosolvent effects to experimental thermodynamic data^{10,11,20,21}, and for the interpretation of thermodynamic data on cosolvent effects on biomolecules²¹⁻²⁴. Here, we present our latest efforts to use KB theory for the analysis of experimental and computer simulation data relating to the interaction of cosolvents with proteins. The system chosen for study is Lysozyme in urea solutions as extensive experimental data exists for this system. We show how this can be used to provide data which is also available from simulation. A

comparison of the simulated and experimental data is then performed for the above system using two different force fields for urea.

2.2 Background and Theory

The notation used here follows the usual definitions for biological systems where the subscripts 1, 2, and 3 refer to the primary solvent (usually water), the biomolecule, and cosolvent, respectively. All equations refer to the limit of an infinitely dilute biomolecule. The same formulation can be applied to systems with finite protein concentrations, but is significantly more complicated. The basic approach is to use KB theory to interpret experimental data from equilibrium dialysis and cosolvent denaturation experiments. The exact details have been outlined elsewhere^{21,23}. Kirkwood-Buff theory provides relationships between particle number fluctuations and derivatives of the chemical potentials in the grand canonical (μ VT) ensemble where the volume (V), temperature (T), and chemical potential (μ) of all species are constant. The primary result used here is that^{25,26},

$$\frac{RT}{V} \left(\frac{\partial N_i}{\partial \mu_j} \right)_{T,V,\mu_{k \neq j}} = \left(\frac{\partial \rho_i}{\partial \beta \mu_j} \right)_{T,\mu_{k \neq j}} = \rho_i \rho_j G_{ij} + \rho_i \delta_{ij} \quad (2.1)$$

where G_{ij} is the Kirkwood-Buff integral between species i and j , $\rho_i = N_i/V$ is the number density (molar concentration) of species i , R is the Gas Constant, $\beta = 1/RT$, and δ_{ij} is the Kroenecker delta function. The KB integrals are defined in terms of the corresponding rdfs (g_{ij}) such that^{21,26},

$$G_{ij} = G_{ji} = 4\pi \int_0^\infty [g_{ij}^{\mu VT}(r) - 1] r^2 dr \approx 4\pi \int_0^{R_c} [g_{ij}^{NPT}(r) - 1] r^2 dr \quad (2.2)$$

An excess coordination number can be defined ($N_{ij} = \rho_j G_{ij} \neq N_{ji}$) which characterizes the excess number of j molecules around an i molecule in the open system above that observed within an

equivalent volume of the bulk reference solution. KB theory can then be used to provide expressions for thermodynamic properties in other ensembles by using suitable thermodynamic transformations²⁶.

The approximation in the above equation is required for evaluating KB integrals in closed systems. Here, a correlation region exists, defined by a distance R_c , within which the local cosolvent and solvent density around the species of interest differs from the bulk density. Beyond the correlation region all $g_{ij}(r) \approx 1$. The correlation region can extend over many molecular solvation shells and therefore provides a potentially different representation of the cosolvent effect from that assumed in the common binding models, where binding is usually limited to the protein surface.

Equilibrium dialysis experiments provide data on the thermodynamic binding of cosolvents to a biomolecule. This data is usually expressed in terms of the preferential interaction (PI) of the cosolvent with the protein (denoted by Γ_{23}), which measures the change in cosolvent molality (m_3) on changing the biomolecule molality (m_2) in a system open to the cosolvent and water, but not the biomolecule. This is also often referred to as the preferential binding parameter. In the infinitely dilute biomolecule limit an exact expression for Γ_{23} in terms of KB integrals can be obtained²³,

$$\Gamma_{23} = \left(\frac{\partial m_3}{\partial m_2} \right)_{T, \mu_1, \mu_3} = \rho_3 (G_{23} - G_{21}) = N_{23} - \frac{\rho_3}{\rho_1} N_{21} \quad (2.3)$$

The value of Γ_{23} is dependent on the cosolvent concentration. If the biomolecule exists as a mixture of different major forms (native and denatured for example), the dialysis experiment provides an average preferential interaction such that,

$$\Gamma_{23} = f_D \Gamma_{D3} + f_N \Gamma_{N3} \quad (2.4)$$

where f_i is the fraction of state i . Hence, the total preferential interaction is simply the sum of the individual preferential interactions. The effects of cosolvents on biomolecules can also be quantified by cosolvent denaturation studies in closed systems. For a biomolecular equilibrium ($K=f_D/f_N$) affected by a cosolvent one can show that²¹,

$$\left(\frac{\partial \ln K}{\partial \ln a_3}\right)_{T,P} = \Delta\Gamma_{23} = \Gamma_{D3} - \Gamma_{N3} = \Delta N_{23} - \frac{\rho_3}{\rho_1} \Delta N_{21} \quad (2.5)$$

where a_3 is the cosolvent activity (on any scale) and $\Delta N_{2i} = N_{Di} - N_{Ni}$. Hence, the denaturation process is driven by the difference in preferential interaction of the cosolvent with the two different states of the protein. When $\Delta\Gamma_{23}$ is positive, negative, or zero, the cosolvent can be classified as a denaturant, an osmolyte, or thermodynamically inert, respectively. Most denaturation studies use cosolvent concentration and not activity. As the biomolecule is infinitely dilute, the transformation involves a property of the cosolvent and water solution only and can also be expressed in terms of KB integrals according to²⁶,

$$\left(\frac{\partial \ln a_3}{\partial \ln \rho_3}\right)_{T,P} = \frac{1}{1 + \rho_3(G_{33} - G_{13})} = a_{33} \quad (2.6)$$

For our purposes we will also assume that the cosolvent denaturation thermodynamics fits the empirical m -value model such that the change in standard free energy of unfolding is given by,

$$\Delta\Delta G = \Delta G^O(\rho_3) - \Delta G^O(0) = -m\rho_3 \quad (2.7)$$

where m is a constant. The approximation is usually good for urea denaturation. Hence, combining Equations 2.5-2.7 provides,

$$\frac{m\rho_3}{RTa_{33}} = \Delta\Gamma_{23} = \Delta N_{23} - \frac{\rho_3}{\rho_1} \Delta N_{21} \quad (2.8)$$

where all the data on the far lhs can be obtained experimentally. A combination of Equations 2.4 and 2.5 (or 2.8) can therefore be used to separate the preferential interaction into components involving the native and denatured states as a function of cosolvent concentration. Xie and Timasheff have also described this procedure²⁷. The results for Lysozyme in urea solutions at pH 7 (where the protein remains folded) and pH 2 (where the protein unfolds with a transition midpoint of 3.7M urea) have been determined previously^{21,27}. The advantage of this approach is that information on cosolvent binding to the native state is available at high urea concentrations where the protein may actually be predominantly unfolded. This is important as the use of computer simulations to study the denatured state is complicated by our limited understanding of the unfolded state of proteins, and our inability to simulate for times long enough to observe unfolding. In contrast, simulations of the native state in high urea concentrations can be performed relatively easily and provide good statistics on the required preferential interactions. This is the approach taken here. An alternative approach is to study cosolvent binding at low denaturant concentrations so as to avoid populating the denatured state²⁸.

In addition, the individual values of N_{21} and N_{23} can be extracted from the above expressions using the KB results for the partial molar volume (\bar{V}) of the solute at infinite dilution in terms of properties of the reference solution,

$$\bar{V}_2^\infty = RT\kappa_T - \rho_1\bar{V}_1G_{21} - \rho_3\bar{V}_3G_{23} = RT\kappa_T - N_{21}\bar{V}_1 - N_{23}\bar{V}_3 \quad (2.9)$$

where κ_T is the isothermal compressibility of the solution. However, we will not pursue this decomposition here.

2.3 Methods

The details of the simulations will be published elsewhere. Briefly, Hen Egg White Lysozyme was simulated by classical dynamics at 300 K and 1 atm using the Gromacs program and the GROMOS 43a1 force field²⁹⁻³¹. The system included Lysozyme, 4096 urea molecules, and 18,112 water molecules in a cubic box of length 9.5 nm. The water model was SPC/E and two different urea force fields (KBFF and OPLS) were investigated^{15,32,33}. Positional constraints were applied to the C^α atoms of the protein to prevent partial unfolding. The Figure 2.1 displays the simulated structure of Lysozyme in 8M urea at pH 7 using KBFF. The simulation corresponding to pH 7 was performed with all residues in their usual protonation state, while the simulation at pH 2 was performed with all carboxylate groups protonated. The total simulation time was 6 ns with the final 5 ns being used for averaging purposes. The properties of the urea and water mixtures have been studied previously and some of the results are displayed in Table 2.1¹⁵. The OPLS and KBFF models display significantly different urea aggregation behavior, with a high degree of self aggregation (large positive G_{ii} values) observed for the OPLS model. We note that the experimental data on Lysozyme denaturation was obtained at 293 K. This is slightly different to the present simulation temperature. We chose 300 K as the simulated properties of urea and water mixtures are known at this temperature. It is expected that the small temperature difference will have little effect on the results considering the errors inherent in both the experimental ($\pm 1-3$) and simulated ($\pm 5-10$) PI data.

The preferential interaction of urea with the protein was determined from the simulations by calculating the number of urea (n_{23}) and water molecules (n_{21}) within a distance R from any atom of the protein. This provides a distance dependent PI according to,

$$\Gamma_{23}(R) = n_{23}(R) - \frac{n_3}{n_1} n_{21}(R) \quad (2.10)$$

The ratio of $n_3/n_1 = \rho_3/\rho_1$ is the bulk cosolvent to solvent ratio. In cases where the n_{23} and n_{21} are large, the bulk cosolvent to solvent ratio should be corrected to account for the cosolvent and water molecules which have moved from the bulk reference solution (beyond R_c) to the local solution environment around the protein. Hence, a more correct expression is,

$$\Gamma_{23}(R) = n_{23}(R) - \frac{n_3 - n_{23}(R)}{n_1 - n_{21}(R)} n_{21}(R) \quad (2.11)$$

The correction involves only a small change in the corresponding bulk concentration ratio, but this can have a significant effect on the calculated PI as n_{21} can be large. The above analysis is conceptually equivalent to assuming a virtual dialysis membrane located at a distance R_c from the protein surface. The local volume then represents an open system in contact with a closed NPT particle bath located beyond the correlation distance. The approximation should be reasonable for systems where the bath volume is far larger than the correlation volume. It cannot be used as R approaches $L/2$, where L is the simulation box length.

2.4 Results

Table 2.1 Simulated properties of 8M urea in water. Data taken from reference 15.

	ρ (g/cm ³)	G_{33} (cm ³ /mol)	G_{13} (cm ³ /mol)	G_{11} (cm ³ /mol)	a_{33}
KBFF	1.121	-39	-48	-1	0.93
OPLS	1.130	391	-282	128	0.16
Experimental	1.119	-56	-39	-5	1.16

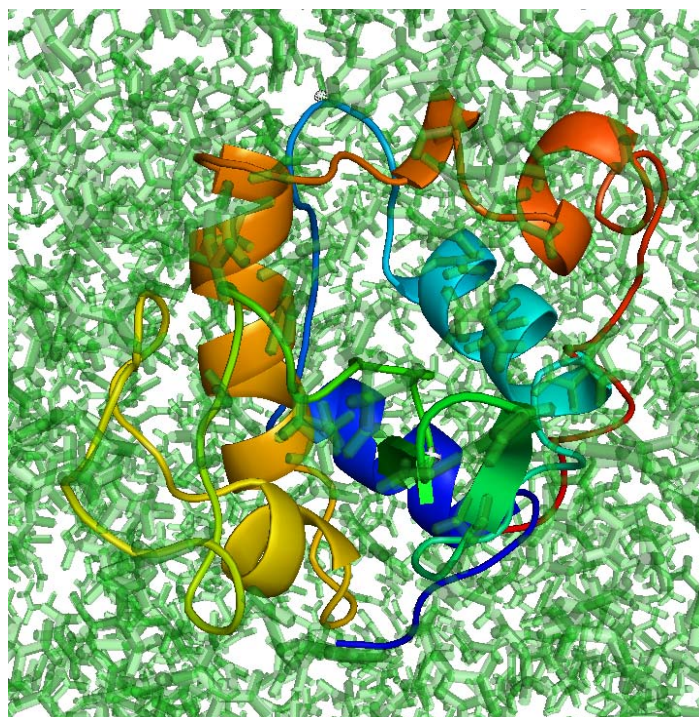


Figure 2.1 The structure of native Lysozyme in 8M urea at pH 7. The cartoon model displays Lysozyme colored by chain, and the green stick model indicates urea. The water molecules are not displayed for clarity.

The preferential interaction (PI) of urea with Lysozyme is displayed in Figure 2.2 for the KBFF and OPLS urea force fields as a function of distance away from the protein surface. The PI is negative at small distances due to the excluded volume effect and the fact that urea is larger than water. The PI then increases sharply due to the presence of an increased number of urea molecules in the first solvation shell, and a corresponding decreased number of water molecules, over that expected from the bulk solution ratio. The OPLS urea model displayed a large preferential interaction of urea with Lysozyme which was several times that of the KBFF urea model, and had not reached the expected plateau value. The correlation volume as defined by R_c was also larger (> 1.5 nm) for the OPLS model than for the KBFF model (1.0 nm). The

difference between using Equations 2.10 and 2.11 was also more significant for the OPLS model as expected based on their respective PI values.

The experimental value of Γ_{23} for this system is determined to be 16 at pH 7 and -10 at pH 2. The results for the KBFF and OPLS models were 50 and > 200 , respectively. Clearly, both urea models display a significantly larger PI than experiment. In our opinion, this is almost certainly due to inaccuracies in the protein force field as our previous studies have shown that common solute force fields used to construct protein force fields do not typically reproduce the experimental KB integrals for solution mixtures^{15,16,19,34}. The KBFF model of urea and water does reproduce the experimental KB integrals (see Table 2.1) and probably explains the improvement of the KBFF urea model over the OPLS model. However, without a reparametrized protein force field it appears that quantitative agreement with experiment will be difficult.

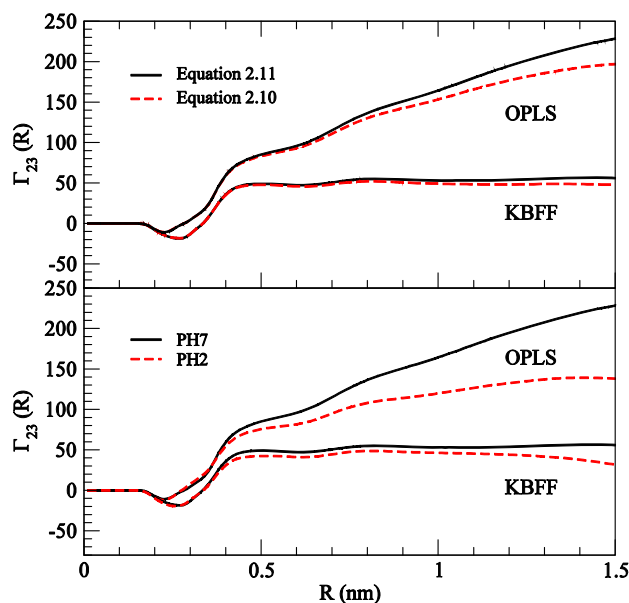


Figure 2.2 The simulated distance dependent PI of 8M urea with native Lysozyme. As a function of urea force field at pH 7 according to Equations 2.10 and 2.11 (top). As a function of urea force field and pH using Equation 2.11 (bottom). The experimentally observed PIs are 16 at pH 7 and -10 at pH 2.

The effect of pH on the simulated PI of urea with Lysozyme is also displayed in Figure 2.2. The simulated effect was small in comparison with the absolute values of Γ_{23} , and the fluctuations in the instantaneous PI observed during the simulations. As observed in the pH 7 simulations, the simulated PI values were larger for the OPLS model and both models produced values which did not agree with experiment. The decrease in PI on lowering the pH was reproduced by both models although the PI values remained positive whereas the experimental value is actually negative. However, the small difference observed for the KBFF model was within the statistical errors associated with the simulated PI values. The rdfs corresponding to the urea and water distribution from the protein surface are displayed in Figure 2.3. The rdfs indicated a significant interaction of urea with the surface groups on Lysozyme (between 0.3 and

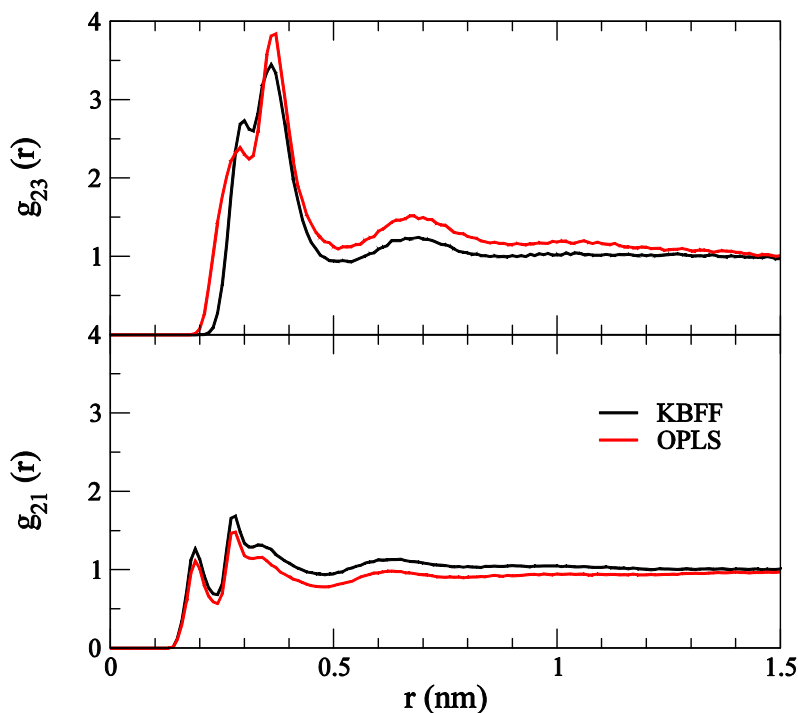


Figure 2.3 The urea (top) and water (bottom) rdfs as a function of the closest distance (r) to any protein atom for the two different urea force fields. The simulated data correspond to 8M urea and a protein at pH 7.

0.5 nm) with a smaller second urea shell at larger distances (between 0.6 and 0.8 nm). As expected, the water molecules penetrated closer to the protein surface. The differences between the two urea models included an enhanced water interaction over all distances for the KBFF model. A corresponding increase in the urea distribution over a larger distance range was observed for the OPLS model. The large differences in the PI values for both urea models were due to the relatively small differences in the rdfs beyond the initial protein contact distance (0.5 nm), which were magnified upon integration.

Figure 2.4 displays the PI of urea with Lysozyme as a function of simulation time. The PI increased from around zero (corresponding to a random initial arrangement of molecules) to fluctuate around a larger positive value of the PI. The time history suggests that between 1-2 ns

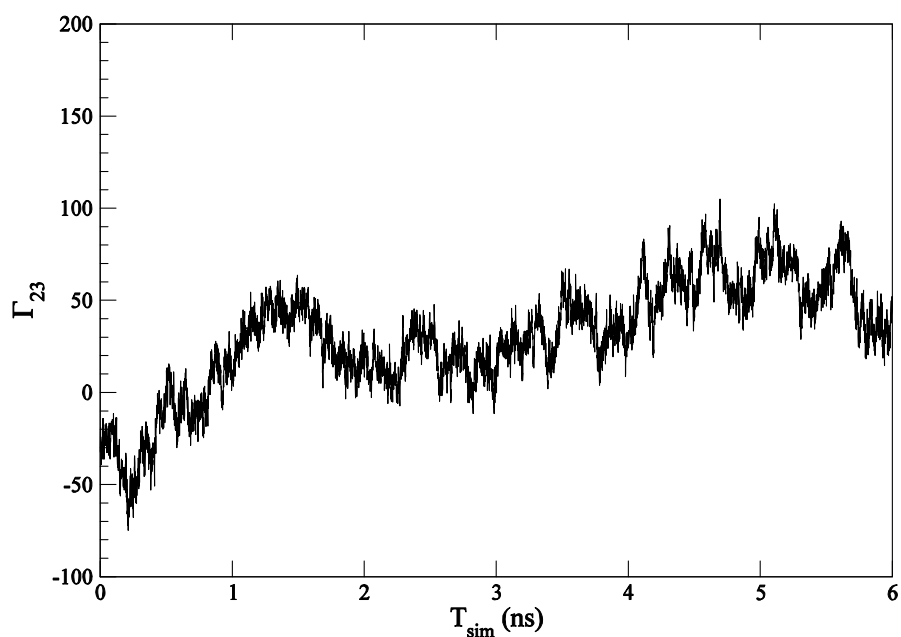


Figure 2.4 The time history of the 8M urea PI corresponding to the KBFF model of urea and a protein at pH 7. The value of Γ_{23} was determined for $R = R_c = 1.0$ nm using Equation 2.11.

of simulation were required for the urea distribution to equilibrate and provide a reasonable estimate of the PI. This is in agreement with our earlier simulation studies on simple solution mixtures³⁵. Long time fluctuations were apparent in the value of the PI which requires averaging over several ns of simulation time to determine PI values with reasonable precision. Slightly longer equilibration times were required for the OPLS urea model.

Most cosolvent molecules are significantly larger than water molecules and cannot therefore approach as close to the protein surface as the water molecules. Consequently, it is well known that this gives rise to an excluded volume effect which tends to stabilize folded proteins. Osmolytes typically enhance the excluded volume effect, whereas denaturants have to overcome the excluded volume effect by binding to the protein in order to promote denaturation. Recently, Schellman has estimated the excluded volume contribution to Γ_{23} for several cosolvents and proteins using some simple approximations involving protein solvent accessible surface areas³⁶. Estimates for the excluded volume contribution to Γ_{23} can also be obtained from the current simulations. From the data provided by Schellman for urea and Lysozyme, the excluded volume (referred to as the gap volume by Schellman) for the native protein is 8000 cm³/mol, if a protein volume of 10100 cm³/mol is assumed³⁶. This leads to an excluded volume contribution to the PI of -64 in 8M urea. The excluded volume contribution can be obtained from the simulations as the value of the first minimum in the distance dependent PI (see Figure 2.2). This provides contributions of -18 for the KBFF and -11 for the OPLS force fields, corresponding to an excluded volume of 2250 and 1375 cm³/mol, respectively. Clearly, the simulated excluded volume contributions are smaller than the estimates provided by Schellman. This is probably due to the fact that the simulations include protein flexibility and the non spherical nature of urea into the calculation. The degree of excluded volume was also dependent on the urea model and the

subsequent interaction with the protein. A larger excluded volume effect was observed with the KBFF model and was consistent with the smaller PI observed for this model.

2.5 Conclusions

It has been shown that calculations of the preferential interaction of urea with native Lysozyme under denaturing conditions provide a convenient way to study cosolvent effects on proteins, especially for comparison of simulated and experimental data. The simulated PI values require several ns of simulation time to equilibrate, and display large fluctuations on the ns timescale. A large positive PI was observed with two different urea models, both of which indicated significantly more urea affinity for the protein than suggested by experiment. In our opinion, the PI represents a model sensitive property of solution mixtures and requires accurate force fields if one requires quantitative agreement with experiment.

The difference between the results obtained for KBFF and OPLS urea models was significant. This is not due to the different water model used here (GROMOS and OPLS were developed to be used with the SPC and TIP3P water models), as we have demonstrated that the KB integrals are relatively insensitive to the water model¹⁴⁻¹⁶. The increased degree of aggregation observed for the OPLS urea model, over the KBFF model, affects the corresponding preferential interactions observed in the presence of an infinitely dilute solute. In an earlier study of cavity formation in urea solutions it was observed that the degree of urea exclusion from the cavity was directly related to the degree of urea aggregation observed in solution; a larger degree of urea aggregation producing a correspondingly larger degree of urea exclusion²⁰. Here, the OPLS model describes a more favorable preferential interaction of urea with Lysozyme than the KBFF model. Therefore, it appears that if the urea and water force field is not correctly balanced,

leading to incorrect descriptions of self aggregation, this will result in an excessive negative preferential interaction if the solute–cosolvent interaction is unfavorable, and an excessive positive preferential interaction if the solute–cosolvent interaction is favorable. Finally, it should be noted that the general appearance of the urea and water rdfs around the protein are in qualitative agreement for both urea models, even though the thermodynamics are not.

In our opinion, the only way one can be confident in the agreement between experimental and simulated data on preferential interactions is to use both a cosolvent and protein force field which have been shown to accurately reproduce the KB integrals observed for solution mixtures.

2.6 References

1. Smith, P. E.; Pettitt, B. M. *Journal of the American Chemical Society* 1991, 113, 6029-6037.
2. Smith, P. E.; Marlow, G. E.; Pettitt, B. M. *Journal of the American Chemical Society* 1993, 115, 7493-7498.
3. TiradoRives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* 1997, 36, 7313-7329.
4. Caflisch, A.; Karplus, M. *Structure with Folding & Design* 1999, 7, 477-488.
5. Zou, Q.; Bennion, B. J.; Daggett, V.; Murphy, K. P. *Journal of the American Chemical Society* 2002, 124, 1192-1202.
6. Timasheff, S. N. *Advances in Protein Chemistry* 1998, 51, 355-432.
7. Tanford, C. *Advances in Protein Chemistry* 1970, 24, 1-95.
8. Schellman, J. A. *Biophysical Chemistry* 1990, 37, 121-140.
9. Smith, P. E. *Journal of Physical Chemistry B* 1999, 103, 525-534.
10. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2001, 105, 11513-11522.
11. Abui, M.; Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 7382-7388.
12. Shimizu, S.; Boon, C. L. *Journal of Chemical Physics*. 2004, 121, 9147-9155.
13. Shulgin, I. L.; Ruckenstein, E. *Journal of Chemical Physics*. 2005, 123, 054909.
14. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 10663-10670.
15. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2003, 107, 3891-3898.
16. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 119, 11342-11349.
17. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2004, 121, 2180-2186.
18. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2005, 109, 15080-15086.

19. Kang, M.; Smith, P. E. *Journal of Computational Chemistry* 2006, 27, 1477-1485.
20. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 5901-5910.
21. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 18716-18724.
22. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 16271-16278.
23. Smith, P. E. *Journal of Physical Chemistry B* 2006, 110, 2862-2868.
24. Smith, P. E. *Biophysical Journal* 2006, 91, 849-856.
25. Kirkwood, J. G.; Buff, F. P. *Journal of Chemical Physics* 1951, 19, 774-777.
26. A. Ben-Naim *Statistical Thermodynamics for Chemists and Biochemists*, Plenum Press: 1992.
27. Timasheff, S. N.; Xie, G. F. *Biophysical Chemistry* 2003, 105, 421-448.
28. Baynes, B. M.; Trout, B. L. *Journal of Physical Chemistry B* 2003, 107, 14058-14067.
29. Scott, W. R. P.; Hunenberger, P. H.; Tironi, I. G.; Mark, A. E.; Billeter, S. R.; Fennen, J.; Torda, A. E.; Huber, T.; Kruger, P.; van Gunsteren, W. F. *Journal of Physical Chemistry A* 1999, 103, 3596-3607.
30. Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. *Computer Physics Communications* 1995, 91, 43-56.
31. Lindahl, E.; Hess, B.; van der Spoel, D. *Journal of Molecular Modeling* 2001, 7, 306-317.
32. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* 1987, 91, 6269-6271.
33. Duffy, E. M.; Severance, D. L.; Jorgensen, W. L. *Israel Journal of Chemistry* 1993, 33, 323-330.
34. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 115, 5521-5530.
35. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 114, 426-435.

36. Schellman, J. A. *Biophysical Journal* 2003, 85, 108-125.

CHAPTER 3 - A Kirkwood-Buff Derived Force Field for Amides[†]

Abstract

A force field for the computer simulation of aqueous solutions of amides is presented. The force field is designed to reproduce the experimentally observed density and Kirkwood-Buff integrals for N-methylacetamide (NMA), allowing for an accurate description of the NMA activity. Other properties such as the translational diffusion constant and heat of mixing are also well reproduced. The force field is then extended to include N,N'-dimethylacetamide and acetamide with good success. Analysis of the simulations of low concentrations of NMA in water indicates a high degree of solvation with only 15% of the NMA molecules involved in solute-solute hydrogen bonding. There is only a weak angular dependence of the solute-solute hydrogen bonding interaction with a minimum at an angle of 65° for the N-H and C=O dipole vectors. The models presented here provide a basis for an accurate force field for peptides and proteins.

[†] Reprinted with permission from "A Kirkwood-Buff Derived Force Field for Amides" by Myungshim Kang and Paul E. Smith, 2006. *Journal of Computational Chemistry*, 27, 1477-1485. Copyright © 2006 Wiley Periodicals, Inc., A Wiley Company.

3.1 Introduction

As part of our continuing effort to develop accurate force fields for the simulation of solution mixtures and their application in biomolecular systems,¹⁻⁵ we present a force field for the simulation of amides. A realistic description of the interaction between amide groups and solvent molecules is fundamental to the accuracy of any peptide or protein force field. Small amides provide simple models for several functional groups observed in peptides and proteins. N-methylacetamide (NMA) represents a model for the peptide group, while N,N-dimethylacetamide (DMA) and acetamide (ACT) provide models for Pro and the Asn/Gln amide side chains, respectively. Hence, many simulation studies have investigated the properties of these molecules in solution.⁶⁻¹⁶

In developing a force field for peptides and proteins it is important to ensure one maintains a correct balance of hydrogen bonding between peptide groups, and the degree of solvation of the peptide groups. Too little solvation will tend to favor self aggregation of the peptide groups, whereas too much solvation will destabilize native state structures. Unfortunately, it has typically been difficult to determine such a delicate balance.¹⁴ Recently, however, the use of Kirkwood-Buff (KB) integrals to quantify solute-solute and solute-solvent interactions has provided a procedure to determine the ability of a particular force field to represent the correct distribution of molecules in solution.^{17,18} This forms the basis of the Kirkwood-Buff derived force field (KBFF) approach which we use here to reproduce the experimental data (KB integrals) corresponding to the intermolecular interactions observed in solution mixtures of NMA, DMA and ACT with water. To our knowledge, only one previous study has determined simulated KB integrals for NMA and water mixtures.¹² That study used the GROMOS force field and was limited to rather short simulation times of 200ps in systems

containing 729 molecules. Since then we have shown that that at least 1 ns of simulation time is usually required for systems of several thousand molecules to correctly determine the values of the KB integrals.^{2,19}

3.2 Methods

All experimental and simulation data refer to systems at 313 K and 1 atm unless stated otherwise. A Kirkwood-Buff analysis of the experimental data for the cosolvent (2) and water (1) mixtures was performed as outlined by Ben-Naim,²⁰ and in our own recent studies.⁵ The Kirkwood-Buff integrals (G_{ij}) are defined as,²¹

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

where g_{ij} is the corresponding radial distribution function (rdf), and the approximation is made for simulations performed in closed systems.²²

Experimental activities²³⁻²⁵ and densities²⁶⁻²⁸ were taken from the literature for NMA, DMA, and ACT. The compressibilities were assumed to follow the usual mixture rule based on volume fractions.²⁹ Excess quantities were fitted to the Redlich-Kister equation,³⁰

$$X_m^E = x_2 x_1 \sum_{i=0}^n a_i (x_1 - x_2)^i \quad (3.2)$$

where a_i are fitting constants, x_i are mole fractions, and X is either the volume V , Gibbs energy βG , or enthalpy βH , and $\beta=1/RT$. Partial molar quantities at any composition are then given by the standard relationship,

$$Y_i = X_m^E - x_j \left(\frac{\partial X_m^E}{\partial x_j} \right)_{p,T} \quad (3.3)$$

Table 3.1 Bonded parameters for KBFF. Potential functions are: angles, $V_\theta=1/2 k_\theta(\theta-\theta_0)^2$; dihedrals, $V_\varphi= k_\varphi[1+\cos(n\varphi-\delta)]$; and impropers, $V_\omega=1/2 k_\omega(\omega-\omega_0)^2$. Force constants are in kJ/mol/rad, angles in degrees, and distances in nm.

bonds	r_0		r_0
C-O	0.1224	C-N	0.1386
C-C	0.1520	N-H	0.1000
N-C(CH ₃)	0.1468		
Angles		k_θ	θ_0
NMA	O-C-C	502.1	124.1
	O-C-N	502.1	121.8
	C-C-N	502.1	114.1
	C-N-H	292.9	110.4
	C-N-C	502.1	119.6
	H-N-C(CH ₃)	376.6	130.0
ACT	O-C-C	502.1	120.9
	O-C-N	502.1	122.4
	C-C-N	502.1	116.7
	C-N-H(N)	292.9	117.0
	C-N-H(N)	292.9	121.5
	H-N-H	334.7	121.5
DMA	O-C-C	502.1	126.8
	O-C-N	502.1	124.5
	C1-C-N	502.1	113.9
	C-N-C(CH ₃)	502.1	117.7
	C(CH ₃)-N-C(CH ₃)	502.1	115.4
Dihedrals	k_φ	δ	n
C-C-N-C(CH ₃)	33.5	180.0	2
C-C-N-H(N)	33.5	180.0	2
Impropers	k_ω		ω_0
N-C-C(CH ₃)-H	167.4		0.0
C-C(CH ₃)-N-O	167.4		0.0
N-H(N)-H(N)-C	167.4		0.0
N-C-C(CH ₃)-C(CH ₃)	167.4		0.0

where X can be V , βG , or βH giving rise to the properties (Y) corresponding of the partial molar volume (\bar{V}), excess chemical potential ($\beta\mu^E = \ln f$), or partial molar enthalpy ($\beta\bar{h}$), respectively. The mole fraction activity derivative (f_{22}) can be expressed in terms of Equation 3.2 and the KB integrals via,²⁰

$$f_{22} = \left(\frac{\partial \ln f_2}{\partial \ln x_2} \right)_{p,T} = x_3 x_1 \beta \left(\frac{\partial^2 G_m^E}{\partial x_2^2} \right)_{p,T} = - \frac{x_2 \rho_1 \Delta G}{1 + x_2 \rho_1 \Delta G} \quad (3.4)$$

where $\Delta G = G_{22} + G_{11} - 2 G_{21}$. The experimental densities, compressibilities, and activity coefficients were then used to determine the corresponding KB integrals using the inversion procedure outlined by Ben-Naim.²⁰ The resulting KB integrals were in agreement with previously reported values,^{12,31} except for the values of G_{ij} observed at very low concentrations of either component, as expected due to the uncertainties in the experimental (and simulated) data in those regions.³²

All mixtures were simulated using classical molecular dynamics techniques using the SPC/E water model.³³ The simulations were performed in the isothermal isobaric (NpT) ensemble at 313 K and 1 atm. The weak coupling technique³⁴ was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively. All bonds were constrained using SHAKE³⁵ and a relative tolerance of 10^{-4} , allowing a 2 fs timestep for integration of the equations of motion. The particle mesh Ewald technique was used to evaluate the electrostatic interactions.³⁶ A real space convergence parameter of 3.5 nm^{-1} was used in combination with twin range cutoffs of 0.8 and 1.5 nm, and a nonbonded update frequency of 10 steps. The reciprocal space sum was evaluated on a 60^3 grid with $\leq 0.1 \text{ nm}$ resolution. Random initial configurations of the solute and water molecules in a cubic box of approximate length 5 nm were used. The steepest descent method was then used to perform 100 steps of minimization.

Table 3.2 Nonbonded parameters for NMA, ACT, DMA, and water. LJ combination rules:

$$\sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}} \text{ and } \varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}}$$

Model	Atom	ε kJ/mol	σ nm	q e
ALL	C	0.330	0.336	0.62
	O	0.560	0.310	-0.62
	CH ₃	0.867	0.375	0.00
	N	0.500	0.311	-0.70
NMA	CH ₃ (N)	0.867	0.375	0.34
	H	0.088	0.158	0.36
ACT	H	0.088	0.158	0.35
DMA	CH ₃ (N)	0.867	0.375	0.35
SPC/E	O	0.6506	0.3166	-0.8476
	H	0.0	0.0	0.4238

This was followed by extensive equilibration which was continued until all intermolecular potential energy contributions and rdfs displayed no drift with time (typically 1 ns). Configurations were saved every 0.1 ps for analysis.

Translational self diffusion constants (D_i) were determined using the mean square fluctuation approach,³⁷ relative permittivities (ε) from the dipole moment fluctuations,³⁸ finite difference compressibilities (κ_T) by performing additional simulations of 500 ps at 250 atm,³⁹ thermal expansion coefficients (α) from additional simulations of 500 ps simulations at 333 K,

and excess enthalpies of mixing (H_m^E) determined from the average potential energies (E_{pot}).³⁹ Errors ($\pm 1\sigma$) in the simulation data were estimated by using two or three block averages.

Table 3.3 Summary of the MD simulations of amide solutions. Symbols are: N_i , number of i molecules; m_2 , amide molality; V , average simulation volume; C , molarity; E_{pot} , average potential energy per molecule, and T_{sim} , total simulation time. All simulations were performed at $T=313K$ and $P=1atm$.

	N_2	N_1	x_2	m_2 mol/kg	V nm ³	C M	E_{pot} kJ/mol	T_{sim} ns
NMA	318	2860	0.1001	6.17	125.717	4.20	-46.58	6
	515	2059	0.2001	13.88	126.029	6.78	-47.35	6
	692	1307	0.3462	29.39	125.958	9.12	-48.16	6
	804	804	0.5000	55.51	125.806	10.61	-48.61	6
	875	471	0.6501	103.12	125.806	11.55	-48.54	6
	924	231	0.8000	222.04	125.889	12.19	-48.17	6
	970	0	1.0000		125.919	12.79	-47.38	6
DMA	300	2705	0.0998	6.16	125.761	3.96	-46.68	6
	699	699	0.5000	55.51	124.861	9.29	-47.57	6
	812	0	1.0000		123.856	10.88	-41.41	5
ACT	190	3613	0.0500	2.92	127.018	2.48	-37.75	6
	343	3083	0.1001	6.18	125.179	4.55	-46.44	6
	640	0	1.0000		55.593	19.11	-60.92	6
Water	0	2170	0.0000		65.734		-45.64	2

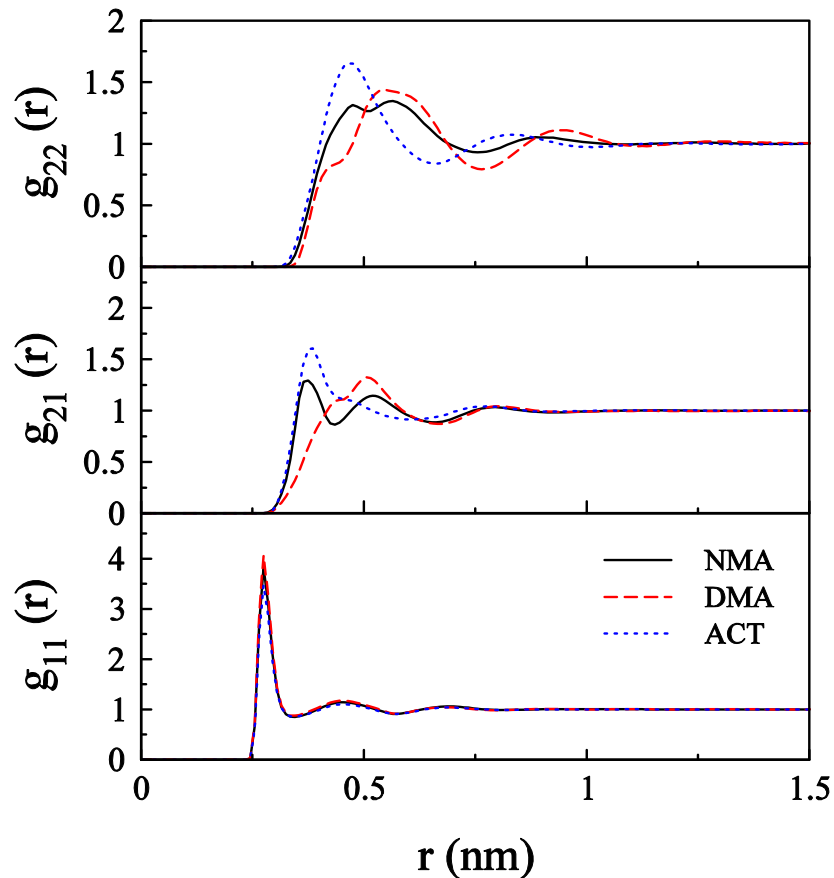


Figure 3.1 Center of mass based rdfs for $x_2 = 0.1$ solutions of NMA, DMA, and ACT

The KBFF force field used in this study involves a Lennard-Jones (LJ) 6-12 plus Coulomb potential, together with the SPC/E water model.¹ Our previous studies used a simple scheme to obtain polar atom parameters for the LJ term,³ and the same approach is used here. United atom methyl group parameters were taken from the literature.⁴⁰ As the parameters characterizing the sizes of the atoms have not been varied, it is possible that the resulting combination of atom sizes and charges is not unique.⁵ However, the current LJ parameters were taken from our previous acetone and urea studies, for which the solution densities were well reproduced even though the atomic charge distributions are quite different.^{2,3} Therefore, in our

opinion, the current set of LJ parameters, while not necessarily unique, should provide very reasonable estimates for the atomic sizes. The molecular geometries were taken from the available crystal structures and the MM4 force field,⁴¹⁻⁴³ with bonded parameters taken from the GROMOS96 force field.⁴⁴ All intramolecular interactions within the NMA, DMA and ACT molecules were removed. The charges on the atoms were then adjusted to best reproduce the density and KB integrals for solution mixtures with $x_2 = 0.20, 0.50,$ and 0.80 . The final parameters are presented in Tables 3.1 and 3.2. Comparison with the CHARMM22 all atom force field was also performed using the parameters presented for peptide group in the protein force field.⁴⁵

3.3 Results

Parametrization of the atomic charges for NMA, DMA and ACT using the solution mixtures displayed in Table 3.3 resulted in the nonbonded parameters presented in Table 3.2. Examples of the center of mass based rdfs obtained for mixtures with $x_2 = 0.1$ are displayed in Figure 3.1. The rdfs indicated that g_{11} was essentially identical for all three solutes while the degree of first shell solute-solute and solute-solvent interactions increased from DMA through NMA to ACT. This is to be expected considering the increase in the number of potential hydrogen bonding sites. DMA, the largest solute, displayed some weak structure in g_{33} between 1.0 and 1.5 nm. Hence, the KB integrals for all systems studied here were obtained from the simulated data by averaging the values obtained between 1.2 and 1.6 nm, a slightly larger distance than our previous studies.^{1,5} First shell center of mass based coordination numbers (n_{ij}) obtained from the simulations are presented in the supporting information. Some of the

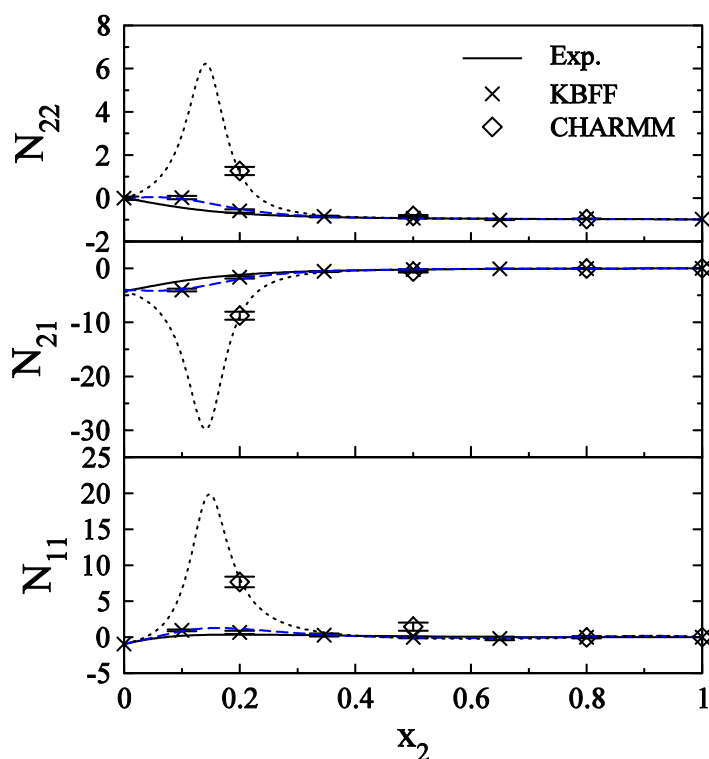


Figure 3.2 Excess coordination numbers (N_{ij}) as a function of composition for NMA solutions. The lines corresponding to the KBFF and CHARMM force fields were obtained after determining the simulated values of f_{22} and then obtaining the fitting constants for Equation 3.2 via a fit of the simulated data to Equation 3.4.

coordination numbers were large in magnitude due to the large integration distance to the first minimum. They display the expected trends with composition; increases for n_{22} and decreases for n_{21} and n_{11} .

The excess coordination numbers ($N_{ij} = \rho_j G_{ij}$) obtained from the simulations of NMA and water mixtures are compared to the experimental data in Figure 3.2. The experimental data was reproduced by the KBFF model for all but low ($x_2 = 0.1$) solute concentrations. In this region there was a small overestimation in the self interactions (N_{22} and N_{11}) suggesting a slight

underestimation in the solvation of NMA (N_{21}). Also included in Figure 3.2 are the corresponding results for the CHARMM22 all atom force field. Here a rather large degree of self aggregation of NMA and water molecules was evident at low compositions, as displayed in Figure 3.3. The presence of excessive self aggregation has also been observed for other force fields and other systems.^{1,3,18,46} From this perspective, it is clear that the KBFF model represents a significant improvement in agreement for the KB integrals over the CHARMM model.

Figure 3.4 displays the composition dependent density and partial molar volumes (pmv) for NMA and water mixtures. The density was consistently underestimated but the error was small ($< 1\%$) and almost identical to the error in the density for pure water. The partial molar volumes were reasonably well reproduced except for low mole fractions ($x_2 < 0.2$) where the NMA pmv was slightly underestimated and the water pmv slightly overestimated. The pmvs

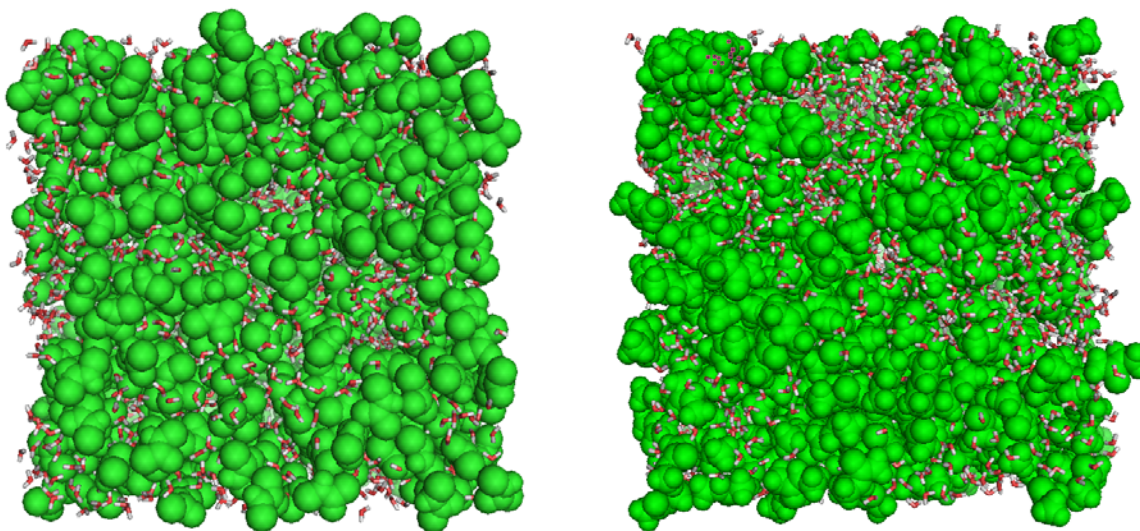


Figure 3.3 The snapshot of the simulated aqueous NMA solution at $x_2=0.2$ with KBFF (left) and CHARMM (right). While the simulation with the KBFF model displayed random distribution, the simulation with CHARMM displayed self aggregation of NMA molecules.

obtained from a fit to the simulated densities are also included in Figure 3.4. The agreement with the pmvs obtained from the simulated KB integrals suggested that the approximation used in Equation 3.1 was reasonable.⁵

Using the simulated KB integrals and Equation 3.4 one can obtain a fit to the parameters of Equation 3.2. The results are displayed in Figure 3.5 along with the enthalpy of mixing. The excess Gibbs energy was reasonably well reproduced with the largest error again appearing for

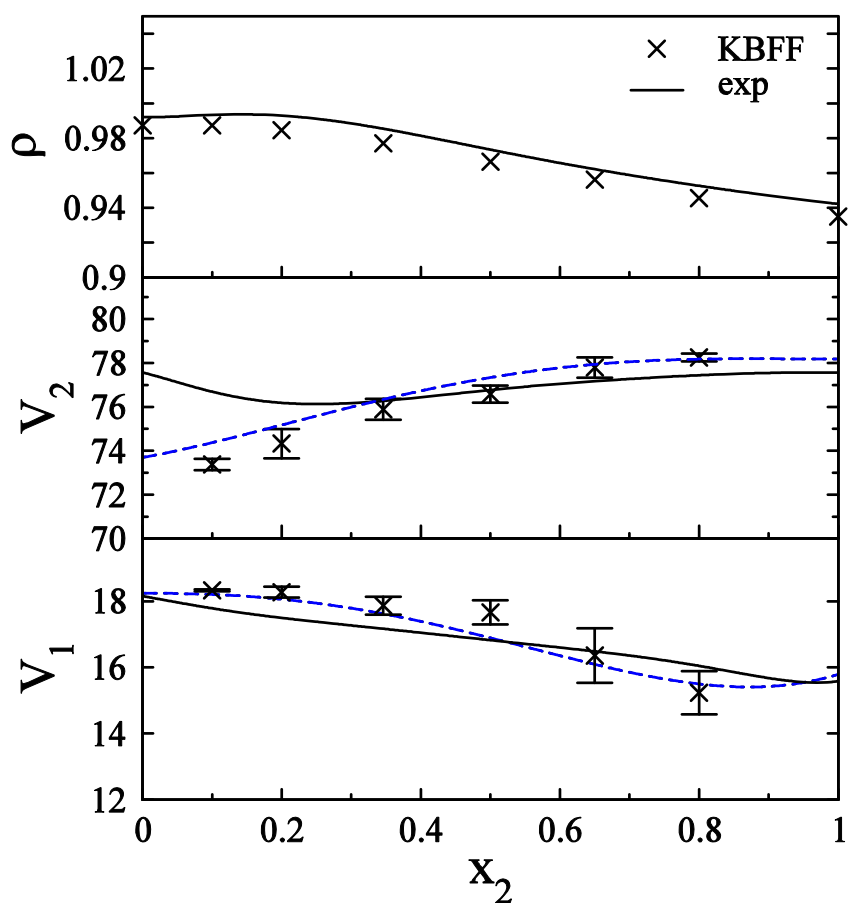


Figure 3.4 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for NMA solutions. The crosses were obtained directly from the KB integrals. The dashed lines were obtained by fitting the simulated excess volumes to Equation 3.2 and then applying Equation 3.3.

low NMA mole fractions. The most interesting observation was for the enthalpy of mixing. The experimental data, at the slightly different temperature of 308K, displays a large and favorable mixing enthalpy between NMA and water.⁴⁷ This was essentially reproduced by the KBFF model with small errors at, somewhat surprisingly, larger NMA mole fractions. In comparison, the CHARMM simulations resulted in a very small mixing enthalpy for the compositions considered here. This suggests that the self aggregation observed at low NMA mole fractions may be a consequence of a low solvation enthalpy and corresponding Gibbs energy. A similar low enthalpy of mixing and high self aggregation was also observed during our studies of NaCl and water mixtures.^{1,48}

The NMA force field was extended to include DMA and ACT. The simulated results for DMA and water mixtures are presented in Figures 3.5 and 3.6. The excess coordination numbers for mole fractions of $x_2 = 0.1$ and 0.5 are compared to the experimental data in Figure 3.6. The same trend of slightly too much self aggregation at low mole fractions observed for NMA was also found in the case of DMA. However, in our opinion the estimated errors in the simulated data, and the unknown but potentially substantial errors in the experimental N_{ij} values for low x_j , did not warrant further changes in the parameters for DMA. The density and pmvs for DMA are displayed in Figure 3.7. The density of pure DMA was observed to be slightly high, although the magnitude of the deviation was similar to our recent model for methanol.⁵ Unfortunately, the error in the density could not be reduced with reasonable modifications to the partial atomic charges. Unfortunately, the only difference between NMA and DMA is the additional methyl group in DMA, which has the same LJ parameters as the other methyl groups in NMA and also those used previously.^{2,5} Therefore, in an effort to maintain some consistency, and as we did not

consider the errors significant enough, new atom types were not introduced to compensate for the difference in density.

Experimental data concerning the activity and density of ACT in water at 313K could not be found. Therefore, limited data describing the solution density and activity of water at 298K was used to determine KB integrals via the Gibbs-Duhem equation.²³ It was then assumed that the difference in temperature has only a small affect on the KB integrals. The experimental data

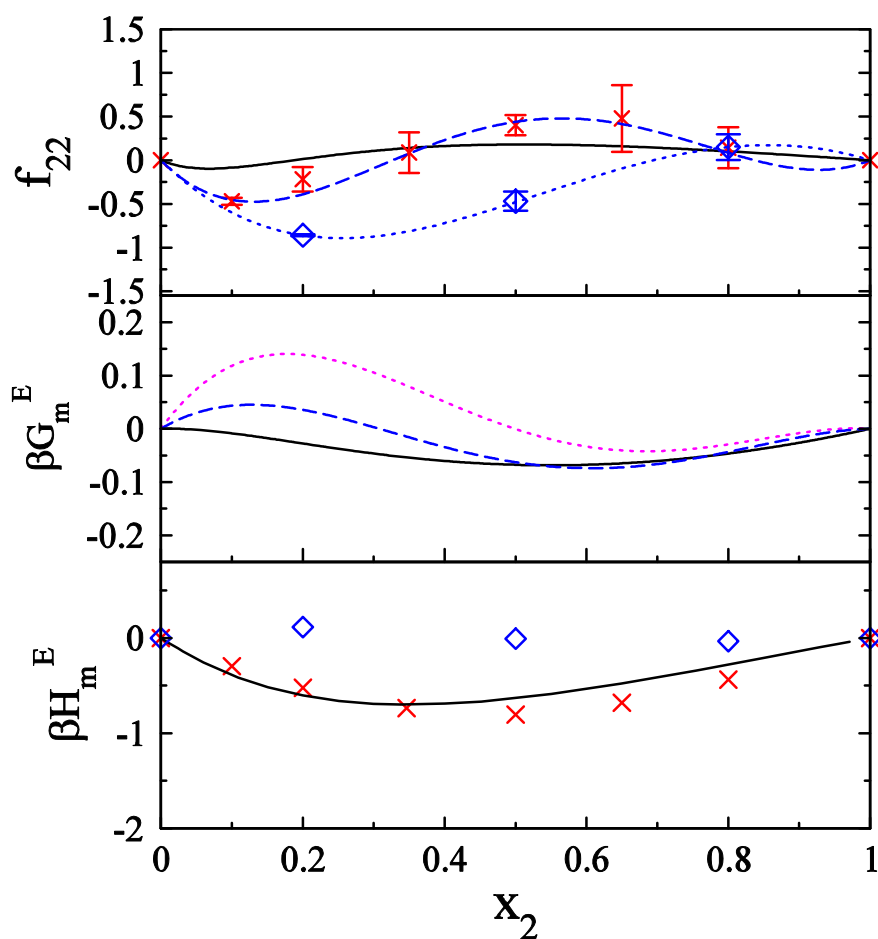


Figure 3.5 Activity derivative (f_{22}) and excess Gibbs energy (G) and enthalpy of mixing (H) for NMA solutions. The solid lines correspond to the experimental data (H at 308K), the crosses and dashed lines to the KBFF model, and the diamonds and dotted lines to the CHARMM model.

are displayed in Figure 3.8 together with the simulated data for ACT and water mixtures of $x_3 = 0.05$ and 0.10 . The results for the KBFF model were in excellent agreement with experiment. As a further test of the ACT force field a simulation of the orthorhombic crystal structure was also performed. The experimental unit cell dimensions were used to build an (approximately) cubic simulation box containing $5 \times 2 \times 4$ unit cells and a box length of 3.8 nm. A simulation of the system using anisotropic pressure scaling resulted in unit cell dimensions of $a = 0.791$, $b = 1.838$ and $c = 0.959$ nm. These compare well to the experimental values of 0.776 , 1.900 , and 0.951 nm,⁴¹ respectively. The corresponding experimental and simulated crystal densities were 1.119 and 1.129 g/cm³, respectively. In particular, the c dimension, which contains an infinite hydrogen bonded chain network, was well reproduced suggesting the amide hydrogen bonding distances and angles were accurately described by the current model.

A summary of the basic properties of the pure liquids of NMA and DMA is presented in Table 3.4. The diffusion constant for pure NMA was slightly higher than experiment, while the density and compressibility were in good agreement. The most significant difference between experiment and simulation occurred for the relative permittivity. Both the KBFF and CHARMM simulations predicted large relative permittivities, but only 30% of the experimental value. The reasons for this disagreement are not fully clear. The total dipole moment fluctuations appeared to have converged sufficiently during the present multi nanosecond runs to provide reasonable estimates of the true values.⁴⁹ A possible problem lies in the relatively long experimental Debye relaxation time for pure NMA. As computer simulations of the permittivity of pure water typically require multiple nanosecond simulations,⁴⁹ a relaxation time of 50 ps for NMA,⁴¹ compared to 9 ps for pure water,⁵⁰ suggests that additional sampling may be required to fully capture the dielectric response corresponding to longer time fluctuations in pure NMA. Alternatively, the

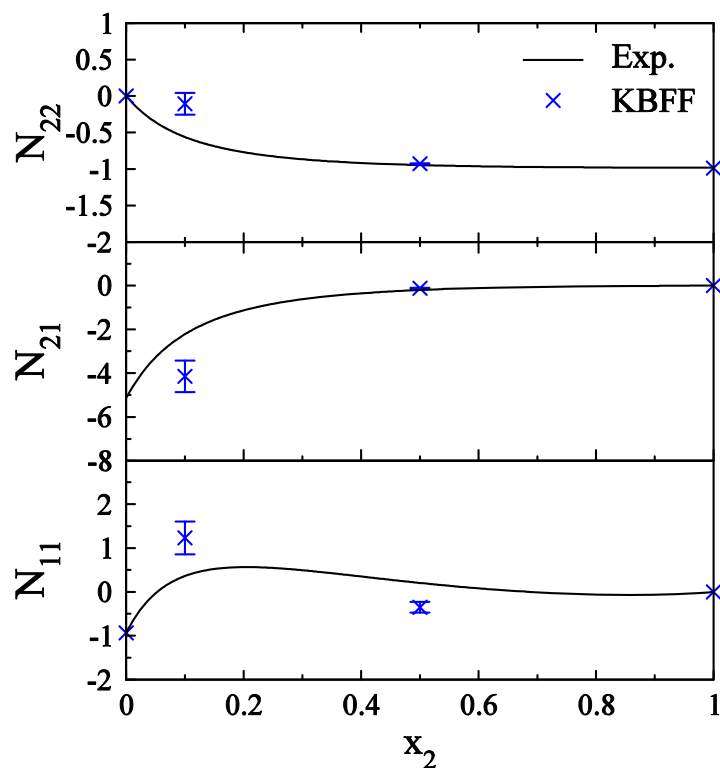


Figure 3.6 Excess coordination numbers (N_{ij}) as a function of composition for DMA solutions.

dipole moment obtained for NMA using our fixed charge distribution is 4.6 D, which is smaller than the average dipole moments obtained from polarizable NMA models,^{7,11} and could be too low. Unfortunately, neither of the polarizable force fields quoted a value for the dielectric constant.

A comparison of the properties obtained from the KBFF model for NMA in combination with different simple three site water models is provided in Table 3.5 for equimolar mixtures. As observed for other solutes,^{3,4} a change in the water model had only a small effect on the excess coordination numbers. The agreement with experiment was improved slightly for the TIP3P model, although the density was not as accurate. Larger deviations between the models were

observed for the diffusion constants, especially for the diffusion of water, with the differences in agreement with the pure water values. However, no experimental data on water and NMA mixtures was available for comparison. In our opinion, the KBFF model of NMA can be used with confidence with any of these simple water models and will provide a reasonable balance between solute-solute and solute-solvent interactions. A similar conclusion can be inferred for DMA and ACT solutions.

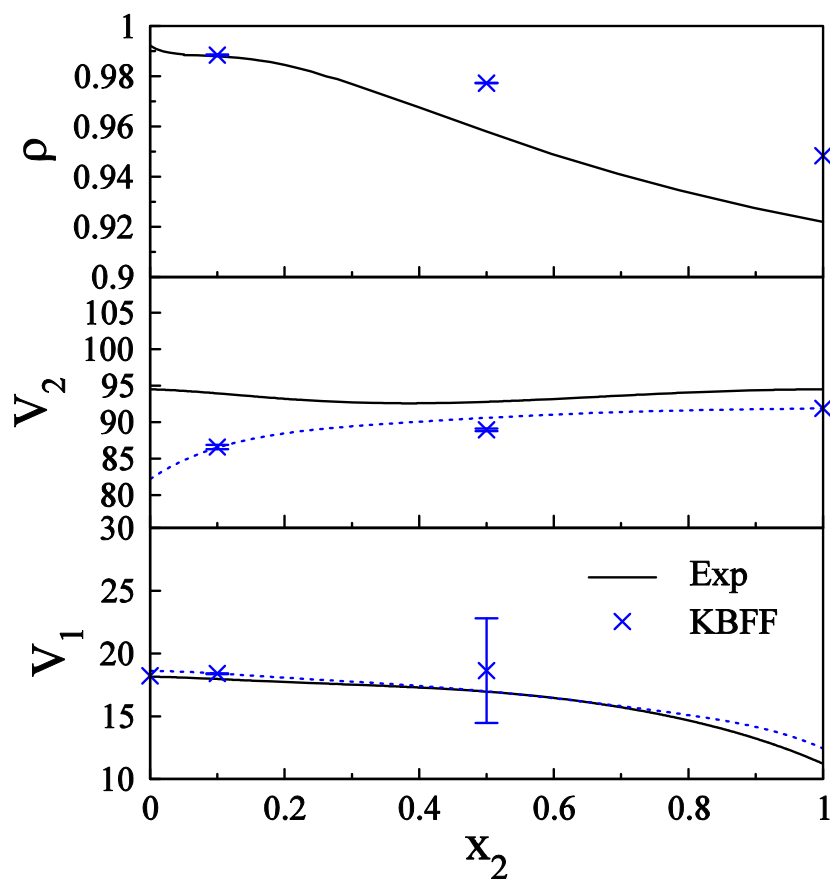


Figure 3.7 Density (g/cm³) and partial molar volumes (cm³/mol) as a function of composition for DMA solutions. The crosses were obtained directly from the KB integrals. The dashed lines were obtained by fitting the simulated excess volumes to Equation 3.2 and then applying Equation 3.3.

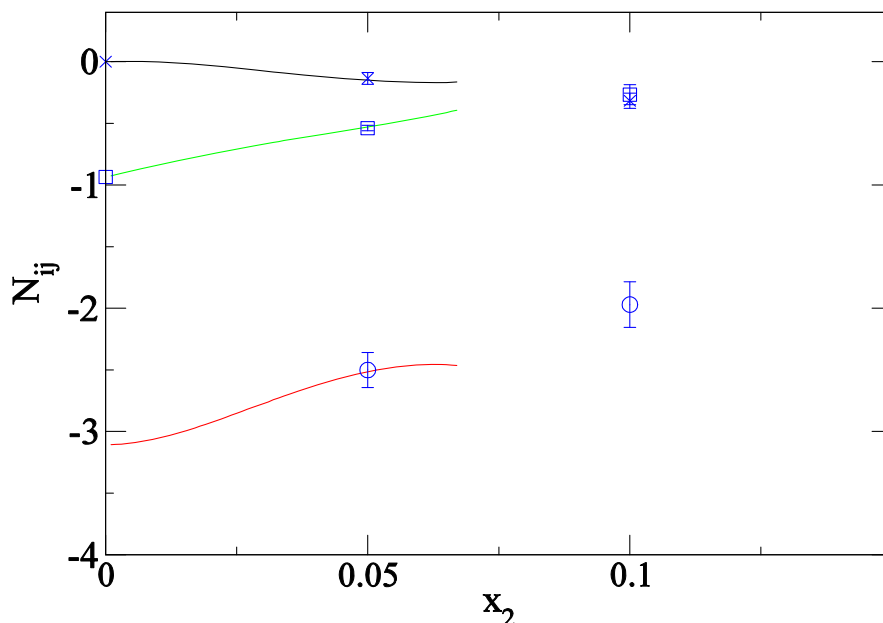


Figure 3.8 Excess coordination numbers (N_{ij}) as a function of composition for ACT solutions. The black, red and green lines correspond to the experimental data at 298K for N_{22} , N_{11} , and N_{21} , respectively. The crosses, diamonds and circles corresponding to the respective data from the KBFF model obtained at 313K.

To further investigate the degree of self aggregation observed in the solutions the atom based rdfs and coordination numbers between NMA molecules, and between NMA and water molecules, were determined and are presented in Table 3.6. The KBFF model displayed a general overall increase in the solvation of both the carbonyl oxygen and amide hydrogen compared to the CHARMM force field at $x_2 = 0.2$. Correspondingly, the KBFF model displayed a lower number of NMA to NMA hydrogen bonds. The largest difference was observed for the solvation of the amide hydrogen by water oxygen. Here, the KBFF model produced an increase of 30% over the CHARMM model. However, it should be noted that while the trends in the first coordination numbers mimicked the differences between the KBFF and CHARMM models the largest differences in solution structure occurred for the longer range distributions, as quantified

by the KB integrals. The interaction of water with the carbonyl group was the same for all three amides and corresponded to two water hydrogen bonds. The same degree of solvation by water was observed for the N-H hydrogens of NMA and ACT, and no significant difference was observed between the number of water oxygens interacting with the cis and trans hydrogens of ACT. In the pure NMA solution an average of one C=O to N-H hydrogen bonds was observed. This was doubled for the ACT crystal simulation. The number of NMA to NMA hydrogen bonds was low (15%) in the NMA and water mixtures.

Table 3.4 Properties of the pure liquids. Experimental data: density from refs 26-28; diffusion constants from refs 57,58; dielectric constants from refs 59,60; predicted compressibilities from refs 60,61; and thermal expansion coefficient from ref 60. Intramolecular contributions to E_{pot} were 0 and -107.9 kJ/mol for the KBFF and CHARMM models, respectively. Average molecular dipole moments were 4.6, 5.5, and 4.7 D for NMA, DMA, and ACT, respectively.

		ρ g/cm ³	D $\times 10^{-9}$ m ² /s	ϵ	κ_T 10^{-5} atm ⁻¹	α $\times 10^{-5}$ K ⁻¹	E_{pot} kJ/mol
NMA	KBFF	0.935	0.7	52	7.5	92	-47.38
	CHARMM	0.954	1.1	53	5.9	94	-155.40
	Exp	0.942	0.46	191,166	6.3		
DMA	KBFF	0.948	1.1	39	5.3	89	-41.41
	Exp	0.922			6.3		
water	SPC/E	0.987	3.5	69	4.6	72	-45.64
	Exp	0.995	3.7	70	4.4	39	

Table 3.5 Properties of NMA and water mixtures ($x_2=0.5$). Experimental data were taken from refs 25,26,47, and 62.

	KBFF SPC/E	KBFF SPC	KBFF TIP3P	CHARMM TIP3P	Exp	Units
ρ	0.966	0.958	0.955	0.954	0.974	g/cm^3
C_2	10.61	10.52	10.48	10.47	10.69	Mol/l
N_{22}	-0.92	-0.93	-0.91	-0.82	-0.93	
N_{21}	-0.21	-0.18	-0.21	-0.60	-0.23	
N_{11}	-0.03	-0.12	0.04	1.45	0.14	
f_{22}	0.40	0.53	0.35	-0.48	0.18	
\bar{V}_2	76.6	76.9	76.1	76.0	77.3	cm^3/mol
\bar{V}_1	17.7	18.1	19.3	19.5	16.9	cm^3/mol
D_1	0.55	0.99	1.31	3.65		$\times 10^{-9} \text{m}^2/\text{s}$
D_2	0.45	0.59	0.72	1.65		$\times 10^{-9} \text{m}^2/\text{s}$
ε	38	48	51	47		
κ_T	5.1	6.0	6.0	5.6		10^{-5}atm^{-1}
H_m^E	-2.09		-1.59	0.02	-1.64	kJ/mol

NMA provides a good model for the interaction of the peptide group with itself and with water.⁵¹⁻⁵³ The interaction between NMA molecules at low concentrations (high water content) should provide a good model for the interaction of peptide groups exposed to solvent, i.e. as observed for the denatured state or in the early stages of protein folding. The pmf for the

interaction of NMA pairs can be obtained from the corresponding center of mass based rdf, using the fact that $W(r) = -RT \ln g_{22}(r)$. The result is presented in Figure 3.9 and displayed a rather shallow interaction minimum indicating a small difference in interactions between solute – solute and solute – solvent hydrogen bonds (< 1 kJ/mol) in dilute solutions. A more pronounced minimum was observed in pure NMA where solvation is absent. The above results are somewhat different from the previous results of -3 kJ/mol from Pranata,⁵² and the pmfs corresponding to the distributions obtained from analysis of the protein database.⁵⁴

Table 3.6 Atom based first shell coordination numbers.

x_2			O...H ₂ O	NH...OH ₂	O...HN
		r_{min}	0.255	0.255	0.285
0.1	KBFF	NMA	1.94	0.78	0.15
	KBFF	DMA	2.02		
	KBFF	ACT	2.04	0.86/0.83	0.15/0.16
0.2	KBFF	NMA	1.68	0.70	0.24
	CHARMM	NMA	1.40	0.55	0.32
1.0	KBFF	NMA			0.98
	KBFF	ACT			1.01/1.01

Further analysis of hydrogen bonded NMA pairs, defined as having a O-H distance of less than 0.275 nm, was performed by determining the probability distribution corresponding to the angle (θ) between the C=O and N-H dipoles. The results are also displayed in Figure 3.8. Most peptide group hydrogen bonding interactions in proteins have a dipole angle of between 0

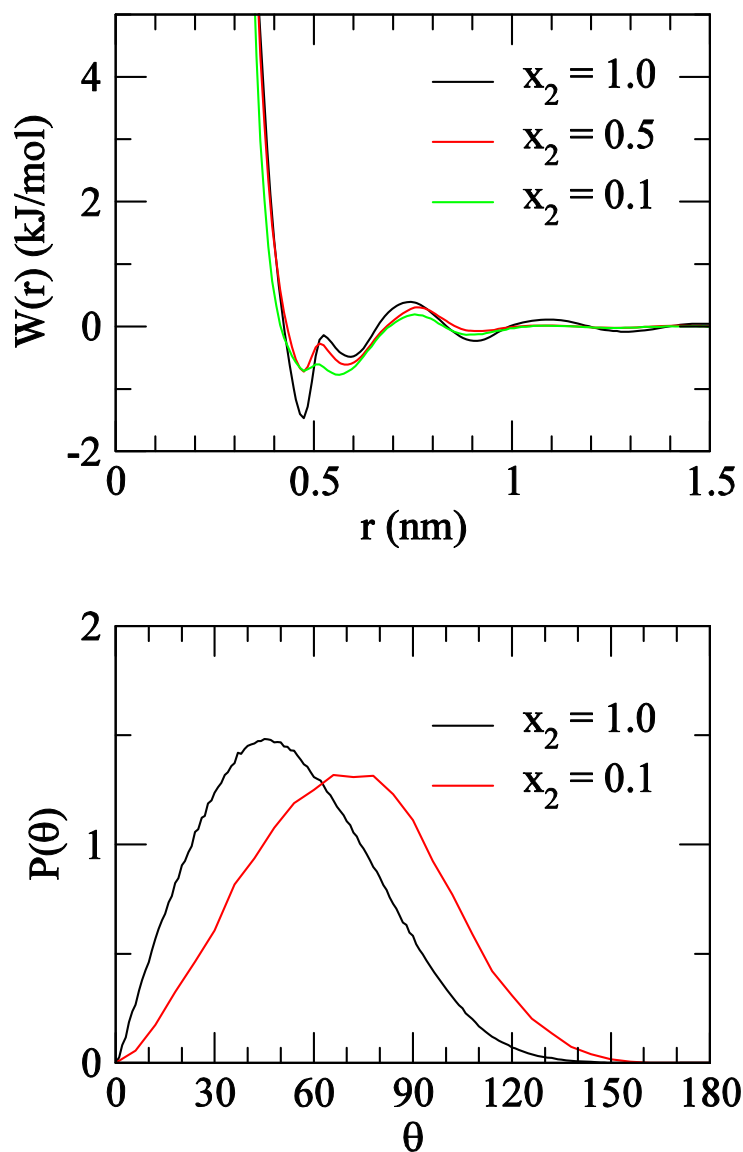


Figure 3.9 Potential of mean force profiles obtained from the center of mass rdfs in NMA solutions (top). Probability distribution for the angle (θ) between the N-H and C=O dipoles of hydrogen bonded ($r_{HO} < 0.275$ nm) NMA molecules (bottom).

and 30 degrees, with a linear hydrogen bond corresponding to 0 degrees. The most probable angle corresponded to 45 degrees in pure NMA, which was increased to 65 degrees for the $x_2 = 0.1$ mixture. However, all angles less than 90 degrees were significantly populated in both

solutions. The distribution indicated that 90% of the hydrogen bonding interactions occurred with angles less than 90 degrees in pure NMA, while 80% of the interactions had an angle of less than 90 degrees in the more dilute solution. The above observations, in conjunction with the NMA-NMA pmf data, suggested that the hydrogen bonding interaction between NMA molecules at low cosolvent compositions was relatively non specific with no strong directional hydrogen bonding interactions.

Another approach to quantify the NMA self interaction in dilute solutions is also based on the KB integrals. The quantity $G_{22} - G_{21}$ describes the preferential interaction of NMA with other NMA molecules as a function of composition. Rewriting the preferential interaction as $n_{21}/\rho_2 (n_{22}/n_{21} - \rho_2/\rho_1)$, indicates that the sign of the preferential interaction is provided by the expression in parenthesis, and is positive if the ratio of NMA to water in the vicinity of an NMA molecule exceeds the ratio of NMA to water in the bulk solvent. The preferential interaction at infinite dilution ($G_{22}^\infty - G_{21}^\infty$) is potentially very informative and describes the preferential interaction between two NMA molecules in pure water. Unfortunately, the experimental and simulated values of G_{22}^∞ vary depending on the exact fitting procedure used. Using Equation 3.2 the KBFF approach gave a value of $G_{22}^\infty = 23 \text{ cm}^3/\text{mol}$, compared to experimental values of $-82 \text{ cm}^3/\text{mol}$ obtained from the analysis presented here, and values of 0 and $-132 \text{ cm}^3/\text{mol}$ described elsewhere.^{12,31,55,56} A value of $G_{21}^\infty = -73 \text{ cm}^3/\text{mol}$ is consistent with all the analyses. Hence, the KBFF approach resulted in a value of $G_{22}^\infty - G_{21}^\infty = 96 \text{ cm}^3/\text{mol}$, compared to experimental values of -56 and $76 \text{ cm}^3/\text{mol}$. Obviously, there is some disagreement in the experimental values which are consistent with either a small net hydration or self association. In this respect, the KBFF value appears reasonable.

3.4 Conclusions

A force field for the simulation of aqueous solutions of NMA, DMA, and ACT has been developed which attempts to reproduce the experimental KB integrals over a range of cosolvent concentrations thereby providing a reasonably accurate description of the balance between solvation and cosolvent self aggregation in these systems. The force field can be used as the basis for a description of peptide and protein backbone interactions. Analysis of relatively dilute solutions of NMA suggests that the molecules are highly solvated with no apparent strong directional hydrogen bonds between NMA molecules. Hence, the results suggest that only as the peptide backbone becomes almost fully desolvated will one tend to observe strong directional intramolecular peptide group hydrogen bonding during the folding processes.

3.5 References

1. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 119, 11342-11349.
2. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 10663-10670.
3. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2003, 107, 3891-3898.
4. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2004, 121, 2180-2186.
5. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2005, 109, 15080-15086.
6. Caldwell, J. W.; Kollman, P. A. *Journal of Physical Chemistry* 1995, 99, 6208-6219.
7. Rick, S. W.; Berne, B. J. *Journal of the American Chemical Society* 1996, 118, 672-679.
8. Patel, S.; Brooks, C. L. *Journal of Computational Chemistry* 2004, 25, 1-15.
9. Zhang, R.; Li, H. R.; Lei, Y.; Han, S. J. *Journal of Physical Chemistry B* 2004, 108, 12596-12601.
10. Sandberg, L.; Casemyr, R.; Edholm, O. *Journal of Physical Chemistry B* 2002, 106, 7889-7897.
11. Iuchi, S.; Morita, A.; Kato, S. *Journal of Physical Chemistry B* 2002, 106, 3466-3476.
12. Zielkiewicz, J. *Physical Chemistry Chemical Physics* 2000, 2, 2925-2932.
13. Decamp, M. F.; DeFlores, L.; McCracken, J. M.; Tokmakoff, A.; Kwac, K.; Cho, M. *Journal of Physical Chemistry B* 2005, 109, 11016-11026.
14. Mennucci, B.; Martinez, J. M. *Journal of Physical Chemistry B* 2005, 109, 9818-9829.
15. Mennucci, B.; Martinez, J. M. *Journal of Physical Chemistry B* 2005, 109, 9830-9838.
16. Morgantini, P. Y.; Kollman, P. A. *Journal of the American Chemical Society* 1995, 117, 6057-6063.
17. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 115, 5521-5530.
18. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 5901-5910.

19. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 114, 426-435.
20. Ben-Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*; Plenum Press: New York, 1992.
21. Kirkwood, J. G.; Buff, F. P. *Journal of Chemical Physics* 1951, 19, 774-777.
22. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 18716-18724.
23. Woldan, M. *Acta Universitatis Lodziensis, Folia Chimica* 1988, 8, 35-43.
24. Zielkiewicz, J. *International DATA Series, Selected Data on Mixtures, Series A: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances* 2004, 32, 232.
25. Zielkiewicz, J. *International DATA Series, Selected Data on Mixtures, Series A: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances* 2000, 28, 287.
26. Zielkiewicz, J. *International DATA Series, Selected Data on Mixtures, Series A: Thermodynamic Properties of Non-Reacting Binary Systems of Organic Substances* 2000, 28, 288.
27. Jelinska-Kazimierczuk, M.; Szydlowski, J. *Journal of Solution Chemistry* 2001, 30, 623-640.
28. Scharlin, P.; Steinby, K. *Journal of Chemical Thermodynamics* 2003, 35, 279-300.
29. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2002, 106, 1491-1500.
30. Redlich, O.; Kister, A. T. *Industrial and Engineering Chemistry* 1948, 40, 345-348.
31. Marcus, Y. *Journal of the Chemical Society, Faraday Transactions* 1990, 86, 2215-2224.
32. Matteoli, E.; Lepori, L. *Journal of Chemical Physics* 1984, 80, 2856-2863.

33. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* 1987, 91, 6269-6271.
34. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *Journal of Chemical Physics* 1984, 81, 3684-3690.
35. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* 1977, 23, 327-341.
36. Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* 1993, 98, 10089-10092.
37. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2000, 104, 5854-5864.
38. Smith, P. E.; van Gunsteren, W. F. *Journal of Chemical Physics* 1994, 100, 3169-3174.
39. Walser, R.; Mark, A. E.; van Gunsteren, W. F.; Lauterbach, M.; Wipff, G. *Journal of Chemical Physics* 2000, 112, 10450-10459.
40. Daura, X.; Mark, A. E.; van Gunsteren, W. F. *Journal of Computational Chemistry* 1998, 19, 535-547.
41. Hamilton, W. C. *Acta Cryst* 1965, 18, 866-870.
42. Callomon, J. H.; Hirota, E.; Kuchitsu, K.; Lafferty, W. J.; Maki, A. G.; Pote, C. S. *Landolt-Börnstein: Numerical Data and Functional Relationships in Science and Technology*; Springer-Verlag Berlin: Heidelberg, New York, 1976.
43. Langley, C. H.; Allinger, N. L. *Journal of Physical Chemistry A* 2002, 106, 5638-5652.
44. van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hunenberger, P. H.; Kruger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. *Biomolecular Simulation: The GROMOS96 manual and user guide.*; VdF Hochschulverlag ETHZ.: Zurich, 1996.
45. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.;

- Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorcikiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* 1998, 102, 3586-3616.
46. Perera, A.; Sokolic, F. *Journal of Chemical Physics* 2004, 121, 11272-11282.
47. Zaichikov, A. M.; Golubinskii, O. E. *Zhurnal Fizicheskoi Khimii* 1996, 70, 1175-1179.
48. Weerasinghe, S.; Smith, P. E.; Pettitt, B. M. *Biochemistry* 1995, 34, 16269-16278.
49. Heinz, T. N.; van Gunsteren, W. F.; Hunenberger, P. H. *Journal of Chemical Physics* 2001, 115, 1125-1136.
50. Hasted J. B. *Water: A Comprehensive Treatise*; Plenum: New York, 1973; Chapter 7.
51. Yu, H. A.; Pettitt, B. M.; Karplus, M. *Journal of the American Chemical Society* 1991, 113, 2425-2434.
52. Pranata, J. *Journal of Physical Chemistry* 1995, 99, 4855-4859.
53. Stultz, C. M. *Journal of Physical Chemistry B* 2004, 108, 16525-16532.
54. Sippl, M. J. *Journal of Molecular Biology* 1996, 260, 644-648.
55. Marcus, Y. *Physical Chemistry Chemical Physics* 2002, 4, 4462-4471.
56. Marcus, Y. *Monatshefte fur Chemie* 2001, 132, 1387-1411.
57. Williams, W. D.; Ellard, J. A.; Dawson, L. R. *Journal of the American Chemical Society* 1957, 79, 4652-4654.
58. Pruppacher, H. R. *The Journal of Chemical Physics* 1971, 56, 101-107.
59. Bonner, O. D.; Woolsey, G. B.; Philips, R. H. *Journal of Chemical and Engineering Data* 1968, 13, 218-220.
60. *Handbook of Chemistry and Physics*; CRC Press, Inc.: Boca Raton, Florida, 1986.

61. Marcus, Y.; Hefter, G. T. *Journal of Molecular Liquids* 1997, 73-4, 61-74.
62. Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *Journal of Chemical Physics* 1995, 102, 5451-5459.

Supplementary Data for Chapter 3

Table 3.7 Center of mass based first shell coordination numbers. **id**; infinite dilution (one solute molecule).

	r_{\min}	id	0.05	0.1	0.2	0.35	0.5	0.65	0.8	1.0
NMA	22	0.765		4.76	7.3	9.76	11.41	12.45	13.13	13.66
	21	0.435	7.08	4.98	3.94	2.89	2.05	1.33	0.77	
	11	0.345	5.11	4.24	3.57	2.75	1.98	1.35	0.71	
DMA	22	0.775		4.5			10.4			12.72
	21	0.665	36.96	23.97			7.32			
	11	0.345	5.11	4.2			1.94			
ACT	22	0.655		1.78	3.14					14.39
	21	0.605	28.43	24.42	21.26					
	11	0.345	5.11	4.70	4.33					

Table 3.8 Fitting constants for Equation 3.2.

	X		a ₀	a ₁	a ₂	a ₃	a ₄	rmsd	units
NMA	βG	Exp	-0.265	0.07301	0.09361	0.08193	0.04369	0.0002	
		KBFF	-0.2510	0.4121	0.6283			0.008	
		CHARMM	1.0292	0.8767	0.0692			0.0	
	βH	Exp	-2.5231	-1.6283	-0.6939	-0.1509	0.11608	0.001	
		KBFF	-3.206	-0.4503	0.9143	0.0700	-0.8576	0.002	
		CHARMM	-0.0236	0.7677	0.76515			0.0	
	V	Exp	-4.2683	1.9691	-0.145	-1.445	1.318	0.004	
		KBFF	-4.3844	-1.0171	0.9132			0.004	cm ³ /mol
		CHARMM	0.5744	1.2905	1.1281			0.0	
DMA	βG	Exp	-0.6107	-0.1356	-0.0329	-0.0094	-0.0031	0.0	
		KBFF							
		V	Exp	-5.762	-2.5203	-0.6182	1.8626	2.5088	0.006
		KBFF	-5.782	-1.372	-1.698			0.0	
ACT	βG	Exp	-0.0094	0.00091	-0.0038	-0.002	0.00245	0.0001	
		V	Exp	22.4	-99.8	237.3	-260.5	106.8	0.008

CHAPTER 4 - A Comparison of Force Fields for Amides and Glycine

Abstract

A Kirkwood-Buff derived force field (KBFF) for the computer simulation of aqueous solutions of amides is presented in chapter 3. Here, the KBFF is compared with results from existing force fields for the aqueous solutions of N-methylacetamide (NMA) and the glycine zwitterionic system. NMA represents a model for a peptide group. Glycine, the simplest amino acid, is selected as an example for ionic interactions. Experimental properties such as density, partial molar volume, and heat of mixing, as well as the excess coordination numbers, are compared for a variety of common biomolecular force fields. No one force field, with the exception of KBFF, accurately reproduces the properties of both solutes.

4.1 Introduction

In computer simulations a force field is a critical factor determining the quality of the simulations. It is important that a force field can reproduce the correct distribution of molecules in a system by keeping the delicate balance of solute-solute interactions and solute-solvent interactions. In particular, for simulations of systems including peptides and proteins, a force field needs to be able to balance hydrogen bonding between peptide groups against the degree of solvation of the peptide groups. Recently, Kirkwood-Buff (KB) integrals have been used to quantify the intermolecular interactions in the solution mixtures. KB integrals provide a measure of the degree to which a force field represents the correct distribution of molecules in solution^{1,2}.

We have continuously endeavored to develop accurate force fields for the simulation of solution mixtures and their application in biomolecular systems³⁻²⁴. As part of such an effort, a Kirkwood-Buff derived force field (KBFF) for the computer simulation of aqueous solutions of amides and other systems is presented in our recent paper⁸. Here the KBFF is compared with existing force fields for aqueous N-methylacetamide (NMA) and the aqueous glycine systems. NMA represents a model for a peptide group. Proteins are large molecules with residues connected with peptide bonds. Hence, even a small error in the peptide-peptide interaction can accumulate to make a big effect. Hence it is important to have accurate force field for NMA. Glycine is the simplest one of amino acids which consists of fundamental building blocks for most biomolecules. Glycine is selected as a model for the interactions of charged side chains.

In addition to the KB integrals, experimental properties such as density, partial molar volume, and heat of mixing, as well as the excess coordination numbers, are compared for a series of force fields.

4.2 Methods

A KB analysis of the experimental data for the cosolvent (2) and water (1) mixtures was performed as outlined by Ben-Naim²⁵, and in our own previous studies⁷. The KB integrals (G_{ij}) are integrals over rdfs in the μVT ensemble as defined as,

$$G_{ij} = 4\pi \int_0^{\infty} [g_{ij}^{\mu VT}(r) - 1] r^2 dr \approx 4\pi \int_0^{R_c} [g_{ij}^{NPT}(r) - 1] r^2 dr \quad (4.1)$$

where g_{ij} is the corresponding radial distribution function (rdf), and the approximation is made for simulations performed in closed system (NPT ensemble)^{10,26}.

All mixtures were simulated by classical molecular dynamics techniques using the Gromacs program²⁷. The simulations were performed in the isothermal isobaric (NpT) ensemble at 313 K and 1 atm for NMA solutions and at 300K and 1atm for glycine solutions. The weak coupling technique²⁸ was used to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively. All bonds were constrained using SHAKE²⁹ and a relative tolerance of 10^{-4} , allowing a 2 fs timestep for integration of the equations of motion. The particle mesh Ewald technique was used to evaluate the electrostatic interactions³⁰. Both distances for Coulomb cutoff and the Lennard-Jones cutoff were 1.5 nm. Random initial configurations of the solute and water molecules in a cubic box of approximate length 5 nm were used. The steepest descent method was then used to perform 100 steps of minimization. This was followed by extensive equilibration which was continued until all intermolecular potential energy contributions and rdfs displayed no drift with time (typically 1 ns). Configurations were saved every 1 ps for analysis. Simulations were performed for 10 ns for each system.

The excess enthalpies of mixing (H_m^E) was determined from the average potential energies (E_{pot}) via

Table 4.1 Summary of the MD Simulations of NMA and Glycine Solutions.

	N_2	N_1	x_2	$m_2, mol/kg$	T_{sim}, ns
NMA	318	2860	0.1001	6.17	10
	515	2059	0.2001	13.88	10
	804	804	0.5000	55.51	10
	970	0	1.0000		10
Glycine	102	3779	0.0263	1.49	10
	201	3779	0.0505	2.95	10

$$H_m^E = E_{pot,mix} - x_2 E_{pot,2} - (1 - x_2) E_{pot,1} \quad (4.2)$$

where $E_{pot,mix}$ is the potential energy of the solution mixture, x_2 the mole fraction of cosolvent, and $E_{pot,1}$ and $E_{pot,2}$ the potential energy of pure water and pure cosolvent, respectively.

The KBFF parameters for amides⁸ and glycine³¹ were compared with the following protein force fields using the parameters presented for the peptide group: AMBER03, CHARMM22 all atom force field, GROMOS45a3, and OPLS all atom force field. For water, each force field was coupled with the corresponding water models as recommended by the developers: KBFF with SPC/E, GROMOS45a3 with SPC, AMBER03, CHARMM22, and OPLS with TIP3P^{8,32,33}.

4.3 Results

The NMA solution mixtures displayed in Table 4.1 were simulated using each force field. Examples of the center of mass based rdfs obtained for mixtures with $x_2=0.1$ are displayed in Figure 4.1. The rdfs indicated that g_{11} was very similar for all four force fields except GROMOS45a3, while more distinct differences among all force fields were observed in g_{22} and

g_{21} . The first shell solute-solute interactions increased in the order of KBFF, AMBER03, CHARMM22, OPLS, and GROMOS45a3. The first shell solute-solvent interactions decreased in the same order. This is to be expected considering the balance of the intermolecular interactions among solute-solute, solvent-solvent and solute-solvent: the stronger the solute-solute and solvent-solvent interactions are, the weaker the solute-solvent interaction is formed. When the solute-solute interaction is stronger than the solvent-solute interaction, it is more likely to

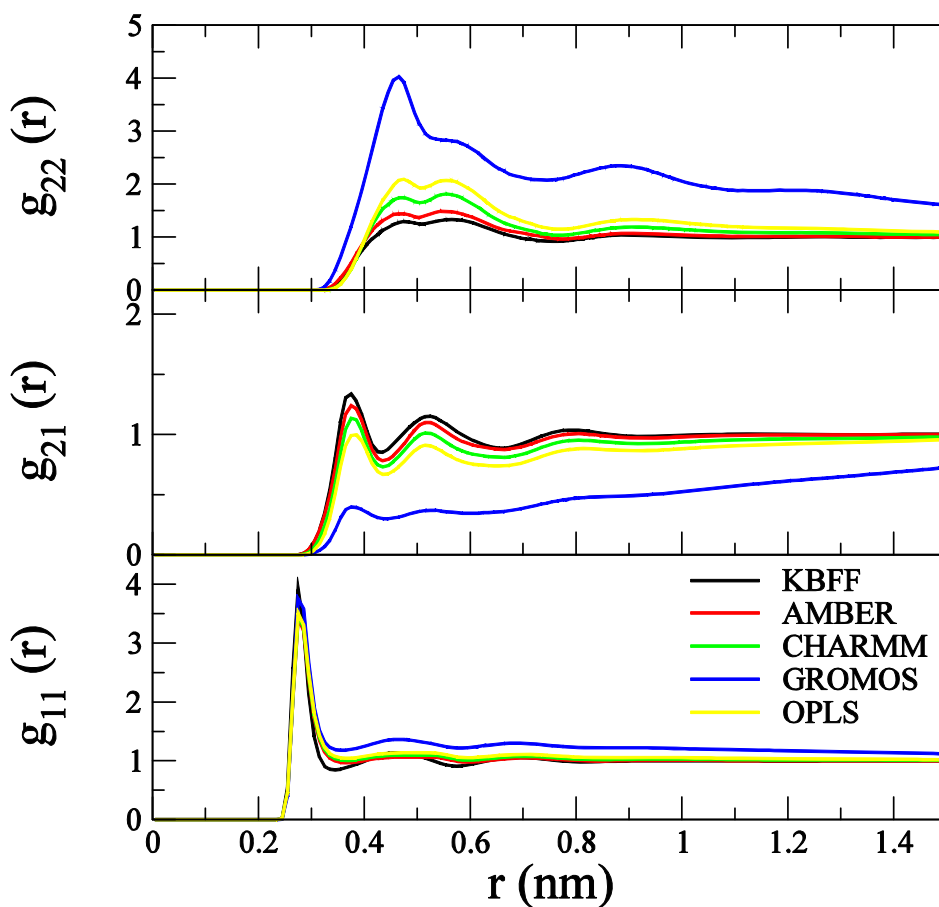


Figure 4.1 Center of mass based rdfs for $x_2 = 0.1$ solutions of NMA (2) in water (1). As the rdf for NMA to NMA, g_{22} , increased, the rdf for NMA to water, g_{21} , decreased. Notice that g_{22} provided by GROMOS45a3 didn't approach unity within the studied distance, indicating high self aggregation of solutes.

observe aggregation of the solute molecules due to a lack of solvation. This trend was the most salient in GROMOS45a3, which may be explained by the fact that it has the smallest charge distribution polarity of all the force fields used here. In particular, g_{22} of GROMOS45a3 didn't approach unity within the studied distance. It remained larger than 1, indicating aggregation of solutes. As in our previous study, 1.2 and 1.6 nm were used as an averaging range for KB integrals in systems with KBFF and AMBER03. CHARMM22, OPLS, and GROMOS45a3 displayed some weak structure beyond 1.6 nm. Hence, the KB integrals for the systems with these three force fields were obtained from the simulated data by averaging the values between 1.8 and 2.2 nm, a larger distance than the other two force fields.

The excess coordination numbers ($N_{ij}=\rho_j G_{ij}$) obtained from the simulations of NMA and water mixtures with each force field are compared to the experimental data in Figure 4.2. The KBFF model reproduced well the experimental data for all solute concentration, including lower solute concentration where the KBFF had displayed small errors as shown in chapter 3. This improvement results from increased cutoffs for the nonbonded interaction. While twin range cutoffs of 0.8 and 1.5nm were used in chapter 3, 1.5nm was used for both Coulomb cutoff and Lennard-Jones cutoff in this chapter. It allows more interactions to be taken into account. Also the total simulation time has been increased from 6 to 10 ns so that it had better chance to capture the average solution distribution from fluctuations over time. AMBER03 displayed very similar values, with slightly larger deviation from the experimental. At low ($x_2=0.1$ and 0.2) solute concentrations an overestimation in the self interactions (N_{22} and N_{11}) were observed suggesting an underestimation in the solvation of NMA (N_{21}). AT high ($x_2=0.5$) solute concentrations, the experimental data was relatively well reproduced by all but GROMOS45a3 which displayed a rather large degree of self-aggregation of NMA and water molecules even at

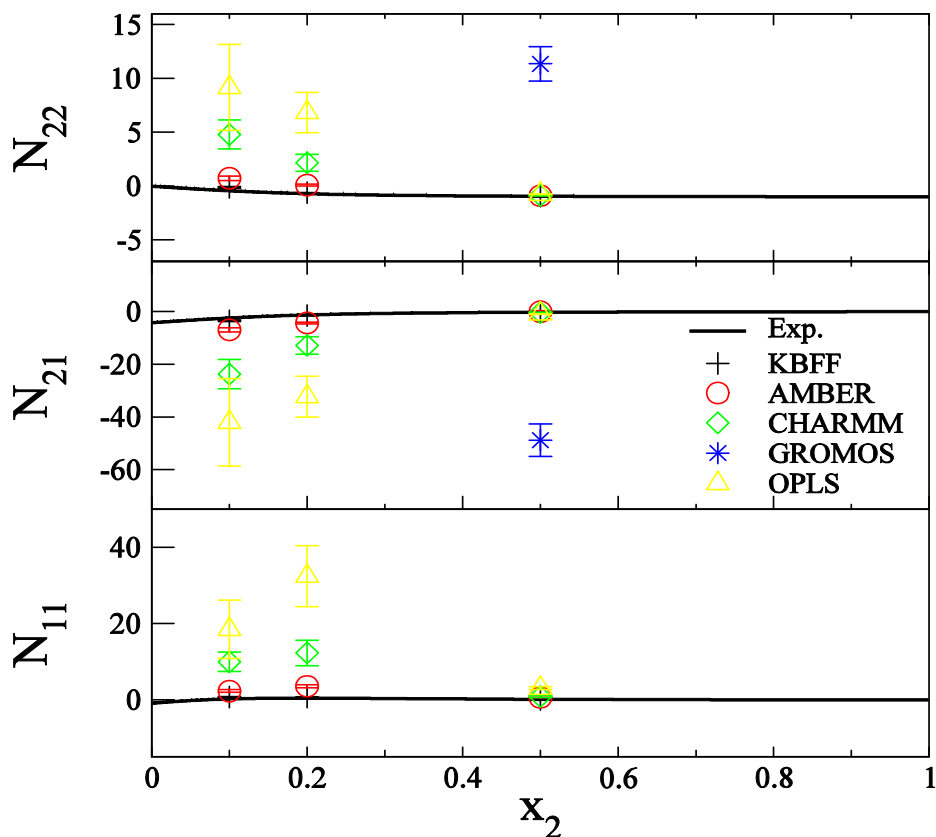


Figure 4.2 Excess coordination numbers (N_{ij}) as a function of composition for NMA (2) in water (1) solutions. The GROMOS values were not displayed for clarity: N_{22} 's at $x_2=0.1$ and 0.2 were 52 ± 5 and 47 ± 2 , respectively. N_{21} 's at $x_2=0.1$ and 0.2 were -209 ± 18 and 189 ± 9 , respectively. And N_{11} 's at $x_2=0.1$, 0.2 and 0.5 were 92 ± 8 , 188 ± 8 , and 193 ± 24 respectively.

these high solute concentrations. All force fields studied here shared the same trend: the error between the experimental and the simulated data became larger as the solute concentration got lower. The difference resulted from the excessive self aggregation. The excessive self aggregation increased in the order of KBFF, AMBER03, CHARMM22, OPLS, and GROMOS45a3. This is the same order as the first shell solute-solute interaction, as expected. A high degree of excessive self aggregation has been reported in other systems^{2,4,5,8,34}. From this

point of view, it is clear that a significant improvement is achieved by KBFF model over the agreement for the KB integrals provided by all the force fields studied here.

Figure 4.4 displays the density and partial molar volumes (pmv) as a function of composition for NMA and water mixtures. For all five force fields, the density of NMA and water mixtures was underestimated. It is consistent with the underestimation of the density of pure water for the selected water models. The density of pure NMA was overestimated with GROMOS45a3. As shown in our previous study⁸, the density from KBFF was underestimated in the whole composition range with the smallest error (<1%) among all five force field. The error was almost identical to the error in the density for pure water with SPC/E model. The partial molar volumes of NMA were reasonably well reproduced at high ($x_2=0.5$) solute concentrations, except for GROMOS45a3 whose density of pure NMA was overestimated. The partial molar

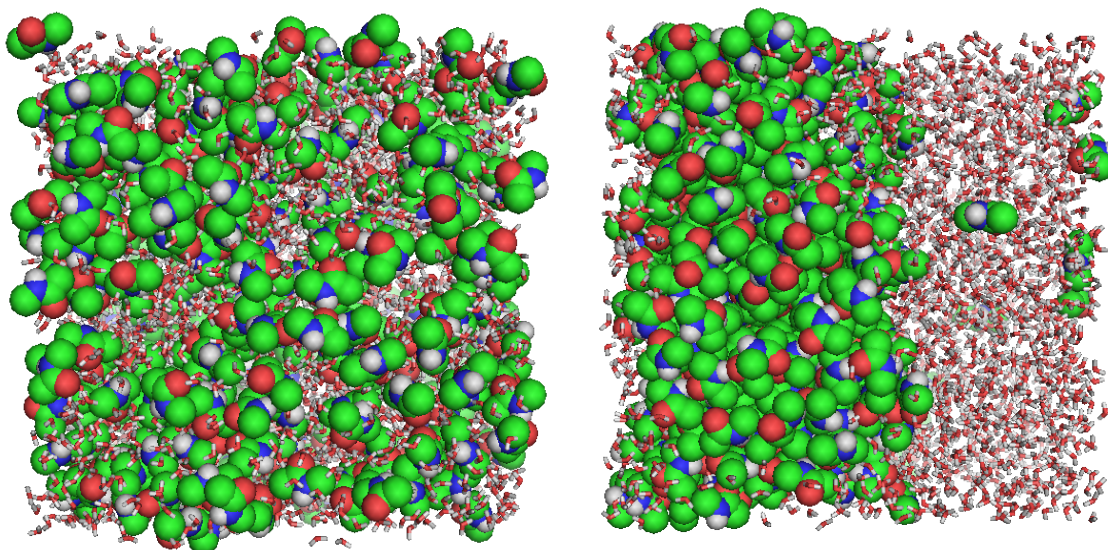


Figure 4.3 The snapshot of the simulated aqueous NMA solutions at $x_2=0.1$ with KBFF (left) and GROMOS45a3 (right). As the large positive N_{22} indicated in Figure 4.2, high degree of aggregation is observed in the system produced by GROMOS45a3.

volumes of water were relatively well reproduced with KBFF, but they were generally overestimated with other force fields.

The composition dependent enthalpy of mixing is displayed in Figure 4.5. According to the experimental data, at a slightly different temperature of 308K, the enthalpy of mixing between NMA and water is large and favorable³⁵. This was reproduced well by the KBFF model with small error. The other force fields displayed a low enthalpy of mixing suggesting that it may be the cause of the observed self aggregation. In particular, high self aggregation observed in

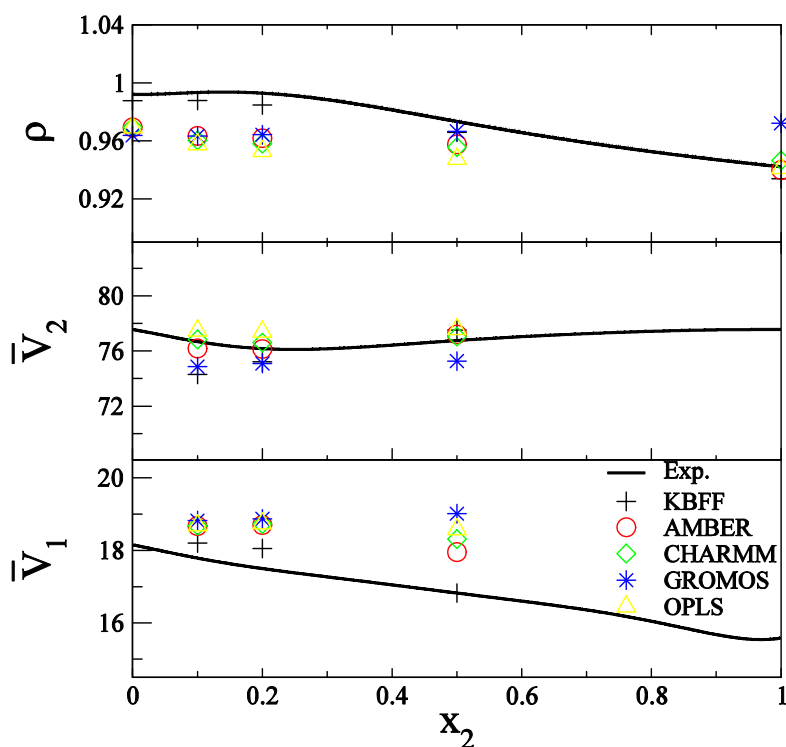


Figure 4.4 Density (g/cm³) and partial molar volumes (cm³/mol) as a function of composition for NMA(2) and water(1) solutions. Underestimation of density of pure water may be the cause of underestimation of density as well as the partial molar volume of water in the whole composition of mixture.

systems with GROMOS45a3 may be explained by the large and positive enthalpy of mixing.

A comparison was also performed for glycine and water mixtures. The simulated results for glycine and water mixtures are presented in Figures 4.6, 4.7, 4.8 and 4.9. Examples of the center of mass based rdfs obtained for mixtures with $m_2 = 1.5$ mol/kg are displayed in Figure 4.6. The rdfs indicate that g_{11} was similar for all five force fields, while the degree of first shell solute-solute interactions increased in the order of GROMOS45a3, KBFF, AMBER03, CHARMM22, and OPLS. The observed distance within which all liquid structure in g_{ij} could be included was dependent on the force field used. Hence, values between the different ranges were averaged to obtain the KB integrals from the simulated data for each force fields studied here:

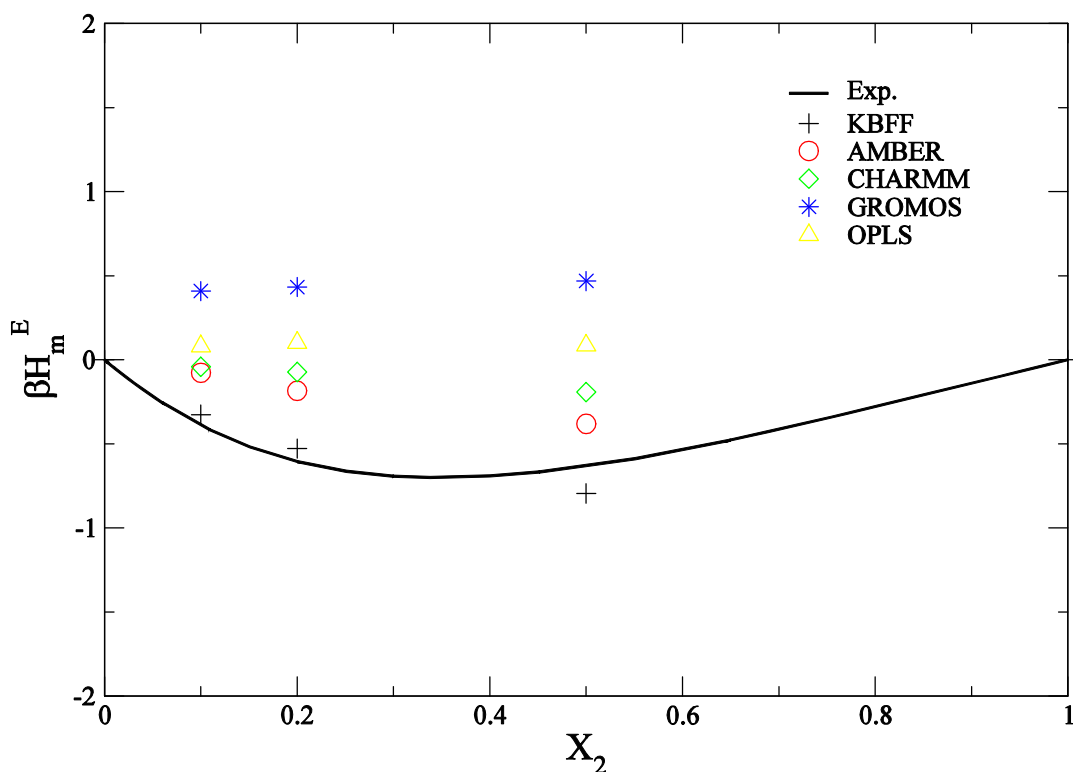


Figure 4.5 Enthalpy of mixing (H_m^E) for NMA solutions. The experimental data is from 308K and the simulations were performed at 313K. A large and positive enthalpy of mixing indicates high self aggregation.

1.2 and 1.6 nm for KBFF and GROMOS45a3, 1.5 and 1.9 for CHARMM22, 1.7 and 2.1 for OPLS, and 1.8 and 2.2 for AMBER03.

The excess coordination numbers N_{ij} obtained from the simulations of glycine and water mixtures are compared to the experimental data in Figure 4.7. The experimental data was relatively well reproduced by KBFF. An overestimation of the self interactions was observed in simulations with OPLS, suggesting self aggregation, as displayed in Figure 4.8. As a result an underestimation of solute-solvent interaction was observed in systems with this force field. Self

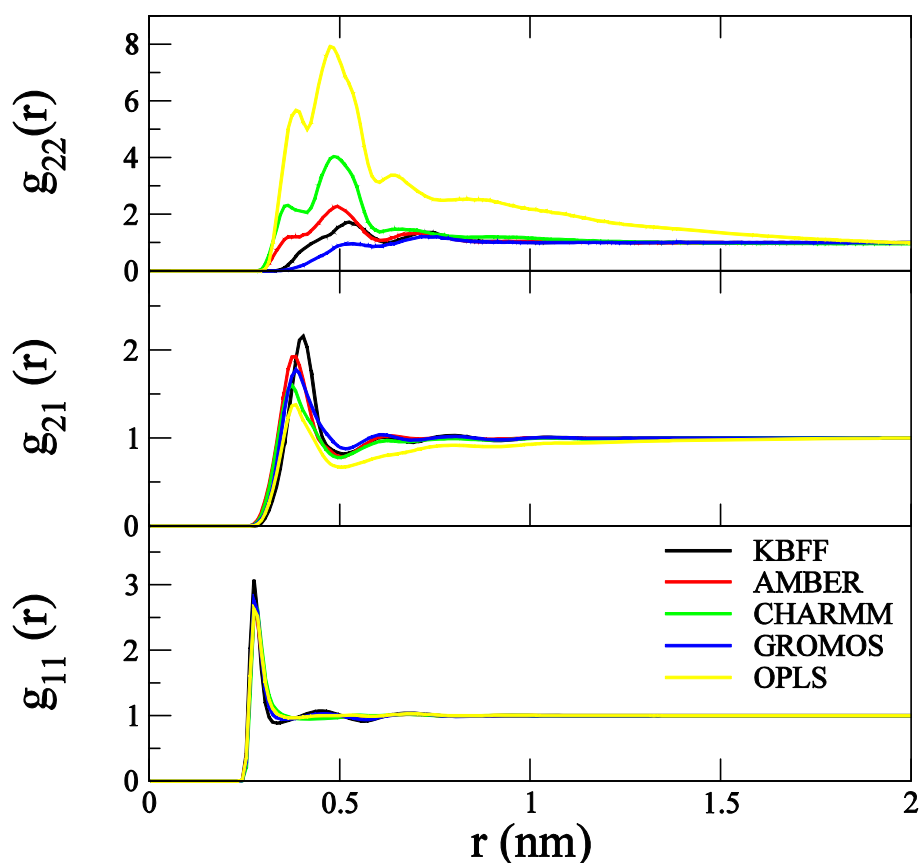


Figure 4.6 Center of mass based rdfs for $m_2=1.5\text{mol/kg}$ solutions of glycine (2) in water (1). The rdfs for water to water, g_{11} , are similar in all force field. The deviation between the force fields is larger in g_{22} .

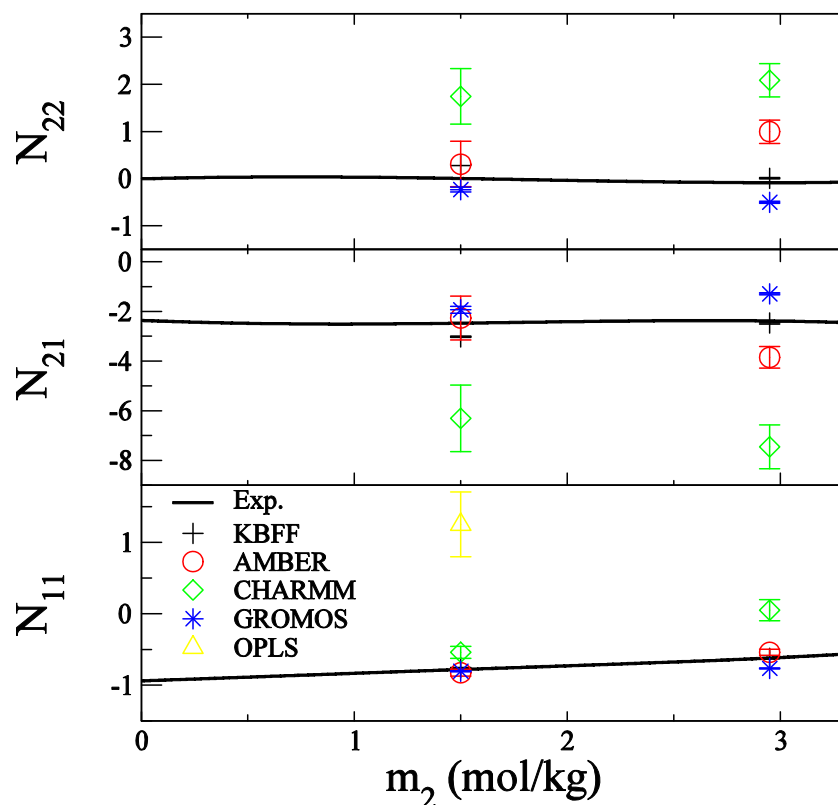


Figure 4.7 Excess coordination numbers (N_{ij}) as a function of composition for glycine (2) in water (1) solutions. The N_{ij} 's for OPLS are not displayed for clarity: N_{22} , N_{21} , and N_{11} were 14 ± 3 , -34 ± 6 , and 1 ± 1 at $m_2 = 1.5$ mol/kg, 25 ± 2 , -62 ± 5 , and 7 ± 1 at $m_2 = 2.95$ mol/kg, respectively.

interactions in solute-solute and solvent-solvent were slightly underestimated by GROMOS45a3. Figure 4.9 displays the composition dependent density and partial molar volumes for glycine and water mixtures. The density was slightly overestimated by KBFF and CHARMM22, and was slightly underestimated by OPLS and GROMOS45a3. The density estimation by AMBER03 was dependent on the solute concentration. But the error was small for all force fields. The errors observed for density were smaller than those observed in the KB integrals suggesting that KB integrals are a more sensitive test of the quality of a force field. The partial molar volume of

glycine was overestimated by GROMOS45a3 and was underestimated by the others. The partial molar volume of water was well reproduced by KBFF, but was overestimated with AMBER03, GROMOS45a3, and OPLS.

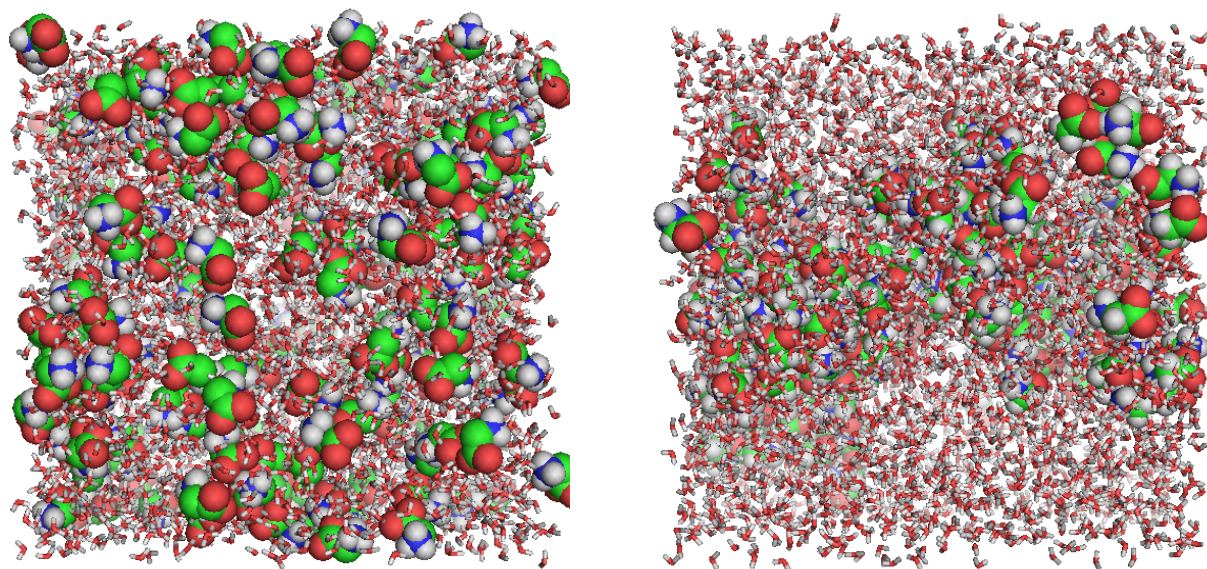


Figure 4.8 The snapshots of the simulated glycine/water systems at $m_2=2.95$ mol/kg with KBFF(left) and OPLS(right). Overestimation of self aggregation is observed in the system with OPLS.

4.4 Conclusion

A comparison between KBFF and other force fields for NMA and glycine was performed over a range of cosolvent concentrations. It was only KBFF that was acceptable for both. This is not surprising since KBFF has been parameterized to best reproduce the KB integrals in those specific systems, while others are for general peptide and proteins. But it is a noticeable advance that KBFF has achieved such a good agreement with the experimental data including the enthalpy of mixing, which is not used in its parameterization, *i.e.* tuning of the partial effective

charges on atoms. It proves again how useful and powerful the KB integrals are in parameterization of a force field. Also the correct KB integrals provided by KBFF indicate the correct molecular distribution of solution mixture, considering that KB theory is exact with no assumptions. It is demonstrated that KBFF can provide a reasonably accurate description of the balance between solvation and cosolvent self-aggregation in these systems over other force fields, suggesting KBFF is a promising force field to study intermolecular interactions in solution mixture.

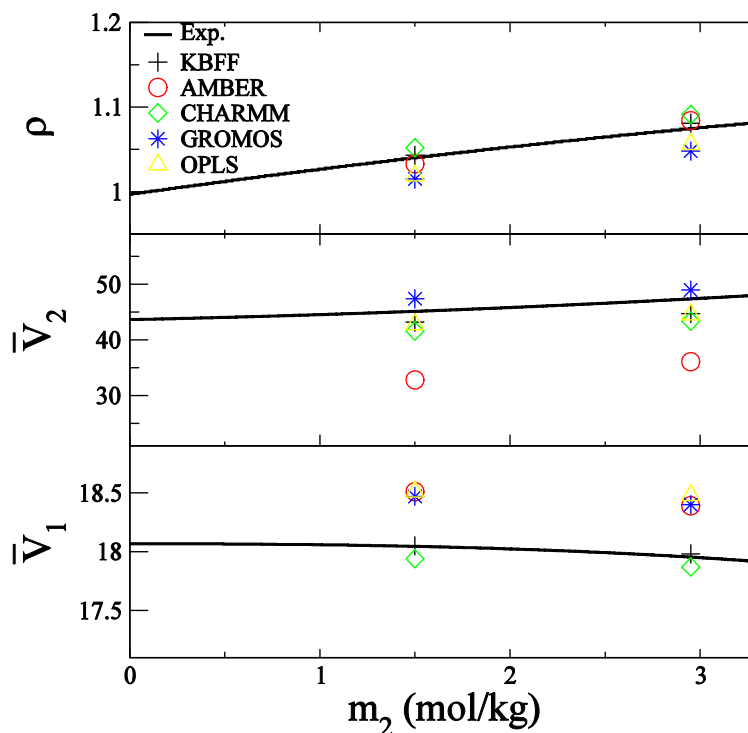


Figure 4.9 Density (g/cm^3) and partial molar volumes (cm^3/mol) as a function of composition for glycine (2) and water (1) solutions. The density was slightly overestimated by CHARMM, and was slightly underestimated by OPLS and GROMOS45a3. KBFF demonstrated the best agreement with the experimental data in partial molar volumes of glycine and water.

4.5 References

1. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 115, 5521-5530.
2. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 5901-5910.
3. Abui, M.; Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 7382-7388.
4. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2003, 107, 3891-3898.
5. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 119, 11342-11349.
6. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2004, 121, 2180-2186.
7. Weerasinghe, S.; Smith, P. E. *Journal of Physical Chemistry B* 2005, 109, 15080-15086.
8. Kang, M.; Smith, P. E. *Journal of Computational Chemistry* 2006, 27, 1477-1485.
9. Smith, P. E. *Biophysical Journal* 2006, 91, 849-856.
10. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 18716-18724.
11. Smith, P. E.; Pettitt, B. M. *Journal of the American Chemical Society* 1991, 113, 6029-6037.
12. Smith, P. E. *Journal of Physical Chemistry B* 2006, 110, 2862-2868.
13. Weerasinghe, S.; Smith, P. E. *Journal of Chemical Physics* 2003, 118, 10663-10670.
14. Kang, M.; Smith, P. E. *Journal of Chemical Physics* 2008, 128, 244511.
15. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 16271-16278.
16. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2002, 106, 1491-1500.
17. Smith, P. E. *Journal of Chemical Physics* 2008, 129, 124509.
18. Smith, P. E.; Mazo, R. A. *Journal of Physical Chemistry B* 2008, 112, 7875-7884.
19. Smith, P. E.; Marlow, G. E.; Pettitt, B. M. *Journal of the American Chemical Society* 1993, 115, 7493-7498.

20. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2001, 105, 11513-11522.
21. Kang, M.; Smith, P. E. *Fluid Phase Equilibria* 2007, 256, 14-19.
22. Chitra, R.; Smith, P. E. *Journal of Chemical Physics* 2001, 114, 426-435.
23. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochemistry and Biophysics* 2008, 50, 1-22.
24. Chen, F.; Smith, P. E. *Journal of Physical Chemistry B* 2008, 112, 8975-8984.
25. Ben Naim, A. *Molecular theory of solutions*, Oxford University Press: New York, 2006.
26. Kirkwood, J. G.; Buff, F. P. *Journal of Chemical Physics* 1951, 19, 774-777.
27. Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *Journal of Physical Chemistry* 1987, 91, 6269-6271.
28. Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *Journal of Chemical Physics* 1984, 81, 3684-3690.
29. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *Journal of Computational Physics* 1977, 23, 327-341.
30. Darden, T.; York, D.; Pedersen, L. *Journal of Chemical Physics* 1993, 98, 10089-10092.
31. The KBFF parameters for glycine are taken from the unpublished data developed by Smith, P. E. and Gee, M. B.
32. MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *Journal of Physical Chemistry B* 1998, 102, 3586-3616.

33. Sorin, E. J.; Pande, V. S. *Biophysical Journal* 2005, 88, 2472-2493.
34. Perera, A.; Sokolic, F. *Journal of Chemical Physics* 2004, 121, 11272-11282.
35. Zaichikov, A. M.; Golubinskii, O. E. *Zhurnal Fizicheskoi Khimii* 1996, 70, 1175-1179.

CHAPTER 5 - Kirkwood-Buff Theory of Four and Higher Component Mixtures[‡]

Abstract

Explicit expressions are developed for the chemical potential derivatives, partial molar volumes, and isothermal compressibility of solution mixtures involving four components at finite concentrations using the Kirkwood-Buff theory of solutions. In addition, a general recursion relationship is provided which can be used to generate the chemical potential derivatives for higher component solutions.

[‡] Reprinted with permission from "Kirkwood-Buff Theory of Four and Higher Component Mixtures" by Myungshim Kang and Paul E. Smith, 2008. *The Journal of Chemical Physics*, 128, 244511. Copyright 2008 by American Institute of Physics.

5.1 Introduction

Kirkwood–Buff (KB) theory is an exact theory of solutions that relates properties of a solution mixture to radial distribution functions (rdfs) between the different components of the solution.^{1,2} KB theory has been widely used to understand the basic properties of solutions,^{2–4} the effects of additives on the solubility of solutes (from small hydrocarbons to proteins)^{5–11} and biomolecular equilibria,^{12–16} to investigate the local composition of solutions in the context of preferential solvation,^{17,18} to study the effects of additives on the surface tension of liquids,^{19,20} to interpret computer simulation data,^{13,21,22} and to develop models for many of the above effects.²³

The central focus of KB theory are the KB integrals (G_{ij}) between the different species i and j in the solution mixture,¹

$$G_{ij} = G_{ji} = 4\pi \int_0^{\infty} [g_{ij}^{\mu VT}(r) - 1] r^2 dr, \quad (5.1)$$

where g_{ij} is the corresponding rdf and r is the intermolecular separation. The above rdfs are defined in a Grand Canonical (μVT) ensemble open to all species. Chemical potential derivatives for closed or semi-open systems in terms of the KB integrals and number densities ($\rho_i = n_i/V$) are then obtained after suitable thermodynamic transformations.^{2,24} The KB integrals, together with the corresponding excess coordination numbers, have provided a simple physical picture of changes in the local solution composition around each species.⁴

Unfortunately, as the number of solution components (n) increases, and/or one moves from open to semi open to closed ensembles, the resulting expressions become more cumbersome and involve significant algebraic manipulation.⁴ Expressions for two component solutions were provided in the original Kirkwood and Buff paper.¹ Subsequently, O’Connell²⁵ presented a general matrix formulation of KB theory, and Ben-Naim¹² developed a method to

simplify the matrices involved for a general n component mixture. O'Connell²⁵ also developed expressions based on the direct correlation function, as defined by the Ornstein-Zernike equation, instead of the total correlation function.¹⁶ This also makes KB theory highly compatible with integral equation theories. However, the physical interpretation of integrals involving the direct correlation function is more complicated than that of the standard KB integral at normal solution densities. Furthermore, the direct correlation function can only be obtained from computer simulations after Fourier transforming the original total correlation function.²⁶

Of course, one could always use the general matrix formulation of KB theory and solve numerically using values for the rdfs or KB integrals provided by some other approach (theory or simulation). However, this tends to obscure the contribution from the different KB integrals and hinder our understanding of specific effects. Therefore, it is often desirable to use explicit expressions that involve combinations of KB integrals and number densities. Explicit expressions for three component solutions have been provided by Ruckenstein and Shulgin (using algebraic software),²⁷ Ben-Naim,⁴ and Smith.²⁴ Ben-Naim¹² also developed expressions for some properties of four component systems, but where several of the components appear at infinite dilution. To our knowledge, explicit expressions for chemical potential derivatives in four or higher component systems have not appeared for the case where all components are present at finite concentrations. Here, we use some of the relationships provided previously by Hall in an alternative derivation of KB theory,²⁸ to generate expressions for four component solutions. A general recursion relationship is then developed for higher component mixtures.

5.2 Theory

5.2.1 General Approach

Hall re-derived KB theory using a different approach to Kirkwood and Buff.²⁸ In doing so Hall produced two primary equations from which many of the expressions required here can be generated. However, his approach was still somewhat involved. Here we present a simpler derivation of the Hall equations. The first focuses on changes in the molar concentration of any component. If we consider the species number densities (or molarities) in the grand canonical ensemble to be functions of T and all the chemical potentials (μ) then we can write

$$d \ln \rho_i = \sum_{j=1}^n \left(\frac{\partial \ln \rho_i}{\partial \mu_j} \right)_{T, \mu_{k \neq j}} d\mu_j \quad \text{constant } T, \quad (5.2)$$

for any component i at constant T . Here, the summation is over all n components of the solution. We note that all the chemical potentials are independent thermodynamic variables in this open ensemble. The above derivatives can be expressed in terms of KB integrals using the fact that^{1,2}

$$\left(\frac{\partial \ln \rho_i}{\partial \mu_j} \right)_{T, \mu_{k \neq j}} = \beta(\delta_{ij} + N_{ij}), \quad (5.3)$$

which is essentially the starting equation for KB theory. Here, δ_{ij} is the Kroenecker delta function, $N_{ij} = \rho_j G_{ij} \neq N_{ji}$, $\beta = 1/RT$, and R is the Gas constant. From the these two equations one finds

$$d \ln \rho_i = \beta \sum_{j=1}^n (\delta_{ij} + N_{ij}) d\mu_j \quad \text{constant } T. \quad (5.4)$$

The above expression is valid for changes in the concentration of any component in any multicomponent system and any (thermodynamically reasonable) ensemble with T constant. This is the equation derived by Hall but using a much longer route.²⁸ If one is interested in expressing

solution compositions in terms of molalities ($m_i = \rho_i/\rho_1$ to within a conversion factor) then using the fact that $d \ln m_i = d \ln \rho_i - d \ln \rho_1$ one can write

$$d \ln m_i = \beta \sum_{j=1}^n (\delta_{ij} + N_{ij} - \delta_{1j} - N_{1j}) d\mu_j \quad \text{constant T,} \quad (5.5)$$

which is also valid for any constant T ensemble. Clearly, in doing so we have labeled component 1 as the primary solvent and therefore component 1 is unique – as it is also experimentally. The consequences of doing this will be discussed later.

In the traditional derivation of KB theory the set of equations presented in Equation 5.4 are converted into matrix form after taking derivatives with respect to $\ln \rho_j$ with all $\rho_{k \neq j}$ held constant.⁴ They can then be solved to obtain a series of expressions involving the quantities

$$\beta \left(\frac{\partial \mu_i}{\partial \ln N_j} \right)_{T, V, N_{k \neq j}} = \beta \left(\frac{\partial \mu_i}{\partial \ln \rho_j} \right)_{T, N_{k \neq j}} . \quad (5.6)$$

These constant volume derivatives then have to be transformed using a series of thermodynamic relationships into the required and experimentally relevant derivatives at constant P as defined by

$$\mu_{ij} = \beta \left(\frac{\partial \mu_i}{\partial \ln N_j} \right)_{T, P, N_{k \neq j}} = \beta \left(\frac{\partial \mu_i}{\partial \ln m_j} \right)_{T, P, m_{k \neq j}} . \quad (5.7)$$

This is clearly the most general approach. However, it has long been recognized that the expressions obtained for higher multicomponent systems ($n \geq 3$) involve considerable algebraic manipulation.^{4,27} In addition, a significant degree of cancellation of terms in the expressions is often found but not easily recognized in the matrix formulation.

Here, we will adopt a different route which we believe is much simpler for mixtures with a large number of components. Eliminating $d\mu_1$ from Equation 5.4 using the corresponding Gibbs-Duhem (GD) relationship at constant T and P , $\sum_{j=1}^n \rho_j d\mu_j = 0$, provides

$$d \ln \rho_i = \beta \sum_{j=2}^n [\delta_{ij} + N_{ij} - m_j (\delta_{i1} + N_{i1})] d\mu_j \quad \text{constant T, P.} \quad (5.8)$$

This can be used to obtain an expression for changes in the molalities,

$$d \ln m_i = \beta \sum_{j=2}^n (\delta_{ij} + N_{ij}^+) d\mu_j \quad \text{constant T, P,} \quad (5.9)$$

where $N_{ij}^+ = N_{ij} + m_j (1 + N_{11} - N_{i1} - N_{j1})$ and $i = 2, n$. The additional constraint of constant P arises from our use of the corresponding GD expression. This is the equation provided by Hall²⁸ for changes in molal concentrations at constant T and P . It also appears in the original KB paper without derivation.¹ The above set of equations can be written in a general $(n-1) \times (n-1)$ matrix form for a mixture of n components so that

$$\begin{bmatrix} 1 + N_{22}^+ & N_{23}^+ & N_{24}^+ & \cdots & N_{2n}^+ \\ N_{32}^+ & 1 + N_{33}^+ & & & \\ N_{42}^+ & & 1 + N_{44}^+ & & \\ \vdots & & & \ddots & \\ N_{n2}^+ & & & & 1 + N_{nn}^+ \end{bmatrix} \begin{bmatrix} \beta d\mu_2 \\ \beta d\mu_3 \\ \beta d\mu_4 \\ \vdots \\ \beta d\mu_n \end{bmatrix} = \begin{bmatrix} d \ln m_2 \\ d \ln m_3 \\ d \ln m_4 \\ \vdots \\ d \ln m_n \end{bmatrix}. \quad (5.10)$$

To continue we will chose our required ensemble and then take derivatives with respect to the molality of one component, in this case $\ln m_j$, keeping T , P , and all other $m_{k \neq j}$ constant. This makes the resulting expressions less general than previous approaches, but one can easily recover derivatives with respect to other species by a simple index change. One finds that

$$\begin{bmatrix} 1+N_{22}^+ & N_{23}^+ & N_{24}^+ & \cdots & N_{2n}^+ \\ N_{32}^+ & 1+N_{33}^+ & & & \\ N_{42}^+ & & 1+N_{44}^+ & & \\ \vdots & & & \ddots & \\ N_{n2}^+ & & & & 1+N_{nn}^+ \end{bmatrix} \begin{bmatrix} \mu_{2j} \\ \mu_{3j} \\ \mu_{4j} \\ \vdots \\ \mu_{nj} \end{bmatrix} = \begin{bmatrix} \delta_{2j} \\ \delta_{3j} \\ \delta_{4j} \\ \vdots \\ \delta_{nj} \end{bmatrix} \quad \text{or} \quad \mathbf{D}_n \boldsymbol{\mu} = \mathbf{d}, \quad (5.11)$$

where the matrix elements are given by $D_{\alpha\beta} = \delta_{\alpha\beta} + N_{(\alpha+1)(\beta+1)}^+$, the vector elements of μ_α are given by $\mu_{(\alpha+1)j}$, and the vector elements of d_α are given by $\delta_{(\alpha+1)j}$ with $\alpha, \beta = 1, n-1$. Hence, we have a set of simultaneous equations which can be solved to give the chemical potential derivatives. Therefore,

$$\boldsymbol{\mu} = \mathbf{D}_n^{-1} \mathbf{d}. \quad (5.12)$$

One can express the inverse in terms of cofactors of the original D_n matrix. The chemical potential derivatives are then given by $\mu_{ij} = D_n^{j-1, i-1} / |D_n|$, for $i, j \neq 1$ and where $D_n^{j-1, i-1}$ is a cofactor of D_n . If the chemical potential derivative of species one is required it can be obtained from the GD equations

$$\mu_{1j} = -\sum_{k=2}^n m_k \mu_{kj} = -m_j \sum_{k=2}^n \mu_{jk}, \quad (5.13)$$

using the solutions to Equation 5.11. In the above expression $j = 1, n$ and $m_1 = 1$.

There are several advantages of this approach. First, it can be applied directly to any number of solution components in any constant T and P ensemble. Therefore, we do not have to transform the subsequent expressions from a constant T and V to a constant T and P ensemble. Second, we have eliminated the chemical potential of species one and therefore the resulting set of equations and the corresponding matrix is reduced. Third, the simplicity of the column vector on the right-hand side of Equation 5.12 indicates that each chemical potential derivative expression involves only one element of the inverse matrix in the numerator, together with the

determinant of D_n in the denominator. This set of combined factors greatly simplifies the resulting expressions.

In the previous sections we have focused on derivatives with respect to molality. Derivatives with respect to mole fraction or molarity can be obtained by noting that

$$d \ln x_i = d \ln \rho_i - \sum_{j=1}^n x_j d \ln \rho_j \quad \text{constant } T \quad (5.14)$$

and

$$d \ln m_i = d \ln \rho_i - d \ln \rho_1 \quad \text{constant } T \quad (5.15)$$

for any number of components at constant T . One could develop Equation 5.14 in terms of the KB integrals. However, it is much easier, especially for closed systems, to convert the molality based derivatives to mole fraction derivatives after the former have been obtained. For closed systems these equations become

$$\left(\frac{\partial \ln x_i}{\partial \ln m_j} \right)_{T,P,m_{k \neq j}} = \delta_{ij} - x_j, \quad (5.16)$$

with,

$$\left(\frac{\partial \ln \rho_i}{\partial \ln m_j} \right)_{T,P,m_{k \neq j}} = \delta_{ij} - \rho_j \bar{V}_j, \quad (5.17)$$

and therefore,

$$\left(\frac{\partial \ln \rho_i}{\partial \ln x_j} \right)_{T,P,m_{k \neq j}} = \frac{\delta_{ij} - \rho_j \bar{V}_j}{1 - x_j}, \quad (5.18)$$

for any number of components at constant T and P .

To determine the corresponding expressions for the partial molar volumes (pmvs) in multicomponent systems it is sufficient to express the pmvs in terms of the chemical potential

derivatives. We will continue to treat species 1 as a unique component. Starting with Equation 5.4 for the differential of the number density of species 1 and eliminating $d\mu_1$ by use of the GD relationship one finds

$$d \ln \rho_1 = -\beta \sum_{j=2}^n m_j (1 + N_{11} - N_{j1}) d\mu_j \quad \text{constant } T, P. \quad (5.19)$$

Obviously, there are a series of similar expressions depending on the initial choice of i in Equation 5.4. Taking the derivative with respect to $\ln m_k$ while keeping T, P , and all other $m_{j \neq k}$ constant provides for $k \neq 1$,

$$\rho_k \bar{V}_k = \sum_{j=2}^n m_j (1 + N_{11} - N_{j1}) \mu_{jk} = m_k \sum_{j=2}^n (1 + N_{11} - N_{j1}) \mu_{kj}, \quad (5.20)$$

where the appropriate chemical potential derivatives are provided by Equation 5.12. If required, the pmv of species 1 can be obtained from the fact that,

$$\sum_{j=1}^n \rho_j \bar{V}_j = 1, \quad (5.21)$$

for all mixtures.

Finally, if one starts from Equation 5.4 and then takes derivatives with respect to pressure with all m_j and T constant one can show that for any component i ,

$$RT\kappa_T = \sum_{j=1}^n (\delta_{ij} + N_{ij}) \bar{V}_j = \bar{V}_i + \sum_{j=1}^n N_{ij} \bar{V}_j, \quad (5.22)$$

where κ_T is the isothermal compressibility. This can be used to derive an expression for the compressibility. If we chose $i = 1$ and then eliminate the pmv of species 1 by using Equation 5.21 then

$$\rho_1 RT\kappa_T = (1 + N_{11}) - \sum_{j=2}^n (1 + N_{11} - N_{j1}) \rho_j \bar{V}_j, \quad (5.23)$$

which can be written as

$$\rho_1 RT \kappa_T = (1 + N_{11}) - \sum_{j=2}^n (1 + N_{11} - N_{j1}) \sum_{k=2}^n m_k (1 + N_{11} - N_{k1}) \mu_{kj}, \quad (5.24)$$

using the corresponding chemical potential derivatives.

Before leaving this section we note that the pmv of a species can be considered to involve two contributions.² The first relates to the change in volume of the solution due to the volume occupied by the additional molecule. The second involves the ideal contribution to the pmv which arises due to the fact that the additional molecule will possess a momentum, corresponding to the particular temperature, which contributes to the pressure of the system. Under constant P conditions this gives rise to a change in volume according to the compressibility of the solution. Therefore, one can isolate the former change by writing

$$V_i = \bar{V}_i - RT \kappa_T. \quad (5.25)$$

From Equation 5.22 we obtain the relationship

$$\rho_i V_i = -N_{1i} - \sum_{j=2}^n (N_{ji} - N_{1i}) \sum_{k=2}^n m_k (1 + N_{11} - N_{k1}) \mu_{kj}, \quad (5.26)$$

which is now a better measure of the actual volume occupied by each species in solution. Another interesting property of solutions is the pseudo chemical potential (μ^*). The pseudo chemical potential plays an important role in solution theory and is defined by the equation²

$$\mu_i = \mu_i^* + RT \ln \Lambda_i^3 \rho_i, \quad (5.27)$$

where Λ is the thermal de Broglie wavelength. From this equation it is quite easy to show that

$$\beta \left(\frac{\partial \mu_i^*}{\partial \ln m_j} \right)_{T, P, m_{k \neq j}} = \mu_{ij} - (\delta_{ij} - \rho_j \bar{V}_j), \quad (5.28)$$

and also

$$\beta \left(\frac{\partial \mu_i^*}{\partial \ln \rho_j} \right)_{T,P,m_{k \neq j}} = \beta \left(\frac{\partial \mu_i}{\partial \ln \rho_j} \right)_{T,P,m_{k \neq j}} - 1 \quad (5.29)$$

which completes our preliminary analysis.

Hence, we have a general set of equations (5.12, 5.20, and 5.24) which can be used to derive the KB expressions for the chemical potential derivatives, pmvs, and compressibility of any multicomponent mixture.

5.2.2 Four Component Mixtures

As an example of the current approach we will generate the expressions for a four component system where all components appear at finite concentrations. To our knowledge the explicit KB expressions for a four component system have not been presented in the literature.

Using the above approach we find

$$\begin{bmatrix} \mu_{22} \\ \mu_{32} \\ \mu_{42} \end{bmatrix} = D_4^{-1} \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad (5.30)$$

and where the inverse of D_4 is given by

$$D_4^{-1} = \frac{1}{|D_4|} \begin{bmatrix} (1 + N_{33}^+)(1 + N_{44}^+) - N_{43}^+ N_{34}^+ & N_{24}^+ N_{43}^+ - N_{23}^+(1 + N_{44}^+) & N_{23}^+ N_{34}^+ - N_{24}^+(1 + N_{33}^+) \\ N_{34}^+ N_{42}^+ - N_{32}^+(1 + N_{44}^+) & (1 + N_{22}^+)(1 + N_{44}^+) - N_{42}^+ N_{24}^+ & N_{24}^+ N_{23}^+ - N_{34}^+(1 + N_{22}^+) \\ N_{32}^+ N_{43}^+ - N_{42}^+(1 + N_{33}^+) & N_{23}^+ N_{42}^+ - N_{43}^+(1 + N_{22}^+) & (1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+ \end{bmatrix}. \quad (5.31)$$

Therefore, the expressions for the chemical potential derivatives in a four component mixture are given by

$$\beta \left(\frac{\partial \mu_2}{\partial \ln m_2} \right)_{T,P,m_3,m_4} = \frac{(1 + N_{33}^+)(1 + N_{44}^+) - N_{43}^+ N_{34}^+}{|D_4|},$$

$$\beta \left(\frac{\partial \mu_3}{\partial \ln m_2} \right)_{T,P,m_3,m_4} = - \frac{(1 + N_{44}^+) N_{32}^+ - N_{34}^+ N_{42}^+}{|D_4|},$$

$$\beta \left(\frac{\partial \mu_4}{\partial \ln m_2} \right)_{T,P,m_3,m_4} = - \frac{(1 + N_{33}^+) N_{42}^+ - N_{43}^+ N_{32}^+}{|D_4|}, \quad (5.32)$$

with

$$|D_4| = (1 + N_{22}^+)(1 + N_{33}^+)(1 + N_{44}^+) - (1 + N_{22}^+) N_{34}^+ N_{43}^+ - (1 + N_{33}^+) N_{24}^+ N_{42}^+ - (1 + N_{44}^+) N_{23}^+ N_{32}^+ + 2 N_{23}^+ N_{34}^+ N_{42}^+, \quad (5.33)$$

and where we have also used the fact that $\rho_i N_{ij}^+ = \rho_j N_{ji}^+$ to simplify the above determinant. The final derivatives (μ_{12} and μ_{11}) can be obtained after application of the GD relationships

$$\beta \left(\frac{\partial \mu_1}{\partial \ln m_2} \right)_{T,P,m_3,m_4} = -m_2 \mu_{22} - m_3 \mu_{32} - m_4 \mu_{42} \quad (5.34)$$

and,

$$\beta \left(\frac{\partial \mu_1}{\partial \ln N_1} \right)_{T,P,N_2,N_3,N_4} = -\mu_{12} - \mu_{13} - \mu_{14}. \quad (5.35)$$

Derivatives with respect to other species molalities (m_3 and m_4) are obtained quite easily by either inspection, by noting that $m_j \mu_{jk} = m_k \mu_{kj}$, or from the fact that,

$$\begin{bmatrix} \mu_{23} \\ \mu_{33} \\ \mu_{43} \end{bmatrix} = D_4^{-1} \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mu_{24} \\ \mu_{34} \\ \mu_{44} \end{bmatrix} = D_4^{-1} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}, \quad (5.36)$$

which are also relatively simple to solve.

Application of Equation 5.20 and the expressions found above provides the following expressions for the pmvs in four component mixtures,

$$\rho_2 \bar{V}_2 = m_2 (1 + N_{11} - N_{21}) \mu_{22} + m_3 (1 + N_{11} - N_{31}) \mu_{32} + m_4 (1 + N_{11} - N_{41}) \mu_{42},$$

$$\begin{aligned}\rho_3 \overline{V}_3 &= m_2(1 + N_{11} - N_{21})\mu_{23} + m_3(1 + N_{11} - N_{31})\mu_{33} + m_4(1 + N_{11} - N_{41})\mu_{43}, \\ \rho_4 \overline{V}_4 &= m_2(1 + N_{11} - N_{21})\mu_{24} + m_3(1 + N_{11} - N_{31})\mu_{34} + m_4(1 + N_{11} - N_{41})\mu_{44}.\end{aligned}\quad (5.37)$$

The above expressions obey Equation 5.21 as required. If necessary, the pmv of 1 can then be obtained using Equation 5.21. Finally, for the isothermal compressibility we obtain

$$\begin{aligned}\rho_1 RT\kappa_T &= 1 + N_{11} \\ &\quad - (1 + N_{11} - N_{21})[(1 + N_{11} - N_{21})m_2\mu_{22} + (1 + N_{11} - N_{31})m_3\mu_{32} + (1 + N_{11} - N_{41})m_4\mu_{42}] \\ &\quad - (1 + N_{11} - N_{31})[(1 + N_{11} - N_{21})m_2\mu_{23} + (1 + N_{11} - N_{31})m_3\mu_{33} + (1 + N_{11} - N_{41})m_4\mu_{43}] \\ &\quad - (1 + N_{11} - N_{41})[(1 + N_{11} - N_{21})m_2\mu_{24} + (1 + N_{11} - N_{31})m_3\mu_{34} + (1 + N_{11} - N_{41})m_4\mu_{44}],\end{aligned}\quad (5.38)$$

which can be simplified to provide

$$\begin{aligned}\rho_1 RT\kappa_T &= 1 + N_{11} - (1 + N_{11} - N_{21})^2 m_2\mu_{22} - (1 + N_{11} - N_{31})^2 m_3\mu_{33} - (1 + N_{11} - N_{41})^2 m_4\mu_{44} \\ &\quad - 2(1 + N_{11} - N_{21})(1 + N_{11} - N_{31})m_3\mu_{32} \\ &\quad - 2(1 + N_{11} - N_{21})(1 + N_{11} - N_{41})m_4\mu_{42} \\ &\quad - 2(1 + N_{11} - N_{41})(1 + N_{11} - N_{31})m_3\mu_{34}.\end{aligned}\quad (5.39)$$

5.2.3 Three Component Mixtures

While three component systems have been studied before,^{4,27} it is interesting and informative to compare the limiting expressions provided here with those currently in the literature, especially due to the different notations involved. In addition, this will aid in the development of a general recursive relationship for the derivatives. The limiting forms are quite easy to obtain as we have $N_{ij}^+ \rightarrow 0$ as $\rho_j \rightarrow 0$, and $m_i \mu_{ij} = m_j \mu_{ji} \rightarrow 0$ as $\rho_i \rightarrow 0$ or $\rho_j \rightarrow 0$. Therefore, as m_4 tends to zero one obtains the derivatives for a ternary system of 1, 2 and 3. The chemical potential derivatives are then given by

$$\beta \left(\frac{\partial \mu_2}{\partial \ln m_2} \right)_{T,P,m_3} = \frac{(1 + N_{33}^+)}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+},$$

$$\beta \left(\frac{\partial \mu_3}{\partial \ln m_2} \right)_{T,P,m_3} = - \frac{N_{32}^+}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+}. \quad (5.40)$$

The corresponding pmv expressions reduce to

$$\rho_2 \bar{V}_2 = m_2 \frac{(1 + N_{11} - N_{21})(1 + N_{33}^+) - (1 + N_{11} - N_{31})N_{23}^+}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+},$$

$$\rho_3 \bar{V}_3 = m_3 \frac{(1 + N_{11} - N_{31})(1 + N_{22}^+) - (1 + N_{11} - N_{21})N_{32}^+}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+}, \quad (5.41)$$

with a compressibility given by

$$\rho_1 RT \kappa_T = 1 + N_{11} - m_2 \frac{(1 + N_{11} - N_{21})^2 (1 + N_{33}^+)}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+}$$

$$+ 2m_2 \frac{(1 + N_{11} - N_{21})(1 + N_{11} - N_{31})N_{23}^+}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+} - m_3 \frac{(1 + N_{11} - N_{31})^2 (1 + N_{22}^+)}{(1 + N_{22}^+)(1 + N_{33}^+) - N_{23}^+ N_{32}^+}. \quad (5.42)$$

5.2.4 Two Component Mixtures

Two component mixtures have clearly been studied before, but not using the present notation. After taking an additional $\rho_3 \rightarrow 0$ limit one finds that for the two component case,

$$\beta \left(\frac{\partial \mu_2}{\partial \ln m_2} \right)_{T,P,N_1} = \frac{1}{1 + N_{22}^+} \quad (5.43)$$

and

$$\rho_2 \bar{V}_2 = m_2 \frac{1 + N_{11} - N_{21}}{1 + N_{22}^+}, \quad (5.44)$$

with

$$\rho_1 RT\kappa_T = 1 + N_{11} - m_2 \frac{(1 + N_{11} - N_{21})^2}{1 + N_{22}^+}, \quad (5.45)$$

which further reduces to the required compressibility equation when $\rho_2 \rightarrow 0$.

5.2.5 Open and Semi-Open Systems

Traditionally, KB expressions for open and semi-open systems have been derived starting from the fully closed system results.^{29, 30} This can be quite tedious. Recently, we suggested starting from expressions for the fully open system and transforming to the required ensemble in a stepwise manner.^{24,31} This made the manipulations easier although several steps were still required. However, it is clear from Equations 5.4 and 5.5 that results for open semi-open systems become almost trivial. As an example we will derive an expression for the preferential binding parameter ($\partial m_3/\partial m_2$) for ternary mixtures in the T, μ_1, μ_3 ensemble, where 1 is the primary solvent, 2 is the biomolecule of interest and 3 is an additive. Starting from Equation 5.5 one immediately finds

$$\left(\frac{\partial \ln m_2}{\partial \mu_2} \right)_{T, \mu_1, \mu_3} = \beta(1 + N_{22} - N_{12}) \quad (5.46)$$

and

$$\left(\frac{\partial \ln m_3}{\partial \mu_2} \right)_{T, \mu_1, \mu_3} = \beta(N_{32} - N_{12}), \quad (5.47)$$

which can be solved to yield

$$\left(\frac{\partial m_3}{\partial m_2} \right)_{T, \mu_1, \mu_3} = \frac{N_{23} - m_3 N_{21}}{1 + N_{22} - N_{12}}, \quad (5.48)$$

and is in agreement with previous results.²⁴ It is clear from Equation 5.5 that the same expression is obtained if we have any number of additional components at a constant chemical potential.

Alternatively, one can start from Equation 5.9 to obtain an expression for the equivalent property in the T, P, μ_3 ensemble. Hence,

$$(1 + N_{22}^+) \beta \left(\frac{\partial \mu_2}{\partial \ln m_2} \right)_{T, P, \mu_3} = 1 \quad (5.49)$$

and

$$N_{32}^+ \beta \left(\frac{\partial \mu_2}{\partial \ln m_2} \right)_{T, P, \mu_3} = \left(\frac{\partial \ln m_3}{\partial \ln m_2} \right)_{T, P, \mu_3}, \quad (5.50)$$

which can be solved quite easily to give

$$\left(\frac{\partial m_3}{\partial m_2} \right)_{T, P, \mu_3} = \frac{N_{23}^+}{1 + N_{22}^+}, \quad (5.51)$$

and is also in agreement with previous results.²⁴ Again, the same expression is valid in the presence of any number of additional species as long as their chemical potentials remain constant.

5.2.6 A General Recursion Relationship for the Chemical Potential Derivatives

Analysis of the chemical potential derivatives for two, three, and four component systems enables a general recursion relationship to be established. It is clear that the denominator will always contain the determinant $|D_n|$ for a general n component system. If we focus on the expressions in the numerator, one immediately observes that the numerator of μ_{ii} is just the determinant of the D matrix for the corresponding $n-1$ system in which component i has been eliminated. Hence, we have

$$\mu_{nn}^n = \frac{|D_{n-1}|}{|D_n|}, \quad (5.52)$$

where the superscript indicates a derivative defined in an n component system. A simple change in indices provide equivalent expressions for any μ_{ii}^n where $i \neq 1$. The numerators of the other derivatives (μ_{nj} , $j \neq 1$ or n) also follow a simple pattern. The chemical potential derivatives for the n th component obey the recursive relationship

$$\mu_{nj}^n = -\frac{|D_{n-1}|}{|D_n|} \sum_{k=2}^{n-1} \mu_{kj}^{n-1} N_{nk}^+ = -\mu_{nn}^n \sum_{k=2}^{n-1} \mu_{kj}^{n-1} N_{nk}^+, \quad (5.53)$$

which is just a factorization of the D_n matrix that one observes due to the relative simplicity of Equation 5.12. Expressions for the μ_{ij} derivatives where $i \neq j \neq 1$ can then be found by inspection.

5.2.7 Five Component Systems

Using the recursion relationship developed above one can generate expressions for the chemical potential derivatives in five component solutions. For simplicity, we will only consider the chemical potential derivatives. The following expression is obtained from Equation 5.52 and 5.32 followed by a simple index change ($5 \leftrightarrow 2$) in the numerator,

$$\begin{aligned} \mu_{22}|D_5| &= (1 + N_{33}^+)(1 + N_{44}^+)(1 + N_{55}^+) \\ &\quad - (1 + N_{33}^+)N_{45}^+N_{54}^+ - (1 + N_{44}^+)N_{35}^+N_{53}^+ - (1 + N_{55}^+)N_{34}^+N_{43}^+ \\ &\quad + 2N_{34}^+N_{45}^+N_{53}^+. \end{aligned} \quad (5.54)$$

Using Equation 5.53 and the set of derivatives for a four component solution provided in Equation 5.32 one finds

$$\begin{aligned}
\mu_{52}|D_5| &= -(1 + N_{33}^+)(1 + N_{44}^+)N_{52}^+ \\
&+ (1 + N_{33}^+)N_{42}^+N_{54}^+ + (1 + N_{44}^+)N_{32}^+N_{53}^+ \\
&- N_{32}^+N_{43}^+N_{54}^+ - N_{34}^+N_{42}^+N_{53}^+ + N_{34}^+N_{43}^+N_{52}^+.
\end{aligned} \tag{5.55}$$

Consequently, using a simple index change of $5 \leftrightarrow 3$ and $5 \leftrightarrow 4$ one finds

$$\begin{aligned}
\mu_{32}|D_5| &= -(1 + N_{44}^+)(1 + N_{55}^+)N_{32}^+ \\
&+ (1 + N_{44}^+)N_{35}^+N_{52}^+ + (1 + N_{55}^+)N_{42}^+N_{34}^+ \\
&- N_{42}^+N_{54}^+N_{35}^+ - N_{45}^+N_{52}^+N_{34}^+ + N_{45}^+N_{54}^+N_{32}^+
\end{aligned} \tag{5.56}$$

and

$$\begin{aligned}
\mu_{42}|D_5| &= -(1 + N_{33}^+)(1 + N_{55}^+)N_{42}^+ \\
&+ (1 + N_{33}^+)N_{45}^+N_{52}^+ + (1 + N_{55}^+)N_{32}^+N_{43}^+ \\
&- N_{32}^+N_{53}^+N_{45}^+ - N_{35}^+N_{52}^+N_{43}^+ + N_{35}^+N_{53}^+N_{42}^+,
\end{aligned} \tag{5.57}$$

respectively. Derivatives with respect to other species can be obtained by inspection. The above expressions are in agreement with those obtained directly via Equation 5.12. If required, the corresponding pmvs and compressibility can be obtained from Equations 5.20 and 5.24.

5.3 Discussion

We have provided general relationships which can be used to develop explicit expressions for chemical potential derivatives, pmvs, and the isothermal compressibility of any mixture of n components in terms of KB integrals. Our choice of the molality concentration scale makes species 1 unique. Consequently, some of the ‘‘symmetry’’ in the expressions that might be observed for molarity or mole fraction based derivatives is lost using the current notation. We consider this to be an acceptable sacrifice in many cases. It is therefore informative to compare and relate the expressions generated here with those developed by other approaches, especially when considering symmetric ideal solutions. To do this we will refer to the expressions of Smith

for ternary solutions which represent the most condensed form for the chemical potential derivatives.²⁴ They are easily expanded to provide the expressions of Ruckenstein and Shulgin²⁷ and Ben-Naim.⁴ Smith provided the following expressions:

$$\begin{aligned}\beta\left(\frac{\partial\mu_1}{\partial\ln N_2}\right)_{T,P,N_1,N_3} &= -\frac{\rho_2 A_3}{\rho_1 A_2 A_3 + \rho_2 A_1 A_3 + \rho_3 A_1 A_2}, \\ \beta\left(\frac{\partial\mu_2}{\partial\ln N_2}\right)_{T,P,N_1,N_3} &= \frac{\rho_1 A_3 + \rho_3 A_1}{\rho_1 A_2 A_3 + \rho_2 A_1 A_3 + \rho_3 A_1 A_2}, \\ \beta\left(\frac{\partial\mu_3}{\partial\ln N_2}\right)_{T,P,N_1,N_3} &= -\frac{\rho_2 A_1}{\rho_1 A_2 A_3 + \rho_2 A_1 A_3 + \rho_3 A_1 A_2},\end{aligned}\tag{5.58}$$

where the A's are given by

$$\begin{aligned}A_1 &= 1 + \rho_1(G_{11} + G_{23} - G_{12} - G_{13}), \\ A_2 &= 1 + \rho_2(G_{22} - G_{23} - G_{12} + G_{13}), \\ A_3 &= 1 + \rho_3(G_{33} - G_{23} + G_{12} - G_{13}).\end{aligned}\tag{5.59}$$

Comparison with the expressions provided in Equation 5.40 indicate that

$$\begin{aligned}A_1 &= \rho_1 N_{32}^+ / \rho_2 = \rho_1 N_{23}^+ / \rho_3, \\ A_2 &= 1 + N_{22}^+ - N_{32}^+ = 1 + N_{22}^+ - \rho_2 A_1 / \rho_1, \\ A_3 &= 1 + N_{33}^+ - N_{23}^+ = 1 + N_{33}^+ - \rho_3 A_1 / \rho_1.\end{aligned}\tag{5.60}$$

Specific combinations of KB integrals often appear repeatedly in other formulations. For instance, one can define for $i \neq j$,

$$\eta_{ij} = \rho_i + \rho_j + \rho_i \rho_j (G_{ii} + G_{jj} - 2G_{ij}) = \rho_i A_j + \rho_j A_i.\tag{5.61}$$

In the current notation it is found that

$$\eta_{ii} = \rho_i (1 + N_{ii}^+).\tag{5.62}$$

We attempted to find a similar factorization and relationships as found in Equations 5.59 and 5.60 for four component systems, but were unsuccessful.

The application of the KB theory to symmetrical ideal solutions is also of interest. Ben-Naim⁴ has shown that for a general n component mixture to display symmetric ideal behavior then one must have $\Delta G_{ij} = G_{ii} + G_{jj} - 2G_{ij} = 0$ for all ij pairs. Therefore, in the current notation one must have $\rho_l N_{ii}^+ = \rho_i$ for symmetric ideal solutions. In addition, one finds $A_1 = A_2 = A_3 = 1$ for symmetric ideal ternary solutions.

5.4 Conclusions

Using a new approach we have developed explicit relationships for KB integrals in four and five component solution mixtures. In our opinion, the use of molalities as concentration variables provides the simplest path to expressions in multicomponent solutions. We are currently using this type of approach to study biologically relevant systems containing five or more components.

For a general n component system there are $n(n+1)/2$ unique G_{ij} integrals. To determine the integrals from experimental data using the KB inversion approach requires 1 isothermal compressibility value, $n-1$ independent pmvs, and $n(n-1)/2$ independent μ_{ij} values as a function of composition. This has been achieved for ternary systems.³² As one moves beyond ternary systems the experimental data becomes increasingly more difficult to obtain. Consequently, we envision the major use for the expressions provided here will involve either, theoretical estimates of the KB integrals, or simulated values of the integrals. In either case, the exact KB expressions provide a solid foundation for investigating these complicated solution mixtures.

5.5 References

1. Kirkwood, J. G.; Buff, F. P. *Journal of Chemical Physics* 1951, 19, 774-777.
2. Ben-Naim, A. *statistical thermodynamics for chemists and biochemists*, (Plenum Press, New York, 1992).
3. Matteoli, E.; Mansoori, G. A. *Fluctuation theory of mixtures*, (Taylor & Francis, New York, 1990).
4. Ben-Naim, A. *molecular theory of solutions*, (Oxford University Press, New York, 2006)
5. O'Connell, J. P. *Aiche Journal* 1971, 17, 658-663.
6. Chialvo, A. A. *Journal of Physical Chemistry* 1993, 97, 2740-2744.
7. Smith, P. E. *Journal of Physical Chemistry B* 1999, 103, 525-534.
8. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2001, 105, 11513-11522.
9. Ruckenstein, E.; Shulgin, I. *Industrial & Engineering Chemistry Research* 2002, 41, 4674-4680.
10. Shulgin, I. L.; Ruckenstein, E. *Biophysical Chemistry* 2005, 118, 128-134.
11. Mazo, R. M. *Journal of Physical Chemistry B* 2006, 110, 24077-24082.
12. Ben-Naim, A. *Journal of Chemical Physics* 1975, 63, 2064-2073.
13. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 18716-18724.
14. Shimizu, S.; Boon, C. L. *Journal of Chemical Physics* 2004, 121, 9147-9155.
15. Shimizu, S.; McLaren, W. M.; Matubayasi, N. *Journal of Chemical Physics* 2006, 124, 234905.
16. Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Biochemistry and Biophysics* 2008, 50, 1-22.

17. Ben-Naim, A. *Cell Biophysics* 1988, 12, 255-269.
18. Marcus, Y. *Monatshefte Fur Chemie* 2001, 132, 1387-1411.
19. Tronel-Peyroz, E., Douillard, J. M.; Benne, R.; Privat, M. *Langmuir* 1989, 5, 54-58.
20. Chen, F.; Smith, P. E. *Journal of Physical Chemistry B* 2008, 112, 8975-8984.
21. Weerasinghe, S; Smith, P. E. *Journal of Physical Chemistry B* 2003, 118, 10663-10670.
22. Aburi, M. and Smith, P. E., *Journal of Physical Chemistry B* 2004, 108, 7382-7388.
23. Smith, P. E. *Journal of Physical Chemistry B* 2004, 108, 16271-16278.
24. Smith, P. E. *Biophysical Journal* 2006, 91, 849-856.
25. O'Connell, J. P. *Molecular Physics* 1971, 20, 27-33.
26. Ramirez, R.; Mareschal, M.; Borgis, D. *Chemical Physics* 2005, 319, 261-272.
27. Ruckenstein, E.; Shulgin, I. *Fluid Phase Equilibria* 2001, 180, 345-359.
28. Hall, D. G. *Transactions of the Faraday Society* 1971, 67, 2516-2524.
29. Schurr, J. M.; Rangel, D. P.; Aragon, S. R. *Biophysical Journal* 2005, 89, 2258-2276.
30. Shulgin, I. L.; Ruckenstein, E. *Journal of Chemical Physics* 2005, 123, 054909/1-9.
31. Smith, P. E. *Journal of Physical Chemistry B* 2006, 110, 2862-2868.
32. Matteoli, E.; Lepori, L. *Journal of the Chemical Society-Faraday Transactions* 1995, 91, 431-436.

CHAPTER 6 - Pairwise Preferential Interaction Model

Abstract

A Pairwise preferential interaction model (PPIM), characterized by KB integrals, is developed to quantify and characterize the interactions between functional groups observed in peptides. The existing experimental data are analyzed to determine preferential interaction parameters for different amino acid and small peptide systems in aqueous solutions. Then, the preferential interactions (PIs) between function groups on those peptides are isolated and quantified using pairwise additivity. The PPIM approach provides consistent estimates for the same pair interactions obtained from different solute molecules. Furthermore, these interactions are chemically intuitive.

6.1 Introduction

It is well known that peptide and protein aggregation are directly involved with many age-related diseases and aging itself¹⁻⁴. A better understanding of protein aggregation would hopefully lead to the prediction and even prevention, of these the undesirable conditions. Hence, a number of studies have been pursued to understand and predict the misfolding and subsequent aggregation of proteins^{4,5}. However, it is still unclear why certain peptides and proteins tend to aggregate.

In principle, aggregation in a solution mixture results from a shifted balance in the intermolecular interactions between solute and solvent. If the solute-solute interactions are larger than solute-solvent interactions, self-association is likely to occur, and vice versa: i.e. the tendency for aggregation can be predicted using the difference between solute-solute and solute-solvent interactions. Hence, it is reasonable to express the difference between solute-solute and solute-solvent interactions using a quantitative term. The concept of preferential interactions (PI) has been introduced previously⁶. KB integrals can play an important role in quantifying these PIs⁶. Proteins are usually large molecules using 20 amino acids as building blocks. Therefore, it would be more useful to quantify the interactions between amino acids, or even between functional groups, rather than to deal with the protein as a whole. This requires that the total interaction between two peptides or proteins can be decomposed to interactions between component functional groups.

Here, a pairwise preferential interaction model (PPIM) is developed to quantify and characterize the interactions between functional groups in peptides and proteins. First, the existing experimental data are analyzed to determine preferential interaction parameters for different amino acid and small peptide systems in aqueous solutions. Then, the PIs between

function groups on those peptides are isolated and quantified using a pairwise preferential approach.

6.2 Pairwise Preferential Interaction Model (PPIM) Approach

6.2.1 Thermodynamics of Solutions and KB Theory

The notation used here follows the common definition where the subscripts 1 and 2 refer to the primary solvent (usually water), and the solute, respectively. Chemical potential (μ) plays an important role in thermodynamic changes in a system. Under thermodynamic control, changes in the chemical potential (μ) of a species in a system reflects how the species can bring about change in the system: both physical and chemical changes. According to statistical mechanics, the chemical potential of a species can be expressed as⁷,

$$\mu = W + RT \ln[\Lambda^3 \rho q^{-1}] = \mu^* + RT \ln[\Lambda^3 \rho] \quad (6.1)$$

Here, q is the internal partition function of a molecule, N the number of the species, V the volume of the system, $\rho=N/V$ a number density (or molar concentration), and Λ the thermal de Broglie wavelength of the species. The first term (W) quantifies contributions of the interactions among molecules to the chemical potential on the addition of a molecule. If there is no interaction in the system, $W = 0$ and only the second term ($RT \ln[\Lambda^3 \rho p^{-1}]$) will be left, simply indicating the chemical potential of an ideal gas at the same temperature and density. Ben-Naim has developed the symbol $\mu^* = W - RT \ln q$ to represent the pseudo chemical potential⁷. The pseudo chemical potential captures the free energy change for transfer of a molecule from a fixed position in a vacuum to a fixed position in the solution. This will be the same as the work

required for the corresponding cavity formation⁷. Using the pseudo chemical potential, the entropy of mixing can be eliminated which is not directly related to the intermolecular interactions⁷.

Kirkwood-Buff (KB) theory plays an important role in the development of this model. KB theory is an exact theory with no restrictions on molecular size and provides a link between the relative solution structure and the corresponding thermodynamics. Expressions for changes and derivatives in the pseudo chemical potential as well as the total chemical potential required in Equation 6.1 are easily obtained.

6.2.2 *Preferential Interactions*

We need to develop a method to analyze the experimental data and a model which can be used to quantify and predict peptide aggregation. Here we present and apply a new approach to quantitatively express features of the interactions between functional groups in amino acids. It is limited in some aspects (discussed below), but in the absence of other predictive approaches it seems worthy of development. In many respects it is somewhat analogous to, and on the same level as, the Chou-Fasman type of approach used to predict secondary structure elements in proteins^{8,9}. Using KB theory it is quite easy to show that for any thermodynamically stable mixture of a solute (2) and solvent (1) we can write that⁷,

$$-\beta \left(\frac{\partial \mu_2^*}{\partial \rho_2} \right)_{T,P} = - \left(\frac{\partial \ln y_2}{\partial \rho_2} \right)_{T,P} = \frac{G_{22} - G_{12}}{1 + \rho_2(G_{22} - G_{12})} \quad (6.2)$$

where y_2 is the molar activity coefficient of the solute. The above expression reduces to the numerator in the limit of infinite dilution of the solute (2). The value of $G_{22}-G_{21}$ at infinite

dilution is the central quantity of interest in this work. We will define this as the preferential interaction (PI) between two infinitely dilute solute molecules in a binary system,

$$P_{22}^{\infty} = G_{22}^{\infty} - G_{21}^{\infty} \quad (6.3)$$

The PI defined here is the same as previous definitions of preferential solvation (PS)¹⁰, except for the infinitely dilute solute restriction. However, it will be used in a different manner. The PI at infinite dilution of solute quantifies the interaction between two solute molecules in a large excess of solvent (no additional many body solute interactions present). It results from a balance of the pair solute-solute and solute-solvent interactions. A positive value indicates a favorable solute-solute interaction which tends towards solute association. A negative value indicates a favorable solvation which tends towards solute hydration and low solute self association. A value of zero indicates a balance of the interactions, *i.e.* an ideal solution. The above expression indicates that if the molar activity coefficient decreases with molarity, then the solute must display a tendency towards self association. The approach therefore provides a way to quantify the degree of molecular association.

6.2.3 *Experimental Data*

It is common practice to perform an analysis of the experimental data (solute or water activity, partial molar volumes, and isothermal compressibilities), to obtain the composition dependent KB integrals and use them to quantify preferential solvation in solutions. In principle, we want to do the same here, but there is a slight problem. We require the KB integrals at infinite dilution of the solute (2). Traditionally, obtaining reliable values of G_{ij} at low concentrations of i or j has been difficult^{11,12}. For instance, the partial pressure of a solute above a solute solvent solution is difficult to determine experimentally at low solute concentrations. Consequently, the

KB integrals tend to be sensitive to the exact fitting procedure used to determine the solute activity. Part of the problem is that the two most common fitting equations for the solute activity, the Wilson and Redlich-Kister equations, are essentially empirical in nature and therefore may not fully capture the physics of the process. Recently, a rigorous statistical mechanical approach based on a semi grand ensemble (open to solute, closed to solvent) was presented which can accurately model the molal activity coefficients of nonvolatile solutes over large concentration ranges using just one or two parameters¹³.

In the original approach by Rosgen et al either the molal or molar activity coefficient was fitted using ratios of polynomials in the solute activity (a)^{13,14}. Here, we will extend the approach to fit both the molal and molar activities coefficient data simultaneously. This requires the corresponding density data. The derivations are quite straight forward but too long to present here. The final result for the solute molality (m_2) and solute molarity (c_2) after including terms up to a^2 is given by,

$$m_2 = \frac{a + 2a^2 A_1^{-1}}{1 + aA_2 + a^2 A_2 A_1^{-1}} \quad \text{and} \quad c_2 = \frac{a + 2a^2 A_1^{-1}}{d_0^{-1} + aV_1 + a^2 V_2 A_1^{-1}} \quad (6.4)$$

where we have used the molality scale solute activity ($a = \gamma_2 m_2$) in both cases and d_0 is the density of pure solvent. At this level the fitting equations involve four unknowns (A_1 , A_2 , V_1 , and V_2). The advantage of this approach is that one can fit both activity coefficients over a large range of compositions with just four parameters, one of which (A_1) appears in both expressions. As the molar and molal concentrations are related through the solution density, this is equivalent to fitting one of the activity coefficients and the density simultaneously. Further analysis of the limiting values of the activity derivatives and comparison with Equation 6.4 indicates that the PI value required for this work is then given by,

Table 6.1 Fitting parameters for Equation 6.4. Errors in the final PIs are estimated as <10%. Several parameters adopted values close to zero during the fitting procedure. In this case, the fit was repeated with these values set at zero. All the experimental peptide data refer to racemic mixture taken from references¹⁵⁻¹⁷.

Solute (2)	Abbrev.	Molal activity range	A ₁ (mol/kg)	A ₂ (kg/mol)	V ₁ (L/mol)	V ₂ (L/mol)	P ₂₂ [∞] (cm ³ /mol)
NMA	N	0.5-2.1	56.82	-0.101	-0.020	-0.134	55
Gly	G	0.1-3.3	6.83	0.075	0.118	0.174	176
Gly ₂	G2	0.2-1.7	2.70	0.164	0.241	0.330	502
Gly ₃	G3	0.1-0.3	1.68	0.129	0.240	0.372	955
Ala	A	0.1-1.4	18.43	0.0	0.073	0.0	36
Ala ₂	A2	0.2-1.0	2.89	0.555	0.651	0.0	43
AlaGly	AG	0.2-1.0	2.09	0.463	0.814	0.0	147
GlyAla	GA	0.2-1.0	2.57	0.366	0.673	0.0	108

$$P_{22}^{\infty} = 2(d_0 A_1)^{-1} - V_1 \quad (6.5)$$

The above fitting procedure involves an expression for the concentration in terms of the activity instead of the normal situation of activity in terms of concentration. While this is somewhat unusual, it is a perfectly valid approach to fit the experimental data. If required, the above equations can be used to express the molal activity coefficient in terms of the molality via solutions to a quadratic equation in γ_2 ¹⁴. This is not necessary here. We will see that the above approach produces excellent results. This is important as it provides significant confidence in the resulting KB integrals, including those evaluated at low solute concentrations.

Using KB theory and analyzing the experimental data one can obtain the preferential interaction between two solute molecules at infinite dilution thus providing fundamental information on the degree of molecular association. In this section we apply the PPIM approach to analyze existing experimental data on activity coefficients and densities¹⁵⁻¹⁷ in order to isolate

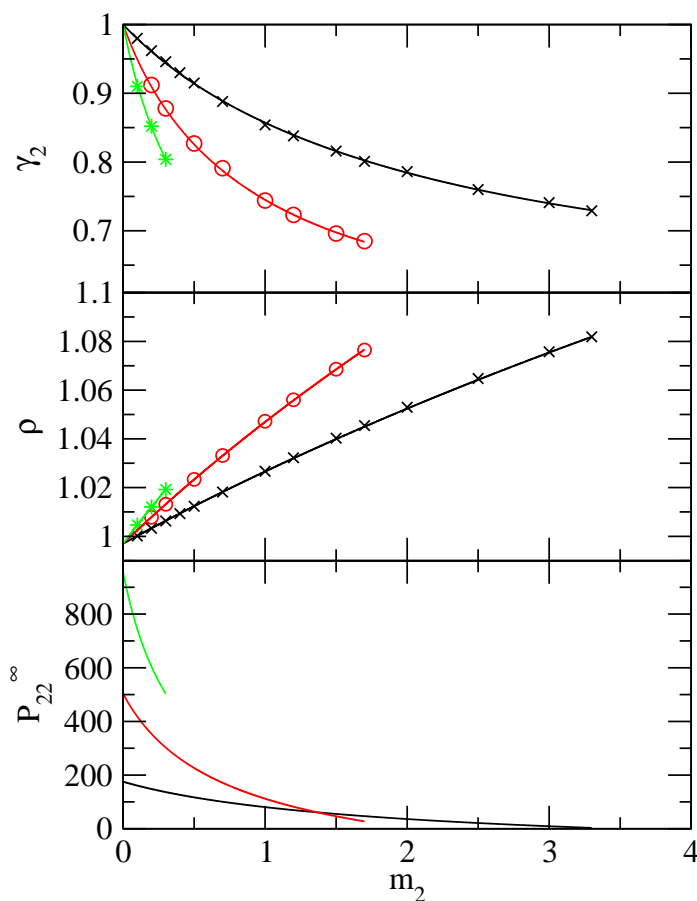


Figure 6.1 Experimental data at 298K and 1atm for the molal activity coefficients (γ_2), solution density (ρ in g/cm^3) and the resulting preferential interactions (P_{22}^∞ in cm^3/mol) for a series of Gly_n peptides as a function of peptide molality (m_2). The symbols indicate the raw experimental data while the solid lines are the corresponding fits after using Equation 6.4 and the parameters presented in Table 6.1. Gly (X), Gly_2 (O) and Gly_3 (*). All PIs are positive indicating a tendency for self association which decreases as the solute concentration increases.

a series of P_{22} 's. Our initial investigations have focused on (N-methylacetamide) NMA and a series of small amino acids and peptides. In the zwitterionic form the amino acids are

nonvolatile, while NMA is a solid at 298 K and a good model for the peptide functional group.

Figure 6.1 provides an example of the quality of fits one can obtain using the above fitting equations. The molal activity and densities are very well reproduced. In contrast, the original activity data alone for Gly was fitted to an expression involving 4 terms and 4 unknowns¹⁵. The corresponding parameters and resulting PIs are presented in Table 6.1. It can be seen that the PI values are positive for all the systems presented here. This indicates aggregation of the solutes at low solute concentrations. The aggregation of the Gly_n peptides increases with n for reasons which will become clear shortly.

6.2.4 *Decomposition Approach*

While the previous results are interesting in themselves, we want to take this a step further and develop a model which can be used to rationalize the available data, and eventually make predictions. To do this we need to be able to decompose the PI between two solute molecules into a combination of effects from the different functional groups present in the molecule. Our approach is to investigate the simplest model possible and then develop the model as it is applied to more systems, where the possibility of small corrections may be required. The basic assumption of the model proposed here is that the preferential interaction between a pair of solute molecules can be decomposed into a series of pairwise preferential interactions between the different chemical groups present in the solutes. This type of approach has been used before for enthalpies of mixing¹⁸. Chemical groups involve the usual chemical functionalities such as hydroxyl, amide, carbonyl, etc, as well as hydrocarbon interactions such as that between methyl groups (Me-Me). Hence, the pairwise preferential interaction model (PPIM) expresses the total PI between two solute molecules (*i* and *j*) at infinite dilution as,

$$P_{ij}^{\infty} = \sum_a \sum_b p_{ab}^{\infty} \quad (6.6)$$

where the sum is over all functional groups (a) on molecule i and groups (b) on molecule j . The p_{ab} 's have then to be determined. This represents one of the main objectives. As there are far fewer functional groups than potential solutes, the PPIM model can be used to predict the degree of molecular association between many different solutes in solution. In the next section we present evidence that the model is reasonable.

To illustrate the model further let us consider a solute molecule which contains two functional groups a and b . Both groups on each molecule interact with each other. In the PPIM model it is assumed that the solute to solute interaction at infinite dilution can be written as a combination of the group interactions and so $G_{22}-G_{21} = p_{aa} + p_{bb} + 2p_{ab}$. The values of p_{aa} and p_{bb} can be obtained from the results for a solute where only one of the groups (a or b) is present. The value of p_{ab} corresponds to the preferential interaction between two different groups and will in general be obtained via a decomposition process involving the data for many solutes (see later). In terms of KB integrals one has $G_{22}-G_{21} = (G_{aa}-G_{a1}) + (G_{bb}-G_{b1}) + 2(G_{ab}-\frac{1}{2}G_{a1}-\frac{1}{2}G_{b1})$. There is no explicit dependence of the pair interaction on the distance between the different groups on the solute molecules in the PPIM approach. This is because the KB integrals quantify changes in the molecular distributions over all distances. However, it is to be expected that the pair interactions will have a limited range, the extent of which remains to be determined.

6.2.5 Evaluation of Group Contributions

The PIs presented in Table 6.1 can be used to evaluate the required group contributions. This is illustrated in Table 6.2 where we have indicated the number and type of the group interactions

Table 6.2 Decomposition of the molecule PIs into group contributions. Solute abbreviations are given in Table 6.1. Group abbreviations are peptide (P), charge (Q) and methyl (M). The QQ group interaction includes all combinations between two zwitterions (i.e. two +/-, one +/+, one -/- interaction). The PQ interaction is a combination of the P+ and P- interactions (similarly for MQ).

Solute	Number of Group Pairs						
	PP	QQ	PQ	MQ	MP	MM	
N	1						
G		1					
G2	1	1	2				
G3	4	1	4				
A		1		2		1	
A2	1	1	2	4	4	4	
AG	1	1	2	2	2	1	
GA	1	1	2	2	2	1	
	Group p_{ab} (cm ³ /mol)						Source
	55						N
	63						½ (G3-2G2+G)
		176					G
			135				½ (G2-G-N)
			144				½ (G3-G2-3N)
					-112		½ (AG-A-PP-2PQ)
						130	½ (A2-AG-A+QQ-2MP)
				-135			½ (A-QQ-MM)
				-137			½ (AG-QQ-PP-2PQ-2MP-MM)
					-131		½ (GA-A-PP-2PQ)
						169	½ (A2-GA-A+QQ-2MP)
				-155			½ (A-QQ-MM)
				-157			½ (GA-QQ-PP-2PQ-2MP-MM)
Average	59	176	140	-146	-122	149	

for each pair of solute molecules. By manipulation of the data one can extract the individual p_{ab} values which are also presented in Table 6.2. There are several features of the resulting data which are encouraging and important. First, comparison of the p_{ab} value for the peptide group obtained from a separate analysis of the Gly_n peptides and NMA indicate essentially the same result of 59 cm³/mol. As these correspond to two totally different types of solute systems it is

satisfying that the same result is obtained for both. A consistent result for the peptide to amino acid termini (PQ) is also obtained from two different decompositions. Inclusion of the Ala-Gly (and Gly-Ala) data provided estimates for the PI between a methyl group and the two charged termini (MQ) as well as the MM and MP interactions. Again, consistent results were obtained for the MP interaction obtained using the same data but from two different routes. Some differences were observed between the Ala-Gly and Gly-Ala results. However, these differences were relatively small (<10%) and hence, to a first approximation, the sequence dependence effect appears to be minor and will therefore be neglected in our initial studies.

6.3 Discussion

The PPIM has several advantages. First, it is very simple. Second, the PIs include both short range and long range contributions to the changes in the solution distribution. Hence, both the effects of direct molecular interactions (hydrogen bonds) and the consequences for the packing of solvent molecules around the solutes are included in the model. The latter is traditionally very difficult to determine. Third, the use of the PIs ensures that one focuses on effects that lead to changes in the pseudo chemical potential, *i.e.* that result solely from intermolecular interactions. Fourth, by performing a simultaneous fitting of the activity and density data using a rigorous statistical mechanical theory one ensures accurate experimental data is obtained at low solute concentrations for the determination of the PIs.

One of the disadvantages of the model is that, in its present form, it is restricted to infinitely dilute solute molecules. Obviously, at finite solute concentrations the solute-solute interactions are modified by the presence of other solute molecules. This changes the value of $G_{22}-G_{21}$. In addition, the solutes cannot be ionic in nature. For ions the distribution at low solute

concentrations is dominated by the electroneutrality constraint,^{19,20} and is therefore independent of the character of the ionic species involved. Furthermore, the current model does not distinguish between different peptide sequences with the same composition of amino acids, such as Gly-Ala and Ala-Gly, or between different chiral molecules. As one of the potential applications of the PPIM is to understand peptide aggregation, this could present a problem. However, the experimentally observed activity coefficients of Gly-Ala and Ala-Gly are very similar,¹⁷ and hence the corresponding p_{ab} 's do not differ significantly (see later). Finally, the present analysis is restricted to that of nonvolatile solutes in water, primarily due to the fitting procedure. This does not mean the model is not applicable to other solvents or to molecules which are volatile. Once the p_{ab} for a particular group has been determined it can be used for any solute in that particular solvent. It is the determination of the p_{ab} 's from representative molecules which requires the use of a fitting equation valid only for nonvolatile solutes. In summary, while there are some restrictions, we feel the model has significant promise, and due to the many potential applications deserves to be developed further. It can certainly provide the first order effect of group pair interactions on the association process.

We also note that the signs estimated by the model are consistent with those intuitively expected from the point of view of physical chemistry. For instance, it is expected the interaction between a methyl group and both the charged group (MQ) and the peptide group (MP) would be unfavorable from a desolvation perspective. This is indicated by the negative values of p_{ab} . In contrast, the positive value of p_{ab} for the MM interaction indicates significant self association, which is in agreement with a simple picture of the hydrophobic effect. The weaker self association of the peptide group is not unrealistic, although it is difficult to predict the exact balance between self association and solvation. This, of course, is precisely what makes the

present model so attractive. The positive value for PP indicates a tendency for the peptide group to self associate. The implications of this for protein folding or peptide aggregation are unclear but deserve further study. The relatively large PI between the peptide and charged groups appears to be just as favorable as the MM self association. In unrelated simulation studies we have observed significant similar interactions between Ser hydroxyl groups and the charged C terminus of small peptides.²¹ The data in Table 6.2 suggest that the reason for the increased self association or aggregation of the Gly_n peptides with increasing n lies in the fact that all the group interactions are positive and thus promote self association, with the dominant effect occurring for the PQ interaction. The self association of Ala is lower than that of Gly as the addition methyl group provides MQ interactions which are unfavorable, presumably due to the effect of the methyl groups on the solvation of the charged groups. Finally, the positive value for p_{ab} between methyl groups is in agreement with the estimates provided previously by analysis of hydrocarbon aggregation in water.¹²

6.4 Conclusion

In summary, the interactions between different groups present on a solute have been quantified using the simple pairwise preferential interaction model. In evaluation, the PPIM provides consistent estimates for the same pair interactions obtained from different solute molecules, and interactions which are chemically intuitive. Further study is needed to generalize the model. The model suggests a role for PQ and MM interactions in peptide aggregation which should be experimentally testable.

6.5 References

1. Truant, R.; Atwal, R. S.; Desmond, C.; Munsie, L.; Tran, T. *FEBS Journal* 2008, 275, 4252-4262.
2. Batchelor, J. D.; Olteanu, A.; Tripathy, A.; Pielak, G. J. *Journal of the American Chemical Society* 2004, 126, 1958-1961.
3. Stefani, M.; Dobson, C. M. *Journal of Molecular Medicine* 2003, 81, 678-699.
4. Cohen, F. E.; Kelly, J. W. *Nature* 2003, 426, 905-909.
5. Dobson, C. M. *Nature* 2003, 426, 884-890.
6. Kang, M.; Smith, P. E. *Fluid Phase Equilibria* 2007, 256, 14-19.
7. Ben Naim, A. *Statistical Thermodynamics for Chemists and Biochemists*, Plenum Press: New York, 1992.
8. Chou, P. Y.; Fasman, G. D. *Trends in Biochemical Sciences* 1977, 2, 128-131.
9. Chou, P. Y.; Fasman, G. D. *Biophysical Journal* 1977, 17, A53.
10. Matteoli, E.; Mansoori, G. A. *Fluctuation Theory of Mixtures*, Taylor & Francis: New York, 1990.
11. Matteoli, E.; Lepori, L. *Journal of Chemical Physics* 1984, 80, 2856-2863.
12. Liu, H.; Ruckenstein, E. *Journal of Physical Chemistry B* 1998, 102, 1005-10012.
13. Rosgen, J.; Pettitt, B. M.; Perkyuns, J.; Bolen, D. W. *Journal of Physical Chemistry B* 2004, 108, 2048-2055.
14. Rosgen, J.; Pettitt, B. M.; Bolen, D. W. *Biochemistry* 2004, 43, 14472-14484.
15. Smith, E. R. B.; Smith, P. K. *Journal of Biological Chemistry* 1937, 117, 209-216.

16. Kuramochi, H.; Noritomi, H.; Hoshino, D.; Nagahama, K. *Journal of Chemical and Engineering Data* 1997, 42, 470-474.
17. Smith, E. R. B.; Smith, P. K. *Journal of Biological Chemistry* 140, 135, 273-279.
18. Savage, J. J.; Wood, R. H. *Journal of Solution Chemistry* 1976, 5, 733-750.
19. Kusalik, P. G.; Patey, G. N. *Journal of Chemical Physics* 1987, 86, 5110-5116.
20. Chitra, R.; Smith, P. E. *Journal of Physical Chemistry B* 2002, 106, 1491-1500.
21. Lei, H.; Smith, P. E. *Biophysics Journal* 2003, 85, 3513-3520.

CHAPTER 7 - Summary and Future Work

Molecular dynamics (MD) simulations have provided great insights into intermolecular interactions in biological systems. Kirkwood-Buff (KB) theory provides a direct relationship between the relative distribution of species in solution and the thermodynamics of that solution. Furthermore, KB theory can be used to interpret both experimental and computational data. Hence, a combination of KB theory and computer simulation provides a powerful tool for the study of a variety of systems. Here, a series of applications of KB theory are demonstrated. KB theory is applied to study preferential interactions in biological system and has illustrated that the common force fields for peptide and protein do not reproduce the delicate balance between solute-solute and solute-solvent intermolecular interactions. Hence, a new force field for amides is developed using KB theory to accurately describe the solution thermodynamics and intermolecular interactions. It has been then shown that KBFF is competitive with the common major force fields for peptide and protein. A general recursion relationship has been then provided which can be used to generate chemical potential derivatives for 4 or higher component solution mixtures which are prevalent in biological systems. In addition, KB theory is used as a foundation for the development of a pairwise preferential interaction model (PPIM) to quantify and characterize the interactions between functional groups observed in peptides.

The Smith group has been working towards the development of the new force field (KBFF) to cover all the amino acids and other common components of biological systems. When the KBFF is complete, it will hopefully lead us to more accurate simulations of biological

systems. In particular, it will be useful for protein folding, rational drug design, and ligand receptor docking studies.

Appendix A - Personal Publications and Copies of Permission Letter from the Publisher

A.1 Personal Publication List

The following is a list of publications that Myungshim Kang has co-authored during her career at Kansas State University.

Kang, Myungshim; Smith, Paul E., A Kirkwood-Buff derived force field for amides. *Journal of Computational Chemistry* (2006), 27(13), 1477-1485. → Chapter 3.

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

Kang, Myungshim; Smith, Paul E., Kirkwood-Buff theory for four and five component mixtures *Journal of Chemical Physics*, (2008), 128, 244511 → Chapter 5.

Pierce, Veronica; **Kang, Myungshim;** Aburi, Mahalaxmi; Weerasinghe, Samantha; Smith, Paul E., Recent applications of Kirkwood-Buff theory to biological systems. *Cell Biochemistry and Biophysics* (2008), 50 (1), 1-22.

Dani, Raj Kumar; **Kang, Myungshim**; Kalita, Mausam, Smith, Paul E.; Bossmann, Stefan H.; Chikan, Viktor, MspA Porin Gold Nanoparticle Assemblies: Enhanced Binding through a Controlled Cysteine Mutation. *Nano Letters*, (2008), 8(4), 1229-1236.

Basel, Matt; Dani, Raj; **Kang, Myungshim**; Pavlenok, Mikhail; Chikan, Viktor; Smith, Paul E.; Niederweis, Michael; Bossmann, Stefan H., Direct Observation of Gold Nanoparticle Assemblies with the Porin MspA on Mica, *ACS Nano*, 2009, ASAP

A book chapter

Weerasinghe, Samantha; Gee, Moon Bae; **Kang, Myungshim**; Benteinitis, Nikolaos; Smith, Paul E., Developing force fields from the microscopic structure of solutions: The Kirkwood-Buff approach, *Modeling Solvent Environments*, submitted. Wiley-VCH, Weinheim, 2009, in press

A.2 Copies of the Permission Letter from the Publisher

A.2.1 ELSEVIER LICENSE

ELSEVIER LICENSE
TERMS AND CONDITIONS

Dec 10, 2008

This is a License Agreement between Myungshim Kang ("You") and Elsevier ("Elsevier").
The license consists of your order details, the terms and conditions provided by Elsevier,
and the payment terms and conditions.

Supplier	Elsevier Limited The Boulevard, Langford Lane Kidlington, Oxford, OX5 1GB, UK
Registered Company Number	1982084
Customer name	Myungshim Kang
Customer address	111 Willard Hall Manhattan, KS 66506
License Number	2082580003078
License date	Dec 05, 2008
Licensed content publisher	Elsevier
Licensed content publication	Fluid Phase Equilibria
Licensed content title	Preferential interaction parameters in biological systems by Kirkwood–Buff theory and computer simulation
Licensed content author	Myungshim Kang and Paul E. Smith
Licensed content date	1 August 2007
Volume number	256
Issue number	1-2
Pages	6
Type of Use	Thesis / Dissertation

Portion	Full article
Format	Both print and electronic
You are an author of the Elsevier article	Yes
Are you translating?	No
Purchase order number	
Expected publication date	Feb 2008
Elsevier VAT number	GB 494 6272 12
Permissions price	0.00 USD
Value added tax 0.0%	0.00 USD
Total	0.00 USD
Terms and Conditions	

INTRODUCTION

1. The publisher for this copyrighted material is Elsevier. By clicking "accept" in connection with completing this licensing transaction, you agree that the following terms and conditions apply to this transaction (along with the Billing and Payment terms and conditions established by Copyright Clearance Center, Inc. ("CCC"), at the time that you opened your Rightslink account and that are available at any time at <http://myaccount.copyright.com>).

GENERAL TERMS

2. Elsevier hereby grants you permission to reproduce the aforementioned material subject to the terms and conditions indicated.

3. Acknowledgement: If any part of the material to be used (for example, figures) has appeared in our publication with credit or acknowledgement to another source, permission must also be sought from that source. If such permission is not obtained then that material may not be included in your publication/copies. Suitable acknowledgement to the source must be made, either as a footnote or in a reference list at the end of your publication, as follows:

"Reprinted from Publication title, Vol /edition number, Author(s), Title of article / title of chapter, Pages No., Copyright (Year), with permission from Elsevier [OR APPLICABLE SOCIETY COPYRIGHT OWNER]." Also Lancet special credit - "Reprinted from The Lancet, Vol. number, Author(s), Title of article, Pages No., Copyright (Year), with permission from Elsevier."

4. Reproduction of this material is confined to the purpose and/or media for which permission is hereby given.

5. Altering/Modifying Material: Not Permitted. However figures and illustrations may be altered/adapted minimally to serve your work. Any other abbreviations, additions, deletions and/or any other alterations shall be made only with prior written

authorization of Elsevier Ltd. (Please contact Elsevier at permissions@elsevier.com)

6. If the permission fee for the requested use of our material is waived in this instance, please be advised that your future requests for Elsevier materials may attract a fee.

7. Reservation of Rights: Publisher reserves all rights not specifically granted in the combination of (i) the license details provided by you and accepted in the course of this licensing transaction, (ii) these terms and conditions and (iii) CCC's Billing and Payment terms and conditions.

8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unrevoked license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions

described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

LIMITED LICENSE

The following terms and conditions apply to specific license types:

15. **Translation:** This permission is granted for non-exclusive world **English** rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website:** The following terms and conditions apply to electronic reserve and author websites:

Electronic reserve: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:

This license was made in connection with a course,

This permission is granted for 1 year only. You may obtain a license for future website posting,

All content posted to the web site must maintain the copyright information line on the bottom of each image,

A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or, for books, to the Elsevier homepage at <http://www.elsevier.com>,

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

All content posted to the web site must maintain the copyright information line on the bottom of each image, and

The permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version,

A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx>,

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:

Authors are permitted to place a brief summary of their work online only.

A hyper-text must be included to the Elsevier homepage at <http://www.elsevier.com>.

All content posted to the web site must maintain the copyright information line on the

bottom of each image

You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at <http://www.sciencedirect.com/science/journal/xxxxx> or, for books, to the Elsevier homepage at <http://www.elsevier.com>.

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other conditions**: None

v1.5

A.2.2 WILEY PERIODICALS, Inc., A WILEY COMPANY

Myungshim Kang

Kansas State University

213 CBC Building

Manhattan, KS 66506-0401

mshkang@ksu.edu

VIA FACSIMILE: 785.532.6666

Dear Myungshim Kang:

RE: Your December 02, 2008 request for permission to republish Journal of Computational Chemistry 2006, 27(13):1477-1485. This material will appear in your forthcoming thesis in print and/or on a password-protected website <<http://krex.ksu.edu>> to be published by Kansas State University.

1. Permission is granted for this use, except that if the material appears in our work with credit to another source, you must also obtain permission from the original source cited in our work.
2. Permitted use is limited to your edition described above, and does not include the right to grant others permission to photocopy or otherwise reproduce this material except for versions made for use by visually or physically handicapped persons. Up to five copies of the published thesis may be photocopied by a microfilm company.

3. Appropriate credit to our publication must appear on every copy of your thesis, either on the first page of the quoted text, in a separate acknowledgment page, or figure legend. The following components must be included: Title, author(s) and /or editor(s), journal title (if applicable), Copyright © (year and owner). Reprinted with permission of John Wiley & Sons Inc.
4. This license is non-transferable. This license is for non-exclusive English print rights and microfilm storage rights by Kansas State University only, throughout the world. This license does not extend to selling our content in any format. *For translation rights, please reapply for a license when you have plans to translate your work into a specific language*
5. Posting of the Material shall in no way render the Material in the public domain or in any way compromise our copyright in the Material. You agree to take reasonable steps to protect our copyright not limited to, providing credit to the Material as specified in Paragraph 3 above. You agree that access to the Material will be deleted from the web page no later than February 19, 2012

Sincerely,

Brad Johnson

Permissions Assistant

201.748.6786

201.748.6008 (fax)

bjohns@wiley.com

A.2.3 AMERICAN INSTITUTE OF PHYSICS LICENSE

AMERICAN INSTITUTE OF PHYSICS LICENSE

TERMS AND CONDITIONS

Jan 16, 2009

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

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

License Number	2095440531511
License date	Dec 24, 2008
Licensed content publisher	American Institute of Physics
Licensed content title	Kirkwood--Buff theory of four and higher component mixtures
Licensed content author	Myungshim Kang
Type of Use	Republish Entire Article
Requestor Type	Author
Title of your work	Molecular dynamics simulations and theory of intermolecular interactions in solution
Publisher of your work	K-State Research Exchange
Publication date of your work	02/19/2009
Billing Type	Invoice
Company	Myungshim Kang
Billing Address	111 Willard Hall Department of Chemistry Manhattan, KS 66506 United States
Customer reference info	
Total	\$0.00

Terms and Conditions

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

American Institute of Physics ("AIP") hereby grants to you the non-exclusive right and license to use and/or distribute the Material according to the use specified in your order, on a

one-time basis, for the specified term, with a maximum distribution equal to the number that you have ordered. Any links or other content accompanying the Material are not the subject of this license.

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

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

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

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 30 days of the license date. Payment should be in the form of a check or money order referencing your account number and this license number 2095440531511.

If you would prefer to pay for this license by credit card, please go to <http://www.copyright.com/creditcard> to download our credit card payment authorization form.

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

If you find copyrighted material related to this license will not be used and wish to cancel, please contact us referencing this license number 2095440531511 and noting the reason for cancellation.

Questions? customercare@copyright.com or 877-622-5543 or +1-978-646-2777.
