PROPERTIES OF BIOLOGICALLY RELEVANT SOLUTION MIXTURES
BY THEORY AND SIMULATION

by

SHU DAI

B.S., Shaanxi Normal University, China, 2009

AN ABSTRACT OF A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Abstract

Molecular Dynamics (MD) simulations have played an important role in providing detailed atomic information for the study of biological systems. The quality of an MD simulation depends on both the degree of sampling and the accuracy of force field. Kirkwood-Buff (KB) theory provides a relationship between species distributions from simulation results and thermodynamic properties from experiments. Recently, it has been used to develop new, hopefully improved, force fields and to study preferential interactions. Here we combine KB theory and MD simulations to study a variety of intermolecular interactions in solution. Firstly, we present a force field for neutral amines and carboxylic acids. The parameters were developed to reproduce the composition dependent KB integrals obtained from an analysis of the experimental data, allowing for accurate descriptions of activities involved with uncharged N-terminus and lysine residues, as well as the protonated states for the C-terminus and both aspartic and glutamic acids. Secondly, the KB force fields and KB theory are used to investigate the urea cosolvent effect on peptide aggregation behavior by molecular dynamics simulation. Neo-pentane, benzene, glycine and methanol are selected to represent different characteristics of proteins. The chemical potential derivatives with respect to the cosolvent concentrations are calculated and analyzed, and the four solutes exhibit large differences. Finally, the contributions from the vibrational partition function to the total free energy and enthalpy changes are investigated for several systems and processes including: the enthalpy of evaporation, the free energy of solvation, the activity of a solute in solution, protein folding, and the enthalpy of mixing. The vibrational frequencies of N-methylacetamide, acetone and water are calculated using density functional theory and MD simulations. We argue that the contributions from the vibrational partition function are large and in classical force fields these contributions should be implicitly included by the use of effective intermolecular interactions.

PROPERTIES OF BIOLOGICALLY RELEVANT SOLUTION MIXTURES
BY THEORY AND SIMULATION

by

SHU DAI

B.S., Shaanxi Normal University, China, 2009

A DISSERTATION

submitted in partial fulfillment of the requirements for the degree

DOCTOR OF PHILOSOPHY

Department of Chemistry
Collage of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2014

Approved by:

Major Professor
Dr. Paul E. Smith

Copyright

SHU DAI

2014

Abstract

Molecular Dynamics (MD) simulations have played an important role in providing detailed atomic information for the study of biological systems. The quality of an MD simulation depends on both the degree of sampling and the accuracy of force field. Kirkwood-Buff (KB) theory provides a relationship between species distributions from simulation results and thermodynamic properties from experiments. Recently, it has been used to develop new, hopefully improved, force fields and to study preferential interactions. Here we combine KB theory and MD simulations to study a variety of intermolecular interactions in solution. Firstly, we present a force field for neutral amines and carboxylic acids. The parameters were developed to reproduce the composition dependent KB integrals obtained from an analysis of the experimental data, allowing for accurate descriptions of activities involved with uncharged N-terminus and lysine residues, as well as the protonated states for the C-terminus and both aspartic and glutamic acids. Secondly, the KB force fields and KB theory are used to investigate the urea cosolvent effect on peptide aggregation behavior by molecular dynamics simulation. Neo-pentane, benzene, glycine and methanol are selected to represent different characteristics of proteins. The chemical potential derivatives with respect to the cosolvent concentrations are calculated and analyzed, and the four solutes exhibit large differences. Finally, the contributions from the vibrational partition function to the total free energy and enthalpy changes are investigated for several systems and processes including: the enthalpy of evaporation, the free energy of solvation, the activity of a solute in solution, protein folding, and the enthalpy of mixing. The vibrational frequencies of N-methylacetamide, acetone and water are calculated using density functional theory and MD simulations. We argue that the contributions from the vibrational partition function are large and in classical force fields these contributions should be implicitly included by the use of effective intermolecular interactions.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

IR	Infrared
MD	Molecular Dynamics
MC	Monte Carlo
QM	Quantum Mechanics
LJ	Lennard-Jones
KB	Kirkwood-Buff
PME	Particle Mesh Ewald
PEN	Neo-pentane
BEN	Benzene
Gly	Glycine
MOH	Methanol
NMA	N-methylacetamide
DFT	Density functional theory
COM	Center of mass
G_{ij}, rdf	Radial distribution function
G_{ij}, KBI	Kirkwood-Buff integral
N_{ij}	Excess coordination number
μ_i	Chemical potential
μ_{ij}	Chemical potential derivative
Γ_{ij}	Preferential binding parameter
Γ_{ij}/ρ_j, $G_{ij}-G_{i1}$	Preferential structure parameter
Q_{vib}	Vibrational partition function
ρ	Number density
γ	Activity coefficient
ν	Vibrational frequency
M	Molarity
m	Molality

Acknowledgements

I would like to give my first and foremost thank to my supervisor Prof. Paul E. Smith, as I could not do all of these without his help and support. Being a brilliant scientist himself, Prof. Smith is also an excellent and respectful mentor, as he always considers tutoring as his responsibility, and he creates an encouraging and inspiring environment that allows us to achieve research goals. Besides providing me financial aid, Prof. Smith helped me adjusting myself from a student to a researcher. He cared about my research, inspired me with new ideas, helped me troubleshoot problems, and taught me how to present as a scientist. I feel very grateful to have Prof. Smith as my supervisor, and I will continue to value his guidance and advice.

I would like to thank previous and current members in Prof. Smith lab, especially Dr. Samantha Weerasinghe, Elizabeth Ploetz, Sadish Karunaweera, Dr. Yuanfang Jiao, Nawavi Naleem, Gayani Pallewela and Nilusha Kariyawasam Manachchige. Thank you for the things you taught me, and the memorable periods we shared together.

I would like to thank Prof. Aikens group members, especially Dr. Emilie B. Guidez—my good friend, and Brian Barngrover. They helped me collect information and data on the areas I was not familiar with and provided me many constructive suggestions and advice when I was having troubles. Their generous help made my research results more meaningful.

I would like to thank Prof. Christine M. Aikens, Prof. Jianhan Chen, Prof. Viktor Chikan, and Prof. Derek A. Mosier for their serving as my Ph. D. committees. Thanks for their valuable time and effort.

I also would like to thank all the members of the Journal Club, especially Prof. Jianhan Chen, Prof. Jeremy D. Schmit, and Prof. Bin Liu. The Journal Club discussion opened my mind and I learned a lot from it.

In the end, I would like to thank my family for their continuous support and sincere care. To my mom, Jun Su, my dad, Hu Dai, and my boyfriend, Dr Weiying Jiang, thank you for your continuous supporting and the fun you brought into my life.

Dedication

To my family

Chapter 1 - Introduction

1.1 General Introduction

Proteins, as the significant functional molecules in life, largely exist and play vital roles in organisms. Their well-organized three dimensional structures are maintained by both intramolecular and intermolecular interaction within themselves or with their surroundings. Alteration of the protein itself or the surrounding environment can disturb the interactions and the structures, which can result in the loss of physical function or diseases, for example, aging, Alzheimer's, Huntington's, Parkinson's, and Amyotrophic lateral sclerosis diseases[1-3]. In physical biochemistry studies, these alterations can be achieved by mutating certain residues within proteins, and by changing in temperature, pH, or adding a cosolvent. Understanding the mechanisms of these structural changes, including protein folding/unfolding, aggregation and translocation, by studying the interactions of protein molecules will be helpful in developing preventions and treatments of these diseases.

Many experimental techniques have been designed to provide protein structural information, such as infrared[4], fluorescence[5], circular dichroism (CD)[6,7], vibrational circular dichroism (VCD)[8,9], dual polarization interferometry[10], and nuclear magnetic resonance (NMR)[11]. Some thermal properties can also be obtained, for example, for protein denaturation, the heat capacity can be monitored by differential scanning calorimeter[12,13], and free energy change can be calculated from the equilibrium constant obtained from structural observations. However, the pathways or mechanisms of protein folding are still unclear, especially at the atomic level, due to their internal complexity. Therefore, computer simulations have been extensively applied to provide detailed information on the relationships between the solution properties and the interactions among the particles. With the rapid and continuing growth of computer power in recent years, more complex data for larger systems can be calculated, compared, and analyzed, and computer simulation has become a useful and indispensable tool in biology, biological chemistry and biophysics.

The Smith group generally focuses on the study of the effects of solvent and cosolvents on the structure and dynamics of biomolecules in solution. In this thesis we describe our efforts to improve simulations in this area. The first goal is to develop an accurate force field for amines

and carboxylic acids, which are commonly found in proteins. The second goal is to evaluate the effect of urea on the aggregation behavior of various molecules. Finally, we estimate the contribution from the vibrational partition function to the free energy changes for several systems and processes, and discuss the inclusion of these contributions in classical force field parameterization.

1.2 Protein Folding/Unfolding and Protein Stability

1.2.1 Changing protein conformational stabilities

The most common protein denaturation methods involve heating or applying high pressure, which changes their physical environment; or adding a cosolvent around a protein or altering the pH of the solution, which changes the chemical environment.

The protein folding/unfolding equilibrium can be considered as a competition between enthalpy favorable processes and entropy favorable processes. Generally, protein folding is described by large negative values of ΔH and ΔS because packing of side chains in the native state is favorable, more hydrogen bonds can be formed in water and the individual water molecules have less freedom of movement.[14] Changes on temperature can alter the proportion of ΔH and −TΔS in ΔG calculation. Therefore, a high temperature will facilitate the protein unfolding process.

It is worthy to mention the cold unfolding process.[14] Compared to simple chemical reactions, the specific heat of protein folding (ΔC_P) is very high, about 12 cal/K/mol per residue.[15] The enthalpy and entropy changes at two different temperatures can be calculated by

$$\Delta H^o_{N \to D}\left(T_2\right) = \Delta H^o_{N \to D}\left(T_1\right) + \Delta C^o_p\left(T_2 - T_1\right) \tag{1.1}$$

$$\Delta S^o_{N \to D}\left(T_2\right) = \Delta S^o_{N \to D}\left(T_1\right) + \Delta C^o_p \ln\frac{T_2}{T_1} \tag{1.2}$$

Therefore, when the temperature decreases, the enthalpy term has a larger decline than the entropy term. So cold unfolding may be observed.

Cosolvent denaturation is very important because of its ability to mimic the protein folding process at physiological temperature and pressure. Cosolvent denaturation displays the following recognized features: 1) the protein unfolding process is reversible without any chemical reaction between the protein and cosolvent; 2) the required concentrations of the

cosolvent are usually high (0.5~10M), which means the interaction between the protein and a cosolvent are weak and nonspecific. More details about cosolvent denaturation will be discussed in Chapter 3.

Protein denaturation can also occur when the pH of the solution is changed to very low or very high values.[14] The primary reason for denaturation is that the pH changes can alter the net charge of the protein, which results in the change of protein stability through charge-charge repulsion or attraction.

1.2.2 The m-value analysis

Denaturants alter the equilibrium between the native (folded) N and denatured (unfolded) D states of a protein, and the stability of the protein is defined as the difference in Gibbs free energy between the denatured and native state[16]:

$$\Delta G^{\circ} = -RT \ln K = -RT \ln \frac{[D]}{[N]} \tag{1.3}$$

The slope of the protein stability with denaturant concentration is called the m-value:[17]

$$\left(\frac{\partial \Delta G^{\circ}}{\partial [\text{denaturant}]} \right) = -m \tag{1.4}$$

Using the m-value concept, the protein folding/unfolding free energy change in denaturant solutions could be described by

$$\Delta G^{\circ} = \Delta G^{\circ}_{H_2O} - m[\text{denaturant}] \tag{1.5}$$

The m-value is an experimentally determined empirical parameter. It assumes a linear relationship between the free energy change and cosolvent concentration,[17] and represents the sensitivity of the protein to chemically induced denaturation. It was also proved to vary linearly with changes in the solvent accessible surface area of a protein.[18] Studies involving experiments with 45 proteins provided expressions for urea and GdmCl can be shown as

$$m_{\text{urea}} = 0.11 + 374 \left(\text{cal} \cdot \text{mol}^{-1} \cdot M^{-1} \cdot \overset{\circ}{A}^{-2} \right) \Delta ASA$$

$$\tag{1.6}$$

$$m_{\text{GdmCl}} = 0.22 + 859 \left(\text{cal} \cdot \text{mol}^{-1} \cdot M^{-1} \cdot \overset{\circ}{A}^{-2} \right) \Delta ASA$$

Here, the solvent accessible surface area change, ΔASA, is in the unit of $\overset{\circ}{A}^2$. These equations have been widely used. For example, with ASA differences calculated from structure changes, the m-value can be estimated and compared to the experiment. Conversely, the m-value can be used to estimate how the structure changes.[19]

However, the m-value approach has some disadvantages. For example, significant deviations from linearity exist, especially in the presence of GdmCl as denaturant.[20,21] A more serious problem is that it is not able to provide specific insights into the denaturation mechanism.

1.2.3 Infrared spectroscopy measurements of protein conformational changes

Infrared (IR) light can be absorbed by molecular vibrations when the frequency of the light is consistent with the frequency of vibration. The frequency of the vibration and the probability of absorption depend on the strength and polarity of the vibrating bonds and are influenced by intramolecular and intermolecular effects.[22] Structural and environmental information concerning bound ligands, amino acid side chains and protein backbones can be obtained from the IR spectrum. Information on protein structure is included in the band position, bandwidth and absorption coefficient parameters. IR spectroscopy has been now a mature technology and has become a valuable tool for the investigation of protein structure.[23-25]

Although amino acid side chains often play key roles in protein structure, only two side chain moieties' absorbance in spectral regions, the SH group of Cys and the carbonyl group of protonated carboxyl groups, can be distinguished from others without further experiments. All other side chain absorptions overlap with those of other side chains or the protein backbone, and further experiment or special treatment, for example, using heavy water as solvent instead of normal water, are needed to assign a band to a specific side chain moiety.[22]

The vibration modes of the amide group include amide A and B, amide I, amide II and amide III bands. The amide A and B vibrations with frequencies over 3000 cm^{-1} are exclusively localized on the NH group and are insensitive to the conformation of the protein backbone.[26] The amide I vibration, absorbing near 1650 cm^{-1}, is mainly from the C=O stretch with a small contribution from the out-of-plane CN stretch, and the CCN deformation and the NH in-plane bend. The amide I vibration is hardly affected by neighboring side chains, but does depend on the secondary structure of the peptide backbone. Therefore, the amide I band is most commonly used for secondary structure analysis. The amide II vibration exists near 1550 cm^{-1} and is the

out-of-phase combination of the NH in-plane bend and the CN stretching vibration with smaller contributions from the CO in-plane bend and the CC and NC stretching vibrations. The amide II vibration is also hardly affected by side chains. However, the correlation between secondary structure and frequency is less obvious than that of amide I vibration.[26] The amide III mode, observed from 1400 to 1200 cm^{-1}, is the in-phase combination of the NH bending and the CN stretching vibration with small contributions from the CO in-plane bending and the CC stretching vibration. The backbone and side chains give strong contributions, and hence the amide II vibration is hardly used for structure analysis.

IR spectroscopy can be widely used to study protein thermal stability because of a large change to the amide I vibrations in the IR spectrum of a protein can be observed for protein transitions between the native state and the denatured or aggregated state. However, for cosolvent protein stability studies, especially with urea or GdmCl, the application of IR spectroscopy is limited due to the overlap of the cosolvent absorption bands with the amide I band of proteins. Using ^{13}C urea is a normal method of avoiding the overlap.[27]

1.3 Molecular Simulation

Molecular simulation is a significant technique which can be used in the area of structural molecular biology. It can be applied to both understand and predict conformational, dynamic and thermodynamic properties of macromolecules based on the knowledge concerning the interactions of their constituent atoms.[28]

Molecular dynamics (MD)[29,30] simulations and Monte Carlo (MC)[31,32] simulations are the two major approaches among a variety of computational techniques used in the world of simulation. In molecular dynamics, atomic motion is simulated by solving Newton's equations of motion simultaneously for all atoms in the system. Compared with Monte Carlo simulation, which generates configurations of a system by making random changes to the positions of the atoms, molecular dynamics reserves all the time-evolution information in the whole trajectory of the simulation, which means not only the thermodynamic or statistical properties, but also the kinetic or dynamic properties are able to be obtained from a MD simulation. This is a remarkable advantage of MD over MC simulation.

Figure 1.1 Simplified flowchart of a typical molecular dynamics simulation[33]

Initial input data:
Interaction function $V\{r\}$: force field
Coordinates r: protein-Protein Data Bank
solvent-random insertion
Velocities v: Boltzmann distribution

Compute potential $V\{r\}$ and
Forces $F_i = -\nabla_i V\{r\}$ on atoms

Update coordinates & velocities
according to equations of motion

Solve constraints
Couple to temperature and pressure bath
-if necessary

Collect statistics and write
energy/coordinates to trajectory files

More steps?

Repeat for millions of steps

Yes

No

Done!

A simplified flowchart scheme for MD simulation is given in Figure 1.1. In molecular dynamics simulations[34], the force on any atom is obtained as

$$F_i = -\frac{\partial V}{\partial r_i}$$ (1.7)

It is computed by summing the forces between nonbonded atom pair involving atom i:

$$F_i = \sum_j F_{ij}$$ (1.8)

together with any restraining and/or external forces if they are needed. The calculation of bonded and nonbonded interactions is described in detail in section 1.4.1. Newton's equations of motion provide the following relationships,

$$\frac{dr_i}{dt} = v_i ; \quad \frac{d^2 r_i}{dt^2} = \frac{F_i}{m_i} = a_i$$ (1.9)

There are many algorithms used to integrate the equations of motion with finite difference methods. The Verlet algorithm is probably the most widely used method for integrating the equations of motion in a molecular dynamics simulation. The Verlet algorithm assumes the new positions can be approximated as a Taylor series expansion for each atom:

$$r(t+\Delta t) = r(t) + v(t)\cdot\Delta t + \frac{1}{2}a(t)\cdot\Delta t^2 + \frac{1}{6}b(t)\cdot\Delta t^3 + \cdots$$

$$r(t-\Delta t) = r(t) - v(t)\cdot\Delta t + \frac{1}{2}a(t)\cdot\Delta t^2 - \frac{1}{6}b(t)\cdot\Delta t^3 + \cdots$$ (1.10)

By adding the two equations in (1.10) we find that the positions and accelerations at time t and from previous step t-Δt can be used to calculate the new positions at t+Δt.

$$r(t+\Delta t) = 2r(t) - r(t-\Delta t) + a(t)\Delta t^2$$ (1.11)

The velocities can be calculated from the difference in positions at times t+Δt and t-Δt
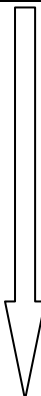
$$v(t) = \left[r(t+\Delta t) - r(t-\Delta t)\right]/(2\Delta t)$$ (1.12)

The value of Δt is known as the time step. A range of biomolecular dynamic motions happen on different time scales and are listed in Table 1.1. In order to obtain a longer time step, the intramolecular bond vibrations are neglected in MD simulation (the errors resulting from this are evaluated in Chapter 4), and the individual time step of MD simulation can be increased from 1 to 2 fs. Therefore, the information relating to bond stretching vibrations is missing. Using

parallel computers makes it possible to perform simulations on a microsecond scale. Some computing techniques, for example replica exchange, or implicit solvent approaches, can enhance sampling capability, improve efficiency and make the simulation of normal protein or even membrane protein feasible.

Although molecular dynamics simulations are performed with several approximations, for instance, all the atoms are considered classical, classical MD simulations still provide useful information concerning biological systems at the molecular level, [35,36] sometimes even beyond that of experimental techniques. Considering the technical obstacles arising in traditional laboratories involving extraction, purification, and preserving 3D structures of proteins, MD simulation is a powerful tool to understand proteins and their behavior.

Table 1.1 Range of time scales for dynamic behavior in biomolecular systems.[33]

Time Scale	Molecular Activity	Simulation Accessibility
10^{-15}s	Bond and Angle Vibration	
	Torsion and Libration	Accessible to
10^{-12}s	Rotation around bonds	atomic-detail
	Water Relaxation	simulation
10^{-9}s	Lipid Rotation	today
	Transport in Ion Channels	
10^{-6}s	Lipid Diffusion	
	Rapid Protein Folding	
10^{-3}s	Normal Protein Folding	
1s	Ribosome Synthesis	
10^{3}s	Membrane Protein Folding	

1.4 Force Fields for Biological Simulations

1.4.1 Common force fields and potential energy functions

The term "force field" in molecular dynamics simulations refers to the combination of mathematical equations and associated parameters which are used to calculate the energy of the system as a function of its atomic coordinates. With the development of increasing computer performance, and the desire for studies on biological systems, various biomolecular protein force fields have been established and improved, for instance, Amber[37], CHARMM[38], GROMOS[39], OPLS-AA[40], ENCAD[41,42], ECEPP[43-46] and UNRES[47]. All the biomolecular force fields are classical, so the electronic degrees of freedom are ignored. Additivity and transferability are two additional assumptions involved in common force fields. The former means that a large molecule can be considered as a sum of various groups. The latter means that potential energy functions developed on a relatively small set of molecules can be applied to a wide range of molecules or environments with similar chemical groups, rather than defining a new set of parameters for each individual molecule or different system. The validity of these two assumptions is assured for all biomolecular force fields.

The total potential energy in most biomolecular force fields can be expressed as a sum of terms

$$V_{potential} = V_{bonds} + V_{angles} + V_{torsions} + V_{improper} + V_{elec} + V_{LJ}$$

$$V_{bond} = \sum_{bonds} \frac{1}{2} k_b \left(r - r_0 \right)^2$$

$$V_{angles} = \sum_{angles} \frac{1}{2} k_\theta \left(\theta - \theta_0 \right)^2$$

$$V_{torsions} = \sum_{torsions} k_\phi \left[1 + \cos \left(n\phi - \phi_s \right) \right] \tag{1.13}$$

$$V_{improper} = \sum_{improper} \frac{1}{2} k_\xi \left(\xi - \xi_0 \right)^2$$

$$V_{elec} = \sum_{elec} \frac{q_i q_j}{r_{ij}}$$

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

Here, the first four terms correspond to the bonded energy, and the last two terms correspond to the nonbonded energy. In the bonded energy terms, especially the bond stretching and angle bending energies, the interaction forces are assumed to be harmonic, and the force constants and equilibrium constants can be derived from experimental data such as gas-phase geometries and vibrational spectra. More refined force fields have a cubic term[48], a quadratic function[49-51], or a Morse function[52] included. However, because of computational inefficiency, they are rarely used in biomolecular studies. The torsional energy surface, which is calculated from the torsional rotation and improper torsion potential equations, are typically supplemented with ab initio results.

The last two nonbonded terms are calculated for pairs of atoms separated by three or more bonds and between atoms in different molecules. The electrostatic interaction is calculated as a sum of interactions between pairs of partial (or effective) atomic charges using Coulomb's law, and the van der Waals interaction is calculated using the Lennard-Jones (LJ) 12-6 potential. It includes a power 12 repulsive term, which is due to the Pauli exclusion principle, and a power 6 attraction term, which due to the dispersion forces generated between instantaneous dipoles arising from fluctuations in electronic charge distributions in all molecules. The constant ε defines the depth of the interaction minimum and σ defines the distance at which the LJ energy is zero.

1.4.2 Force field parameterization

At the beginning of a force field parameterization it is usually necessary to assign an atom type to each atom in the system. The atom type is not only the atomic number of an atom, but also contains information about its hybridization state and sometimes even the local environment. All of the force field parameters depend on the atom types. Since the parameterization is a computationally demanding and labor-intensive process, it is sensible to spend more time on optimizing the parameters, such as nonbonded terms and torsional terms, which are sensitive to the performance of the force field, rather than parameters which do not have a great influence on the simulation results, for example, the bond-stretching and angle-bending terms.

The first step is to select the target data used to guide the parameterization. The LJ parameters and Coulomb partial atomic charge parameters are adjusted to reproduce different sources of data, including molecular volumes, experimental enthalpy of mixing, compressibility, density, and ab initio/quantum mechanics (QM) calculated dipole moments values. For example, the partial charges in OPLS-AA[40] were optimized with LJ parameters to fit the experimental heats of vaporization and densities, GROMOS[53] aimed to reproduce thermodynamic data including solvation free energies, while the Amber force field[54] determined partial charges to target the computed QM electrostatic potential surface. Generally, the QM calculated values are used because there is not enough experimental data available. However, due to its high expense, the QM calculations are limited to the molecules in gas phase which display some deviations from those in the condensed liquid phase, and in some ways, are inappropriate to be used to parameterize. Therefore, some more convincing thermodynamic property data are often required.

One of the commonly used protocols is to first establish a set of van der Waals parameters, and then determine the electrostatic model by using electrostatic potential fitting. Finally, the torsional potentials are determined by reproducing the experimental conformational properties. However, force field development is an iterative self-consistent process. Changes in nonbonded parameters affect not only intermolecular interactions, which characterizes condensed-phase properties, but also intramolecular interactions, which characterized in conformational energies, consequently requiring adjustment of dihedral parameters. On the other hand, the dihedral parameters affect intramolecular energy, which determines the preferred intramolecular geometries and molecular volumes, thereby requiring the validation or adjustment of nonbonded parameters. In principle, at any stage certain parameters are changed, it is necessary to check the ability of the force field to reproduce all of the target data and update some other parameters accordingly. This is the main reason why force field parameterization is a complicated and time-consuming procedure.

1.4.3 Hydrogen bonding and the hydrophobic effect

Hydrogen bonding is a local, directional, and strong (>20kJ/mol for isolated bonds)[14] interaction, which makes a significant contribution to the protein and solvent energy and is a major factor in determining protein structure. In order to improve the accuracy with which the geometry of hydrogen-bonding systems is predicted, some force fields have replaced the LJ 12-6

terms between hydrogen-bonding atoms by an explicit hydrogen-bonding term, for example a LJ 12-10 term, to described the interaction between the donor hydrogen atom and the heteroatom acceptor atom[55]. However, in most of the biomolecular force fields, hydrogen-bonding are incorporated via the electrostatic and van der Waals terms instead of an explicit term. The hydrogen bond energy in these models is largely due to the interaction between a dipole formed by the donor proton and bound electronegative atom on one side of the hydrogen bond and an aligned dipole formed by the electronegative acceptor and bound atom on the other side. Most biomolecular force fields, in which interaction sites for bonded and nonbonded interactions are located at the atomic nuclei, are generally very good at being able to reproduce hydrogen bond energies and geometries, although they can lead to deviations from QM results for the angular dependence of the hydrogen bonding energy in certain cases.[33] Including additional interaction sites at the position of lone pairs, for instance the sulfur atoms in early Amber force field and the oxygen in TIP5P water model[56], can lead to improved thermodynamics and structural properties. Nonetheless, because of the computational expense, current biochemistry simulations generally use three-particle water models and protein force fields without lone pair positions.

The hydrophobic effect plays an important role in protein folding and aggregation events. The hydrophobic interaction between molecules is a complex process which results primarily from entropic effects involving the change in the orientation of solvent molecules in the solvation shell surrounding solute molecules, and also from the bulk solvent molecules.[105] Therefore, in commonly used biomolecular force fields, the hydrophobic interaction is implicitly included in the nonbonded terms of potential energy functions and the inclusion of explicit solvent molecules. Several attempts have been made to quantify the hydrophobicity, for example, the HINT model[57] of hydrophobic interactions generated a hydrophobic field in analogy to the electrostatic field based on the solubility data, and program MOLCAD[58] projects the hydrophobic and hydrophilic property distributions in the extramolecular space onto a molecular surface by introducing a molecular lipophilicity potential, which can be regarded as a pendant to the molecular electrostatic potential.

1.5 Kirkwood-Buff Theory

Kirkwood and Buff developed a new solution theory in 1951[59]. The basis of Kirkwood-Buff (KB) theory is the relation between the number fluctuations in an open system and the thermodynamic properties of that system. Some expanded derivations[60,61] of KB theory have been provided since then. In 1970's, when the radial distribution functions were available from computer simulations, and Ben-Naim[62] illustrated how to use KB theory to analyze experimental thermodynamic properties on solution mixture, it began to be extensively used in the chemistry and chemical engineering fields. It provides a bridge which connects a real system and a model system for computer simulation as shown in Figure 1.3. The specific remarkable advantages of KB theory include:[63] 1) it is an exact theory which is more valid than other theories, e.g. McMillan-Mayer theory. 2) It can be applied to any molecules of various sizes and complexity, and any stable solution mixture involving any number of components. 3) No assumption of pairwise-additive interactions was made in it. 4) It is very suitable for the analysis of computer simulation data.

Figure 1.2 The role of Kirkwood-Buff theory in both experimental and simulated data.
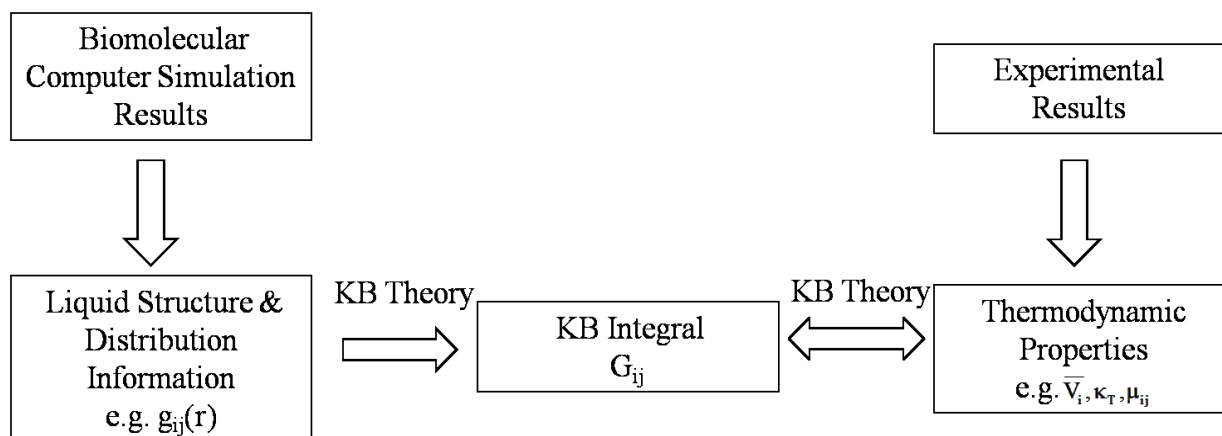
The radial distribution function[64] (rdf) g_{ij}, which is typically calculated from the saved particle coordinates, measures the relative probability of finding an atom at a distance r away from a central atom, and plays a central role in KB theory.

The rdf can be obtained from both computer simulations. In a canonical ensemble of a volume V and a temperature T with N particles, according to Boltzmann distribution, the probability of finding particle 1 in dr_1 at r_1, and particle 2 in dr_2 at r_2, etc., can be represented as

$$P(r_1, r_2, \cdots, r_N) dr_1 \cdots dr_N = \left(e^{-\beta V_N} dr_1 dr_2 \cdots dr_N \right) \Big/ Z_N \tag{1.14}$$

where $\beta = 1/kT$, V_N is N-particle the potential energy and Z_N is the configurational integral.

For a subset of n particles in the N-particle systems, the probability distribution for these n particles can be expressed as

$$P^{(n)}(r_1, r_2, \cdots, r_n) = \frac{\int\int \cdots \int e^{-\beta V_N} dr_{n+1} \cdots dr_N}{Z_N} \tag{1.15}$$

As a result, the probability of finding any particle at distance dr_1, and any particle at distance dr_2, etc., can be expressed as

$$\rho^{(n)}(r_1, r_2, \cdots, r_n) = \frac{N!}{(N-n)!} P^{(n)}(r_1, r_2, \cdots, r_n) \tag{1.16}$$

The probability of finding a particle anywhere in a homogeneous system could be written as

$$\frac{1}{V} \int \rho^{(1)}(r_1) dr_1 = \rho^{(1)} = \frac{N}{V} = \rho \tag{1.17}$$

Consequently, a series of correlation functions, $g^{(n)}\{r\}$, can be defined as

$$\rho^{(n)}(r_1, \cdots r_n) = \rho^n g^{(n)}(r_1, \cdots r_n) \tag{1.18}$$

Hence,

$$g^{(n)}(r_1, r_2, \cdots, r_n) = \frac{V^n N!}{N^n (N-n)!} \frac{\int\int \cdots \int e^{-\beta U_N} dr_{n+1} \cdots dr_N}{Z_N}$$
$$= \left[V^n \left(1 + O(N^{-1}) \right) \int \cdots \int e^{-\beta U} dr_{n+1} \cdots d_N \right] \Big/ Z_N \tag{1.19}$$

For two particles, $g^{(2)}(r_1, r_2)$, known as the pair distribution function, can be obtained from experiment data or simulation results. Using the distance between two particles, the radial distribution function, $g^{(2)}(r_{12})$, can be used to calculate the KB integrals.

Let us consider the 2-Aminoethanol and water radial distribution function (rdf) as an example (see details in Chapter 2) displayed in Figure 1.3 top. Due to the strong repulsion force between two molecules within very short distances, the rdf starts from 0. It then proceeds through a series of fluctuations around unity, which are generally known as solvation shells. The

first and also the highest peak indicates the first solvation shell, where it is most likely to find a molecule compared to other distances. With an increase in the distance r, the rdf approaches unity beyond 2.0 nm, indicating close to a bulk solution random distribution.

The KB integral (G_{ij}) between species i and j, which is the primary quantity of interest in KB theory, can be expressed as

$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}^{\mu VT}(r) - 1 \right] r^2 dr \qquad (1.20)$$

where the g_{ij} is the corresponding center of mass (COM) based radial distribution function in the μVT ensemble, and r is the COM-COM distance. The KB integrals can be calculated from simulation results, which are usually in NpT Ensemble, with the assumption that

$$G_{ij}(R) = 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1 \right] r^2 dr \qquad (1.21)$$

Here, R defines a region within which the molecular distribution differs from the bulk solution random distribution. Previous studies have showed that a combination of KB theory and NpT simulations are able to provide quantitative information of thermodynamics for solutions.[70-73] An example of a typical KB integral G_{ij} as a function of integration distance R is displayed in Figure 1.3 bottom. A positive G_{ij} corresponds to an excess of species j in the vicinity of species i over a random distribution, which means a favorable (attractive) net interaction between species i and j; while a negative G_{ij} indicates a depletion of species j surrounding i, which means an unfavorable (repulsive) net interaction.

The excess coordination numbers are defined and calculated from the KB integrals

$$N_{ij} = \rho_j G_{ij} \qquad (1.22)$$

where ρ_j is the number density of species j giving by

$$\rho_j = {N_j}/{V} \qquad (1.23)$$

A positive N_{ij}, value which is related to favorable (attractive) interactions, corresponds an excess number of species j in the vicinity of species i over a random distribution, while a negative N_{ij} value, which is interpreted as unfavorable (repulsive) interactions, indicates a depleted number of species j surrounding i. Figure 1.4 shows the excess coordination numbers obtained for 2-Aminoethanol and water mixtures (see Chapter 2 for details). The excess coordination numbers

change with solution composition and provide quantitative information on the relative distributions of components in solution.

Figure 1.3 Radial distribution function (g_{ij}) (top) and KB integral (G_{ij}, cm^3/mol) (bottom) between 2-Aminoethanol and water in a 2-Aminoethanol and water system with the 2-Aminoethanol mole fraction of 0.5 at 300K and 1atm
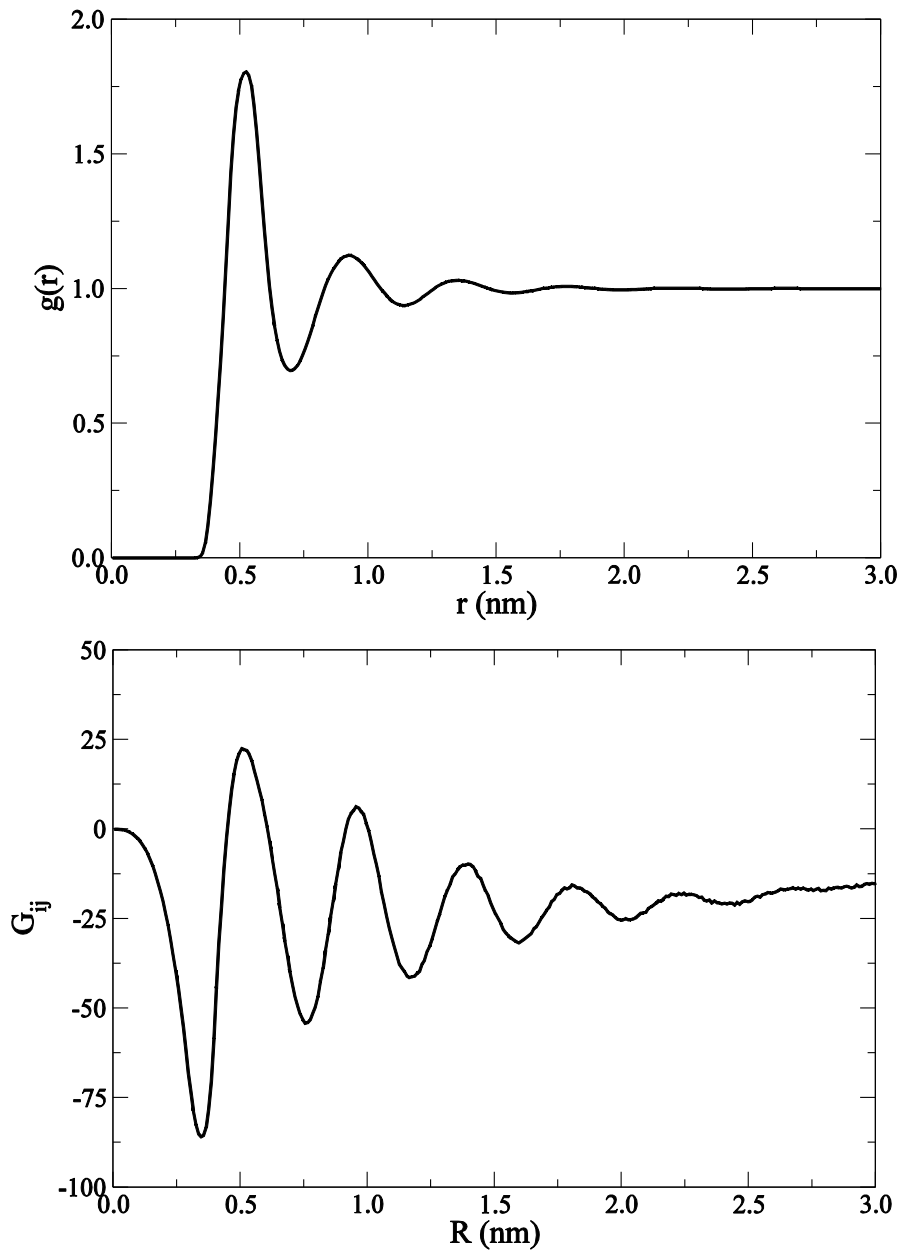
Due to the capability of providing a quantitative description of molecular distributions and thermodynamic properties from KB integrals (see equations in Chapter 2 and 3), more and more chemists and physicists, including Smith[65-79], Ruckenstein[80-87], etc, have been developing KB theory and applying KB theory to study the properties of solution mixtures. For example, Rosgen et al. used KB theory to study molecular crowding effects on macromolecules and small molecules and to understand the structural thermodynamics of protein preferential solvation[88]. Lenhoff and coworkers have applied KB theory to explain experimental results and to understand the concentration dependence of partial specific volumes of proteins in aqueous solution[89]. Matubayasi et al. have applied KB theory to calculate the free energy of molecular binding into lipid membranes[90] and characterize the preferential interactions in bovine serum albumin in the presence of salts in various concentrations[91]. Hirata and coworkers used KB theory to calculate the partial molar volume changes associated with coil-to-helix transition of peptides[92,93]. Some other examples using KB theory to study the cosolvent effect on protein folding are listed in Chapter 3.

Figure 1.4 Excess coordination numbers N_{ij} of 2-Aminoethanol and water system over the entire composition range using available experimental data at 300K and 1atm[94, 95]. The subscript 1 refers to the solvent water, and 2 refers to solute 2-Aminoethanol.
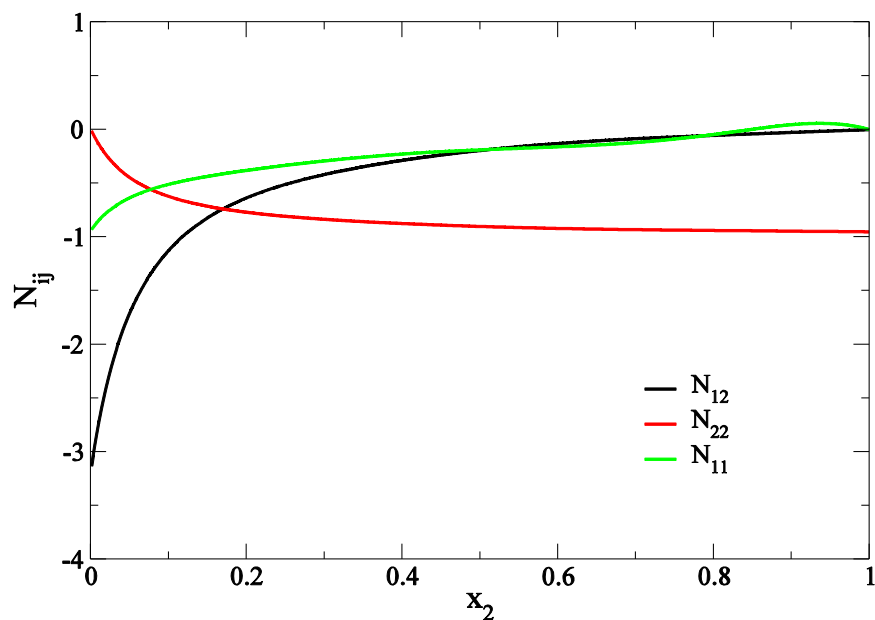
The KB theory has only two major limitations, which involve converge problems and chemical potential calculation. An appropriate system size (>5 nm) and long enough sampling time are needed to allow the KB integrals converge with time and become unity in the bulk area. The changes of chemical potential with solution composition can be calculated from KB theory. However, the absolute chemical potential value is not directly available.

1.6 Kirkwood-Buff Derived Force Fields

Because KB theory can play the role of a bridge between simulated data and experimental results, the quality of a force field in computer simulation can be evaluated by comparing the KB integrals and thermodynamic properties obtained from simulations to the KB integrals and thermodynamic properties from experiments. In addition, it is well known that the KB integrals are much more sensitive to the force field,[67,72] especially the effective atomic charge parameters, than many other thermodynamic experimental data. Several tests using commonly available force fields showed problems in reproducing KB integrals from experiments,[77] which indicated that it is necessary to build a optimized force field. Since the KB integral provide quantitative information concerning the interaction between a pair of components in solutions, it can be used as a target property to be reproduced in the parameterization procedure of a new force field.

The development of the Kirkwood-Buff force field (KBFF) can be traced back to 2003 and a list of recent publications for KBFF is shown in Table 1.2. The KBFF models are simple nonpolarizable classical force fields. The general approach in parameterization is to optimize the effective charge distributions on solute molecules, and the primary aim is to mimic condensed phase polarization effects by reproducing the KB integrals in the whole composition range. The details of an amines and carboxylic acid parameterization are shown in Chapter 2. Generally, the most effective charges on atoms can be determined with the help of the KB integrals, and the resulting force fields display a better performance than other common nonpolarizable force fields at the same level of computational expense[70-72,96,97]. Other parameters are similar to most existing force fields as discussed in section 1.4. For example, the bond parameters parameters are taken from the GROMOS96[98]. Since it is observed that some bias or limitation in reproducing the appropriate secondary structure of peptides or proteins has existed in common force fields[99-101],

the torsional parameters are also modified in KBFF in order to provide accurate protein backbone Φ/Ψ potentials to represent the correct protein conformational distributions[102].

Table 1.2 KBFF Models developed to date

Solute	Relevant species in biomolecules
Acetone[71]	Cosolvent
Urea[72]	Cosolvent
GdmCl[96]	Cosolvent, Arg
NMA[77]	Asn, Gln, Peptide backbone
Methanol[97]	Ser, Thr, Tyr
Thiols, Sulfides[65]	Met, Cys, disulfide
Alkali halides[67,70]	Cosolvent
Pyrrole, Pyridine, Histidine[76]	Phe, Tyr, His, Trp
Amines, Carboxylic acid	Lys, Asp, Glu, termini
Methyl acetate, Dimethyl Phosphate, Glycerol*	Phosphates, lipids

* in progress

Some of the experiences or observations we learned in the development of KBFF parameters include[74]: the KBFF is able to reproduce not only the experimental KB integrals, but also the traditional solution properties (e.g. diffusion constant, relative permittivity) as well[72]; the KB integrals from both the experiments and simulations display significant uncertainty at the extremely low solute or solvent concentration; the properties of a solute and solvent in binary solutions affect the interactions between the solute and additional solutes (cosolvent) in ternary systems.[73,78,103]

1.7 Summary and Outline

Computer simulations, molecular dynamics simulations in particular, have been extensively used in studying biological systems by providing information at the atomic level.

Kirkwood-Buff theory can be a helpful tool to interpret the thermodynamics and other properties of interested systems and quantify bulk solution and interfacial properties. Our long term aim is to establish a full biomolecular force field and to use it to study protein folding/unfolding or aggregation processes using KB theory, or fluctuation theory[75,104], which is developed based on KB theory.

In Chapter 2, KB theory is applied for the parameterization of a united atom nonpolarizable force field for amines and carboxylic acids suitable for modeling uncharged N-termini and lysine residues, and the protonated states for the C-terminus, and both aspartic and glutamic acids, respectively. The parameters were developed to reproduce the experimental KB integrals as a function of solute mole fractions. Some other thermodynamic and physical properties of these solutions, for example, density, dielectric constants and diffusion coefficients, are also calculated.

In Chapter 3, four ternary systems involving various solute molecules in urea aqueous solution are investigated using Kirkwood-Buff theory. Neo-pentane, benzene, glycine and methanol are chosen to represent different moieties in proteins, and are studied with the presence of urea, which is considered as a protein denaturant, over a wide solute and denaturant concentration range. KB theory is used to calculate the transfer free energy and preferential interactions on the molarity scale in order to understand the mechanism of urea cosolvent effect on proteins at the atomic level.

The contribution from the vibrational partition function is investigated in Chapter 4. The vibrational partition functions are estimated to give a remarkablely large contribution for several processes in solution, including the free energy of solvation, protein folding, the enthalpy of mixing, and the enthalpy of vaporization. It is argued that these contributions have to be implicitly included in the classical force field parameterization.

In Chapter 5, a brief summary and conclusion of present work is drawn. A short to-do list with a description of future improvements is also provided.

1.8 References

(1)   Batchelor, J. D.; Olteanu, A.; Tripathy, A.; Pielak, G. J. *J. Am. Chem. Soc.* **2004**, *126*, 1958.

(2)   Creighton, T. E. *Proteins, Structures, and Molecular Properties*; W. H. Freeman and Company, 1996.

(3)   Truant, R.; Atwal, R. S.; Desmond, C.; Munsie, L.; Tran, T. *Febs. J.* **2008**, *275*, 4252.

(4)   Hamm, P.; Lim, M. H.; Hochstrasser, R. M. *J. Phys. Chem. B* **1998**, *102*, 6123.

(5)   Cremer, C.; Cremer, T. *Microsc. Acta.* **1978**, *81*, 31.

(6)   Whitmore, L.; Wallace, B. A. *Biopolymers.* **2008**, *89*, 392.

(7)   Greenfield, N. J. *Nat. Protoc.* **2006**, *1*, 2876.

(8)   Malon, P.; Kobrinskaya, R.; Keiderling, T. A. *Biopolymers* **1988**, *27*, 733.

(9)   Yasui, S. C.; Keiderling, T. A. *Mikrochim. Acta.* **1988**, *2*, 325.

(10)   Boudjemline, A.; Clarke, D. T.; Freeman, N. J.; Nicholson, J. M.; Jones, G. R. *J. Appl. Crystallogr.* **2008**, *41*, 523.

(11)   Wuthrich, K. *J. Biol. Chem.* **1990**, *265*, 22059.

(12)   Attri, P.; Venkatesu, P.; Lee, M. J. *J. Phys. Chem. B.* **2010**, *114*, 1471.

(13)   Shah, M. A.; Mishra, S.; Chaudhuri, T. K. *Eur. Biophys. J .Biophy.* **2011**, *40*, 877.

(14)   Fersht, A. *Structure and Mechanism in Protein Science*; W. H. Freeman and Company, 1997.

(15)   Privalov, P. L.; Makhatadze, G. I. *J. Mol. Biol.* **1990**, *213*, 385.

(16)   Holthauzen, L. M. F.; Auton, M.; Sinev, M.; Rosgen, J. *Methods in Enzymology* **2011**, *492*, 61.

(17)   Greene, R. F.; Pace, C. N. *J. Biol. Chem.* **1974**, *249*, 5388.

(18)   Myers, J. K.; Pace, C. N.; Scholtz, J. M. *Protein. Sci.* **1995**, *4*, 2138.

(19)   Scholtz, J. M.; Grimsley, G. R.; Pace, C. N. *Methods in Enzymology* **2009**, *466*, 549.

(20)   Pace, C. N. *Methods Enzymol.* **1986**, *131*, 266.

(21)   Makhatadze, G. I. *J. Phys. Chem. B* **1999**, *103*, 4781.

(22)   Barth, A.; Zscherp, C. *Q Rev. Biophys.* **2002**, *35*, 369.

(23)   Arrondo, J. L. R.; Muga, A.; Castresana, J.; Goni, F. M. *Prog. Biophys. Mol. Bio.* **1993**, *59*, 23.

(24)   Jackson, M.; Mantsch, H. H. *Crit. Rev. Biochem. Mol.* **1995**, *30*, 95.

(25)   Arrondo, J. L. R.; Goni, F. M. *Prog. Biophys. Mol. Bio.* **1999**, *72*, 367.

(26)   Krimm, S.; Bandekar, J. *Adv. Protein. Chem.* **1986**, *38*, 181.

(27)   Fabian, H.; Mantsch, H. H. *Biochemistry* **1995**, *34*, 13651.

(28)   Goodfellow, J. M. *Computer Modelling in Molecular Biology*; VCH, 1995.

(29)   Alder, B. J.; Wainwright, T. E. *J. Chem. Phys.* **1959**, *31*, 459.

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

(31)   Metropolis, N.; Ulam, S. *J. Am. Stat. Assoc.* **1949**, *44*, 335.

(32)   Mosegaard, K.; Tarantola, A. *J Geophys. Res-Sol. Ea.* **1995**, *100*, 12431.

(33)   Kukol, A. *Molecular Modeling of Protein*; Humana Press, 2008.

(34)   http://www.gromacs.org/Documentation/Manual: 2009.

(35)   Shea, J. E.; Brooks, C. L. *Annu. Rev. Phys. Chem.* **2001**, *52*, 499.

(36)   Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 788.

(37)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.

(38)   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. *J. Phys. Chem. B* **1998**, *102*, 3586.

(39)   Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656.

(40)   Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(41)   Levitt, M. *J Mol Biol* **1983**, *168*, 595.

(42)   Levitt, M.; Hirshberg, M.; Sharon, R.; Daggett, V. *Comput. Phys. Commun.* **1995**, *91*, 215.

(43)   Momany, F. A.; Carruthe.Lm; Mcguire, R. F.; Scheraga, H. A. *J. Phys. Chem.* **1974**, *78*, 1595.

(44)   Momany, F. A.; Mcguire, R. F.; Burgess, A. W.; Scheraga, H. A. *J. Phys. Chem.* **1975**, *79*, 2361.

(45)   Nemethy, G.; Pottle, M. S.; Scheraga, H. A. *J. Phys. Chem.-Us* **1983**, *87*, 1883.

(46)   Nemethy, G.; Gibson, K. D.; Palmer, K. A.; Yoon, C. N.; Paterlini, G.; Zagari, A.; Rumsey, S.; Scheraga, H. A. *J. Phys. Chem.-Us* **1992**, *96*, 6472.

(47)   Oldziej, S.; Czaplewski, C.; Liwo, A.; Chinchio, M.; Nanias, M.; Vila, J. A.; Khalili, M.; Arnautova, Y. A.; Jagielska, A.; Makowski, M.; Schafroth, H. D.; Kazmierkiewicz, R.; Ripoll, D. R.; Pillardy, J.; Saunders, J. A.; Kang, Y. K.; Gibson, K. D.; Scheraga, H. A. *P Natl. Acad. Sci. USA* **2005**, *102*, 7547.

(48)   Allinger, N. L. *J. Am. Chem. Soc.* **1977**, *99*, 8127.

(49)   Allinger, N. L.; Yuh, Y. H.; Lii, J. H. *J. Am. Chem. Soc.* **1989**, *111*, 8551.

(50)   Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8566.

(51)   Lii, J. H.; Allinger, N. L. *J. Am. Chem. Soc.* **1989**, *111*, 8576.

(52)   Morse, P. M. *Phys. Rev.* **1929**, *34*, 57.

(53)   Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. *Eur. Biophys. J. Biophy.* **2011**, *40*, 843.

(54)   Weiner, S. J.; Kollman, P. A.; Case, D. A.; Singh, U. C.; Ghio, C.; Alagona, G.; Profeta, S.; Weiner, P. *J. Am. Chem. Soc.* **1984**, *106*, 765.

(55)   Leach, A. R. *Molecular Modelling Principles and Applications*; Prentice Hall, 2001.

(56)   Mahoney, M. W.; Jorgensen, W. L. *J. Chem. Phys.* **2000**, *112*, 8910.

(57)   Kellogg, G. E.; Semus, S. F.; Abraham, D. J. *J. Comput. Aid. Mol. Des.* **1991**, *5*, 545.

(58)   Heiden, W.; Moeckel, G.; Brickmann, J. *J. Comput. Aid. Mol. Des.* **1993**, *7*, 503.

(59)   Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

(60)   Hall, D. G. *T. Faraday. Soc.* **1971**, *67*, 2516.

(61)   Oconnell, J. P. *Mol. Phys.* **1971**, *20*, 27.

(62)   Bennaim, A. *J. Chem. Phys.* **1977**, *67*, 4884.

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

(64)   McQuarrie, D. A. *Statistical Mechanics*; Harper&Row, 1976.

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

(66)   Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2002**, *106*, 1491.

(67)   Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011**, *7*, 1369.

(68)   Gee, M. B.; Smith, P. E. *Abstr. Pap. Am. Chem. S.* **2008**, *236*.

(69)   Gee, M. B.; Smith, P. E. *J. Chem. Phys.* **2009**, *131*.

(70)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342.

(71)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 10663.

(72)   Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2003**, *107*, 3891.

(73)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 5901.

(74)   Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilibr.* **2010**, *290*, 43.

(75)   Ploetz, E. A.; Smith, P. E. *J. Chem. Phys.* **2011**, *135*.

(76)   Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18154.

(77)   Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477.

(78)   Kang, M.; Smith, P. E. *Fluid Phase Equilibr.* **2007**, *256*, 14.

(79)   Kang, M.; Smith, P. E. *J. Chem. Phys.* **2008**, *128*.

(80)   Ruckenstein, E.; Shulgin, I. *Fluid Phase Equilibr.* **2001**, *180*, 281.

(81)   Shulgin, I.; Ruckenstein, E. *Ind. Eng. Chem. Res.* **2002**, *41*, 6279.

(82)   Shulgin, I.; Ruckenstein, E. *Polymer.* **2003**, *44*, 901.

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

(84)   Shulgin, I. L.; Ruckenstein, E. *J. Phys. Chem. B* **2006**, *110*, 12707.

(85)   Shulgin, I. L.; Ruckenstein, E. *J. Phys. Chem. B* **2007**, *111*, 3990.

(86)   Shulgin, I. L.; Ruckenstein, E. *Fluid Phase Equilibr.* **2007**, *260*, 126.

(87)   Shulgin, I. L.; Ruckenstein, E. *J. Phys. Chem. B* **2008**, *112*, 3005.

(88)   Auton, M.; Bolen, D. W.; Rosgen, J. *Proteins* **2008**, *73*, 802.

(89)   Pjura, P. E.; Paulaitis, M. E.; Lenhoff, A. M. *Aiche. J.* **1995**, *41*, 1005.

(90)   Matubayasi, N.; Shinoda, W.; Nakahara, M. *J. Chem. Phys.* **2008**, *128*.

(91)   Shimizu, S.; Matubayasi, N. *Chem. Phys. Lett.* **2006**, *420*, 518.

(92)   Imai, T.; Harano, Y.; Kovalenko, A.; Hirata, F. *Biopolymers* **2001**, *59*, 512.

(93)   Harano, Y.; Imai, T.; Kovalenko, A.; Kinoshita, M.; Hirata, F. *J. Chem. Phys.* **2001**, *114*, 9506.

(94)   Page, M.; Huot, J. Y.; Jolicoeur, C. *Can. J. Chem.* **1993**, *71*, 1064.

(95)   Touhara, H.; Okazaki, S.; Okino, F.; Tanaka, H.; Ikari, K.; Nakanishi, K. *J. Chem. Thermodyn.* **1982**, *14*, 145.

(96)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2004**, *121*, 2180.

(97)   Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005**, *109*, 15080.

(98)   Daura, X.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **1998**, *19*, 535.

(99)   Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712.

(100)   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.

(101)   Best, R. B.; Buchete, N. V.; Hummer, G. *Biophys. J.* **2008**, *95*, 4494.

(102)   Jiao, Y. F.; Chen, F.; Smith, P. E. *Abstr. Pap. Am. Chem. S.* **2012**, *243*.

(103)   Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2001**, *105*, 11513.

(104)   Smith, P. E.; Matteoli, E.; OConnel, J. P. *Fluctuation Theory of Solutions*; CRC Press, **2013**.

(105)   Holtje, H. D.;Sipple, W.; Rognan, D.; Folkers, G. *Molecular Modeling, Basic Principles and Applications*; Wiley-VCH, **1996**.

Chapter 2 - Kirkwood-Buff Derived Force Fields for Amines and Carboxylic Acid

2.1 Abstract

In our attempts to derive a extended full protein force field based on the Kirkwood-Buff (KB) theory of solutions, we present a force field for amines and carboxylic acids suitable for modeling uncharged N-terminus and lysine residues, the protonated states for the C-terminus, and both aspartic and glutamic acids, respectively. The parameters were developed to reproduce the composition dependent KB integrals obtained from an analysis of the experimental data (activities, densities, and compressibilities) for a variety of binary solution mixtures. The binary mixtures studies include: n-butylamine with both ethanol and n-propanol, 2-aminoethanol with water, and acetic acid with both methanol and benzene. The KB theory of solutions provides exact expressions which relate the thermodynamics of a solution mixture to the underlying distribution between the different species in solution. Other thermodynamic and physical properties of these solutions, including the density, dielectric constants, diffusion coefficient were also calculated.

2.2 Introduction

The accuracy of a protein force field is the foundation of a successful simulation of large biomolecular systems. Therefore, the development and improvement of force field are very significant. Several protein force fields have been proposed in the last ten years,[1-3] and the non-polarizable force fields are dominant because of their high efficiency compared to the polarizable ones. However, they have some problems, for example, overestimation of solute-solute and solvent-solvent interactions,[4-6] which could induce more serious errors of preferential interactions[7-9] or other properties that have no available experimental data to compare with. And if being simulated for long enough time, aggregation or phase separation could happen spontaneously in fully miscible systems. [4,10,11]

The approximate description of electrostatic interactions is considered as one of the origins of these misbalanced interactions. The Coulomb partial atomic charge parameters, which are used to calculate the electrostatic interactions, are parts of the contributions to the heat of

vaporizations and densities,[12] and for this reason, these two properties were used as target data in the force field parameterization.[13] Meanwhile, some researchers have used computed QM electrostatic potential surfaces as the target data in parameterization to reproduce gas-phase electrostatic potentials of molecules.[14-17] Some other researchers have noticed the misbalancing water-protein, water-water and protein-protein interaction energies in the condensed phase. Therefore, they have used computed QM data to reproduce them. However, these data are minimum-energy interaction distances and dimerization energies for small molecule-water dimers,[18] and these force fields have showed varying degrees of divergences from experimental data of other properties. Under these circumstances, we build our protein force field based on KB theory.

The KB theory of solutions provides exact expressions which relate the thermodynamics of a solution mixture to the underlying distribution between the different species in the solution.[19] This is achieved via composition dependent integrals over the various pair distribution functions, known as Kirkwood-Buff integrals (G_{ij}), which can be extracted from the experimental data, or from a simulation of a mixture at an appropriate composition.

We have argued for using the experimental KB integrals as target data for the development of accurate force field parameters for simple solutes in water and other solvents.[10,20-22] The advantages of this approach include high accuracy in the KB theory and the KB integral calculation, the use of data representing solution mixtures at finite concentrations, the sensitivity of the KB integrals to the force field parameters (especially effective partial atomic charges), and the access to changes in the solute activity with composition. The disadvantages involve the requirement of larger system sizes and simulation times to precisely determine the integrals and the inability to obtain the chemical potential.

Based on KB theory, a protein force field has been developed using KB integrals as target properties. The models for acetone, urea, NaCl and guanidinium chloride have been built as cosolvents, while the models for amides, methanol, sulfides, and aromatics are used to mimic the amino acid side chains. In order to keep the consistency of the KB force field, the models of neutral amines and carboxylic acid are needed to mimic the uncharged N- and C- terminals of proteins. Therefore, five binary mixtures, including n-butylamine with both ethanol and n-propanol, 2-aminoethanol with water, and acetic acid with both methanol and benzene, were studied in this work. N-butylamine, the smallest liquid amine at normal temperature and

pressure, was chosen to study the amino group, while the 2-aminoethanol and various solvents were used to test the transferability of the amino group in different molecules and solvents. The model for carboxyl group was tested in polar and nonpolar solvents. These organic molecules, ethanol, propanol, benzene, and methanol, were chosen as solvent instead of water, because of the solubility and ionization problems of n-butylamine and acetic acid.

2.3 Methods

2.3.1Kirkwood-Buff analysis

The detailed Kirkwood-Buff (KB) theory is thoroughly described in Chapter 1. The KB integrals (KBIs) between the different solution components

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

express the thermodynamics properties of the solution mixture. Here G_{ij} is the KB integral between species i and j, $g_{ij}^{\mu VT}(r)$ is the corresponding radial distribution function (rdf) in the grand canonical (μVT) ensemble, and r is the distance between the centers of masses of the two molecules. The KBI could be determined from the simulation data in an NpT ensemble by assuming that

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

where R represents a correlation region within which the solution composition differs from the bulk composition. All rdfs are assumed to be unity beyond R. Schnell and his coworkers have declared that they derive exact expressions for finite-volume KB integrals, which scale linearly with inverse system size and converge much better than the traditional expressions.[23] However, considering the good converge of rdf in these systems and the high computational cost of the new expressions, the new expressions show no obvious advantage over the traditional one in these systems. Therefore, the KB integrals were obtained from the traditional equation (2.2).

For a binary solution consisting of solvent (1) and solute (2), the KB integral G_{11}, G_{22} and $G_{12}=G_{21}$ can be used to define some thermodynamics quantities. The excess coordination numbers can be defined as $N_{ij}=\rho_j G_{ij}$, where $\rho_j=N_j/V$ is the number density of j particles. The excess coordination number is able to help to avoid the inherent uncertainties in both

experimental and simulated G_{ij} integrals at low j concentrations. A large positive value of N_{ij} signifies an increase in the local distribution of j around a central i molecule above what is expected for a random distribution based on the bulk composition. This in turn can be interpreted as resulting from a net attraction between i and j molecules.

The composition-dependent KBIs can be extracted from the corresponding experimental thermodynamic properties, for instance solvent activities, partial molar volumes and the solution compressibility used in the KB inversion procedure. The relationships used for the present work are[24]

$$1 + N_{11} = \rho_1 RT\kappa_T + \rho_1\rho_2 \frac{\overline{V_2}^2}{\mu_{22}}$$

$$1 + N_{22} = \rho_2 RT\kappa_T + \rho_1^2 \frac{\overline{V_1}^2}{\mu_{22}} \tag{2.3}$$

$$N_{12} = \rho_2 RT\kappa_T - \rho_1\rho_2 \frac{\overline{V_w V_c}}{\mu_{22}}$$

here K_T is the isothermal compressibility, $\overline{V}$ is the partial molar volume, and μ_{ij} represents derivatives of chemical potential (or molar activity), calculated by

$$\mu_{22} = \beta\left(\frac{\partial\mu_2}{\partial\ln m_2}\right)_{T,P} = x_1\beta\left(\frac{\partial\mu_2}{\partial\ln x_2}\right)_{T,P} \tag{2.4}$$

where m_2 and x_2 are the molality and molar fraction of species 2, respectively. Therefore, the KB theory builds a bridge between measurable experimental data and the species distributions in solution, which can be quantified in terms of KBIs, and provides a good test of a particular force field. Our previous simulations and simulations done by others have indicated that quantitative information concerning the thermodynamics of solutions can be obtained from KB theory and NpT simulations.

2.3.2 Molecular dynamics simulation

The different n-butylamine, 2-aminoethanol, and acetic acid solute solutions with either water, methanol, ethanol, propanol or benzene solvents were simulated by classical molecular dynamics techniques. The enhanced simple point charge water model (SPC/E) was used for

simulations in water.[25] Other force field models were either developed here (n-butylamine, 2-aminoethanol and acetic acid), or taken from previous work using a similar approach as presented here (methanol, ethanol, propanol and benzene).[20,26] Butylamine systems were performed at 313 K and 1 atm, while all others were simulated at 300 K and 1 atm in an effort to match the experimental activity data. The relaxation times for temperature and pressure were 0.1 and 0.5 ps respectively using the weak coupling technique. Lincs was used to constrain all bonds, allowing a 2 fs time step for integration of the equations of motion. The Particle Mesh Ewald (PME) approach was used for all electrostatic interactions and twin range cutoffs of 1.0 and 1.5 nm were employed for the Coulomb and VDW interactions, respectively. The initial configurations were generated by random placement of molecules into a cubic box of length 10 nm. All simulation times were 15 ns in length, and the final 14 ns were used to calculate ensemble averages. Diffusion constants were calculated by the mean-square fluctuation approach[27]. The excess enthalpy of mixing (ΔH_m) was determined using an established procedure which used the average potential energies and the energies from the pure solute and solvent molecules.[28]

2.3.3 Parameter development

Bonded interactions in terms of bond lengths, angles and dihedrals are determined by the experimental geometry. Nonbonded interactions were described by the Lennard-Jones 6-12 potential and a Coulomb potential. The σ and ε parameters and force constant for the bonded terms used in the simulations were taken from GROMOS force field.[29] The effective solution partial atomic charges in amino and carboxyl groups used in the Coulomb potential were parameterized to reproduce the KBIs of binary mixtures from experimental data.[30-38] The final force field parameters are displayed in Tables 2.1, 2.2, and 2.3.

Table 2.1 Final L-J Nonbonded Parameters.

atoms	ε(kJ/mol)	σ(nm)	Ref
CH_3	0.8672	0.3748	[29]
CH_2	0.4105	0.4070	[29]
C	0.3300	0.3360	[10]
H	0.0880	0.1580	[10]
N	0.5620	0.3370	[39]
$O(sp^3)$	0.6506	0.3192	[20]
$O(sp^2)$	0.5600	0.3100	[10]

Table 2.2 Final Effective Partial Atomic Charges

Compound	Atoms	Ref
n-butylamine	CH_3(0.00) CH_2(0.00) N(-0.90)	This work
Ethanol	CH_3(0.00) CH_2(0.00) O(-0.82)	[20]
n-propanol	CH_3(0.00) CH_2(0.00) O(-0.82)	[20]
2-Aminoethanol	H(0.52) O(-0.82) CH_2(0.30) CH_2(0.20) N(-0.90) H(0.35)	[20] This work
Water	O(-0.8467) H(0.4238)	[25]
Acetic Acid	CH_3(0.00) C(0.62) H(0.55) $O(sp^2)$ (-0.55) $O(sp^3)$(-0.62)	This work
Benzene	C(-0.13) H(0.13)	[26]
Methanol	CH_3(0.30) O(-0.82) H(0.52)	[20]

Table 2.3 Dihedral Force Field Parameters

atom	Force $k_\phi(kJ\ mol^{-1})$	Multiplicity n	Ref
CH$_2$-CH$_2$-N-H	1.75	3	[13]
CH$_2$-CH$_2$-CH$_2$-N	-0.75	1	[13]
	0.50	2	[13]
	8.50	3	[13]
N-CH$_2$-CH$_2$-O	27.0	1	This work
	0.5	2	[13]
	8.5	3	[13]
CH$_3$-CH$_2$-CH$_2$-CH$_2$	-0.25	1	[13]
	-0.50	2	[13]
	7.25	3	[13]
CH$_2$-CH$_2$-O-H	0.75	1	[13]
	0.50	2	[13]
	3.00	3	[13]
CH$_3$-C-O-H	4.25	1	[13]
	-20.00	2	[13]

*The phase shift for all the molecules are 0.00.

2.4 Results and Discussion

2.4.1 n-Butylamine (a) + Ethanol (e)/n-Butylamine (a) + n-Propanol (p)

In order to model neutral amines, we have focused on a single linear amine for which experimental data could be found. The partial atomic charges on the amino groups were varied until a charge distribution that could reproduce the target experimental data was obtained. The center-of-mass radial distribution functions (COM rdfs) for n-butylamine mixture with both

ethanol and n-propanol are shown in Figure 2.1. As would be expected for these two similar systems, the COM rdfs displayed similar positions for the maxima and minima, and all rdfs were essentially unity beyond 1.5 nm. The first maxima increased slightly with the solute mole fraction. The only small difference between these two systems is that the first solvent-solvent solvation shell increased more significantly and appeared at shorter distance for ethanol system than for n-propanol system due to the smaller volume of ethanol molecules. This suggests an increasing degree of solvent self-association with increasing n-butylamine mole fraction.

Figure 2.1 Center-of-mass radial distribution functions (g) as a function of distance for n-butylamine (a) with ethanol (e) (left) and n-propanol (p) (right) at 313K and 1atm. The legends correspond to mole fractions of n-butylamine.
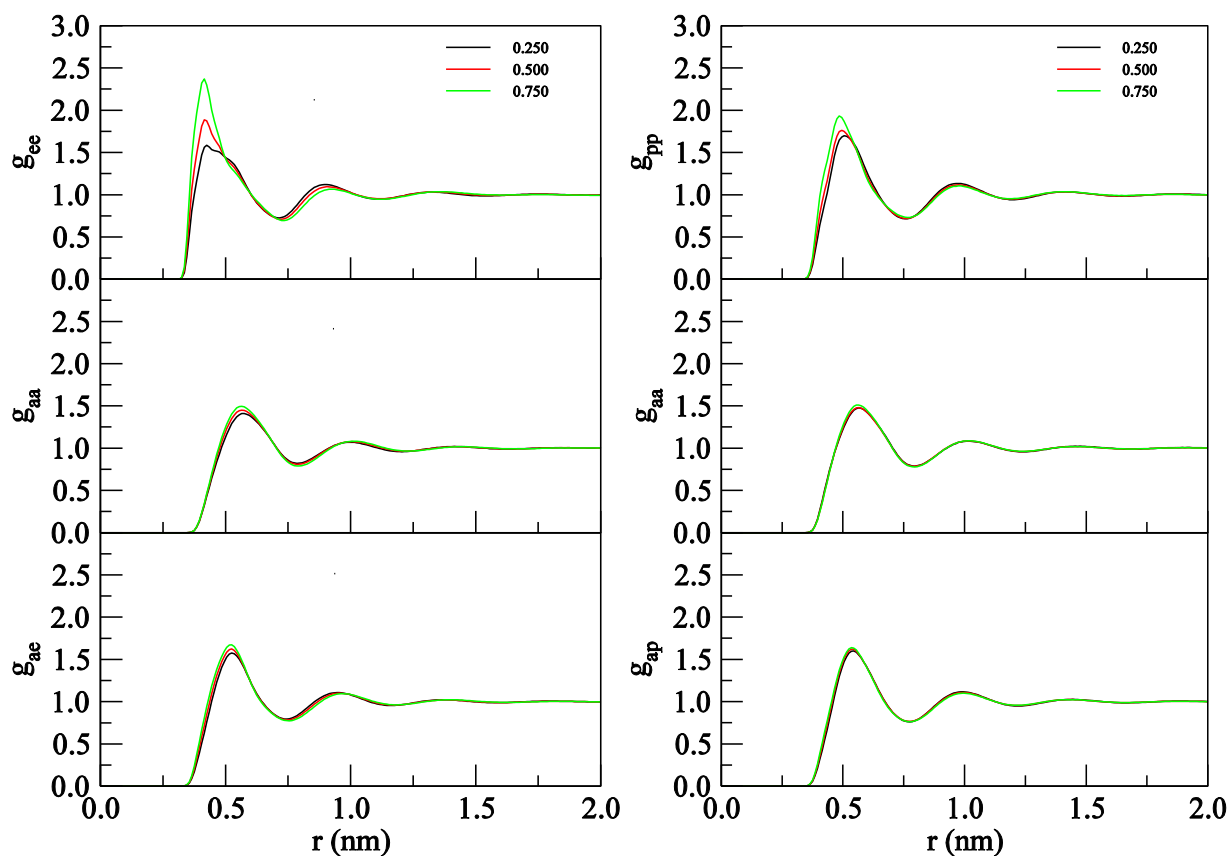
The simulated and experimental excess coordination numbers as a function of n-butylamine mole fraction for both systems are displayed in Figure 2.2. The simulated results quantitatively reproduced the experimental data over the whole composition range, especially for ethanol system. The butylamine-butylamine excess coordination numbers decreased, while the butylamine-ethanol (or propanol) and ethanol-ethanol (or propanol-propanol) excess coordination numbers increased with the increasing butylamine mole fraction, which showed the same trend as in experimental data. This indicated that the changes in molecule distributions determine the solution behavior to a large extent.

Figure 2.2 Excess coordination numbers (N) as a function of n-butylamine mole fraction in n-butylamine (a) with ethanol (e) (left) and n-propanol (p) (right) systems at 313K and 1atm. Lines represent the experimental data, dots are the simulation results and error bars show the standard deviation between multiple 5 ns block averages.
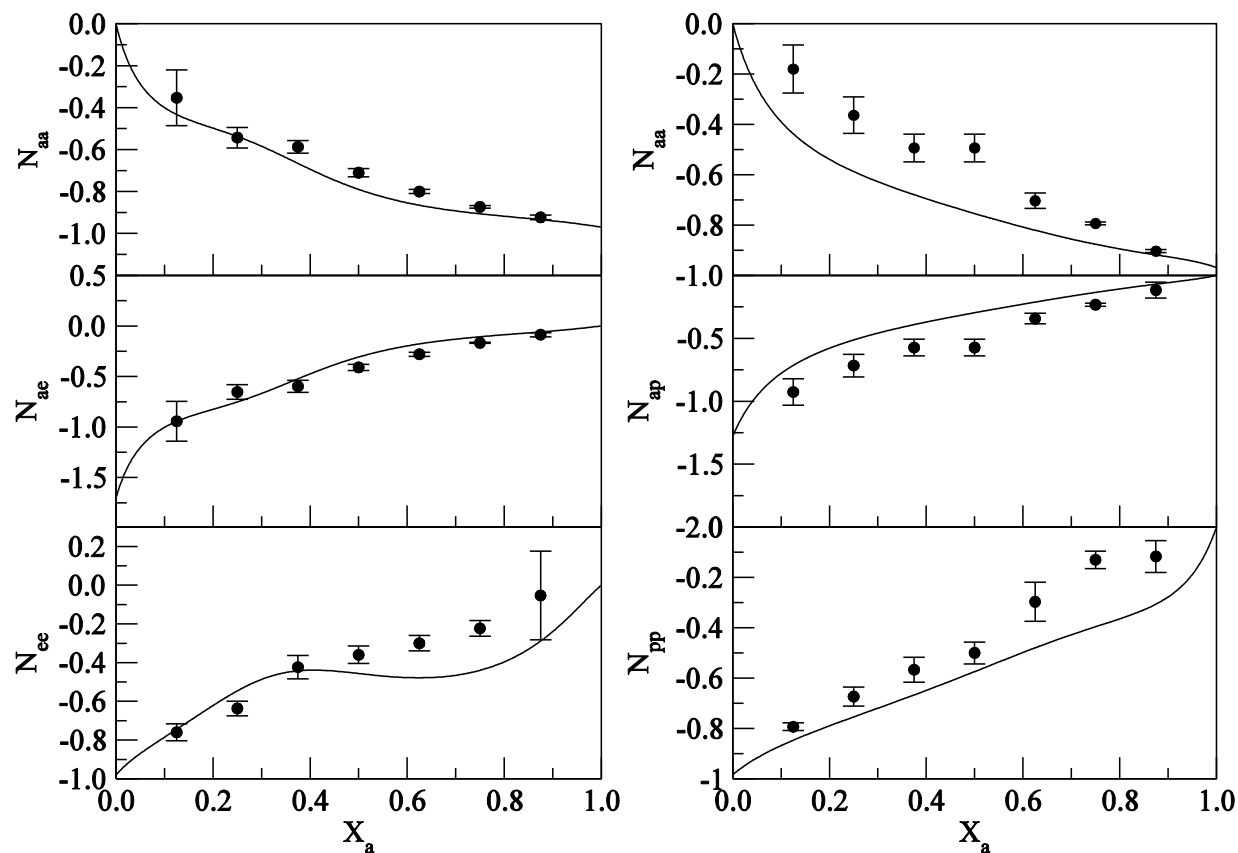
The KBFF models were designed to reproduce solution properties, particularly the experimental KB integrals (or excess coordination numbers). However, these models have been reported with the ability to reproduce other properties of solution mixtures and pure liquids. Therefore, we used the KBFF models to calculate the experimental diffusion coefficient (D), relative permittivity (ε), enthalpy of mixing (ΔH_m) and compared the results with experimental data if available.

Figure 2.3 Diffusion constants (D, $10^{-9} m^2/s$) as a function of n-butylamine mole fraction in n-butylamine (a) with ethanol (e) (top) and propanol (p) (bottom) systems.
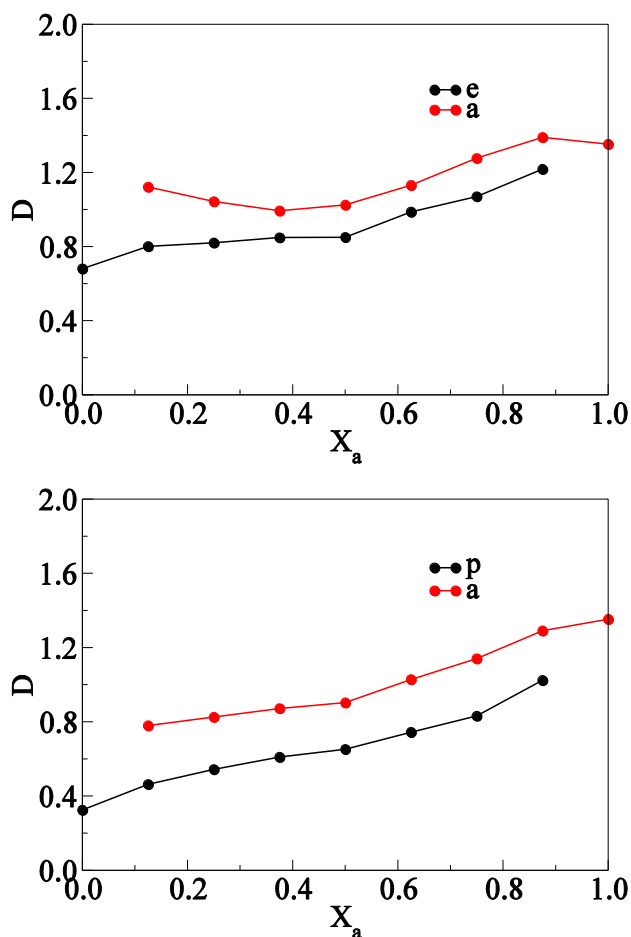
The diffusion constants for both the solute and solvent, which were calculated using the mean square fluctuation approach, are displayed in Figure 2.3. They exhibited the similar smooth expected trends with composition. However, experimental results for these systems were unavailable. The relative permitivities for both systems, calculated from the dipole moment fluctuations, are displayed in Figure 2.4. The dielectric constants for n-butylamine and ethanol system decreased with the increasing n-butylamine mole fraction. Considering the fluctuation, which is shown as large error bars in the figure, the dielectric constant for n-butylamine and n-propanol did not change much with the n-butylamine mole fraction. Again, experimental data for mixtures are unavailable.

Figure 2.4 Relative permittivities as a function of n-butylamine mole fraction in n-butylamine (a) with ethanol (e) (top) and n-propanol (p) (bottom) systems.
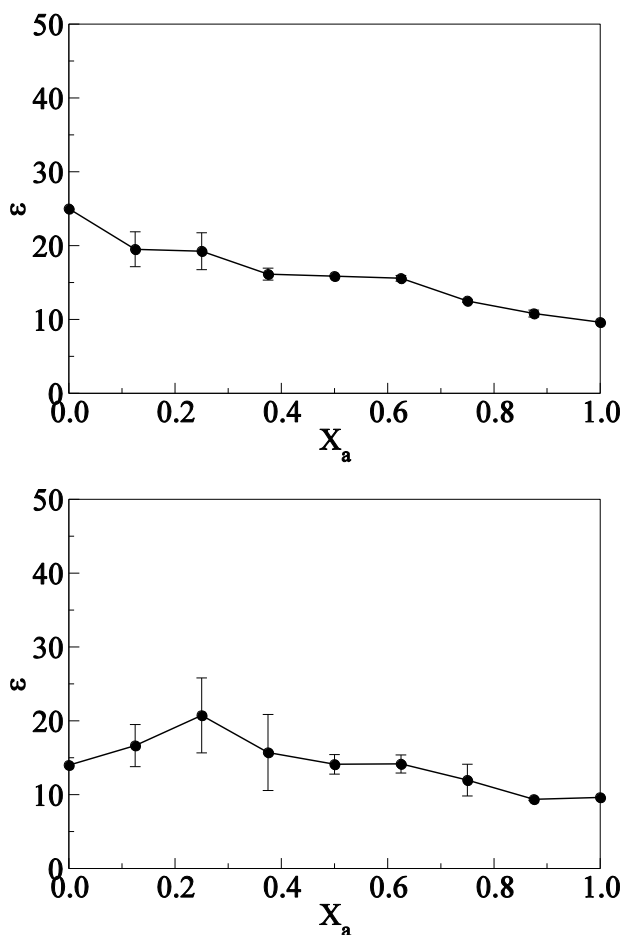
2.4.2 2-Aminoethanol (a) + Water (w)

In order to model some residues with polar side chains, e.g. serine, we performed simulations of 2-aminoethanol originally using the same amino charges as used for n-butylamine. However, it was not possible to accurately fit the experimental data with this restriction. This is probably due to the close vicinity of two polar groups in the molecule, which can then significantly polarize each other. Hence, a new set of amino charges were developed for the 2-aminoethanol mixtures with water. The COM rdfs and excess coordination numbers compared with experimental data are presented in Figure 2.5 and 2.6.

Figure 2.5 Center-of-mass radial distribution function (g) as a function of distance for 2-aminoethanol (a) with water (w) at 300K and 1atm. The legends correspond to mole fractions of 2-aminoethanol.
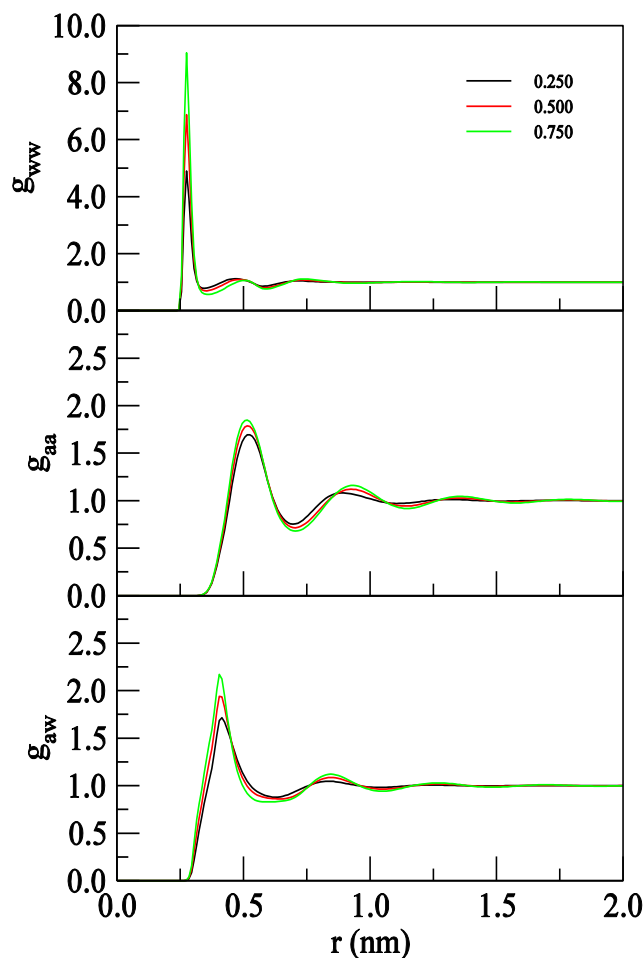
The solvent-solvent rdfs shows sharp peaks around 0.28nm. The peaks increased with increasing 2-aminoethanol composition, while the valley decreased and slightly shifted to a larger distance. The same structural patterns were obtained in Gubskaya and Kusalik's work.[40] They performed molecular dynamics simulation to study 2-aminoethanol structure in aqueous solution using OPLS-based potential models, and explained this as a small tendency of water molecule association in aminoethanol-rich mixtures. The maxima of aminoethanol-aminoethanol and aminoethanol-water rdfs increased a little when the aminoethanol composition increased, and did not show any remarkable shift with the increased aminoethanol composition, which means the aminoethanol-aminoethanol and aminoethanol-water structures are not affected greatly by the aminoethanol concentration.

Figure 2.6 Excess coordination numbers (N) as a function of 2-aminoethanol mole fraction in 2-aminoethanol (a) and water (w) system at 300K and 1atm. Lines represent the experimental data, dots are the simulation results and error bars show the standard deviation between multiple 5 ns block averages.
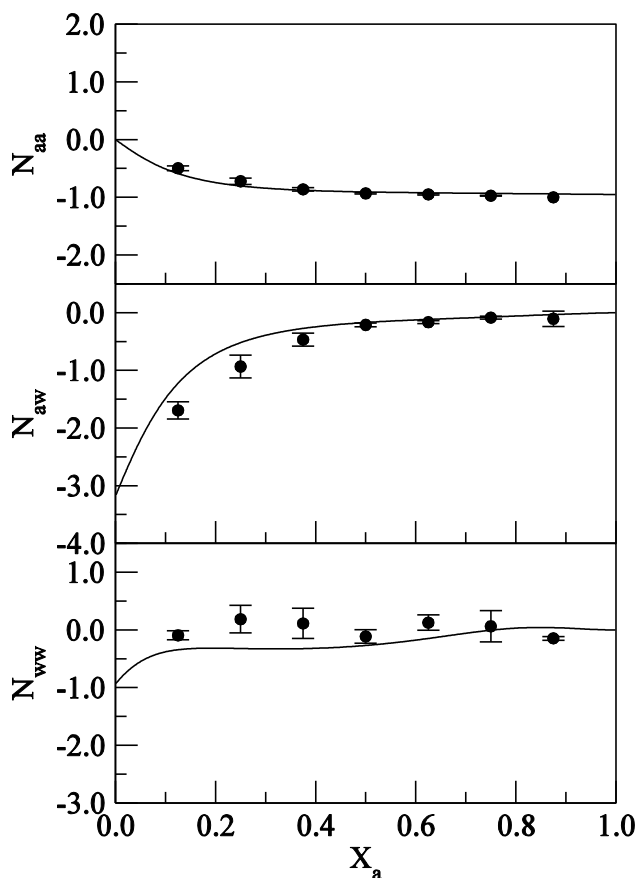
Figure 2.6 showed that our 2-aminoethanol model with SPC/E water model perfectly reproduced the experimental excess coordination numbers, especially for aminoethanol-aminoethanol interactions. The aminoethanol-aminoethanol and aminoethanol-water excess coordination numbers changed very slightly with the increasing aminoethanol concentration in the aminoethanol-rich mixture, which indicated that 2-aminoethanol did not show obvious preference for self association in the whole concentration range. This is in an agreement with Gubskaya's conclusion. KB integral for 2-aminoethanol system (not shown), where these excess coordination numbers were calculated from, should be mentioned here. Matteoli and Lepori did

an analysis through the Kirkwood Buff integrals for organic solutes including 2-aminoethanol,[41] and their result of KB integrals exhibited a subtle difference from ours in very high or low aminoethanol composition. This is probably due to the experimental accuracy at extreme high or low concentration, and the different experimental data or KB inverse fitting procedure we use.

Figure 2.7 Diffusion constants (D, $10^{-9}m^2/s$) as a function of 2-aminoethanol mole fraction in 2-aminoethanol and water system.
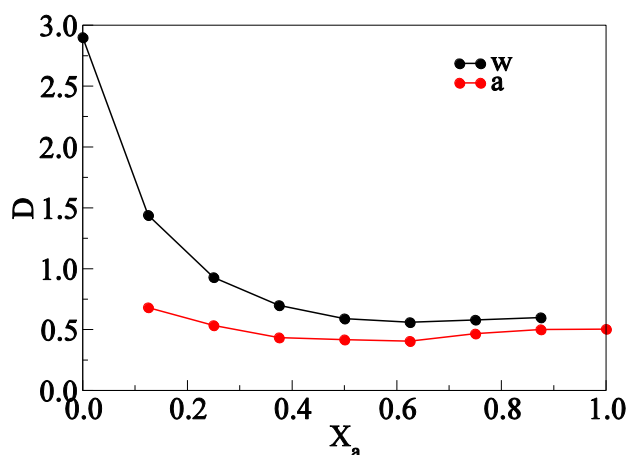
The diffusion constants for both the aminoethanol and water for 2-aminoethal-water system are displayed in Figure 2.7. They exhibited the similar smooth expected trends with composition. Unfortunately, we could not find any appropriate mixture experimental data for comparison. Our results show the same scale as Gubskaya's simulated results, but a smaller range. The relative permitivities are displayed in Figure 2.8. The dielectric constants for system decreased when 2-aminoethanol composition increased in the whole concentration range.

Figure 2.8   Relative permittivities as a function of 2-aminoethanol mole fraction in 2-aminoethanol (a) and water (w) system.
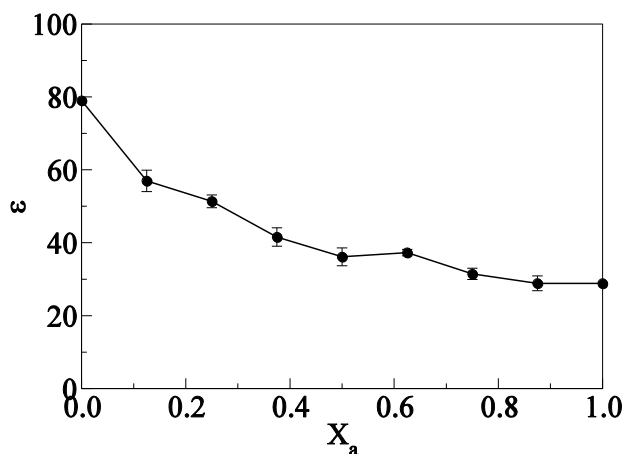
Figure 2.9 2-Aminoethanol (N-C-C-O) gauche fraction as a function of 2-aminoethanol mole fraction in 2-aminoethanol (a) and water (w) system. Lines represent the experimental data, and dots are from our simulation results.
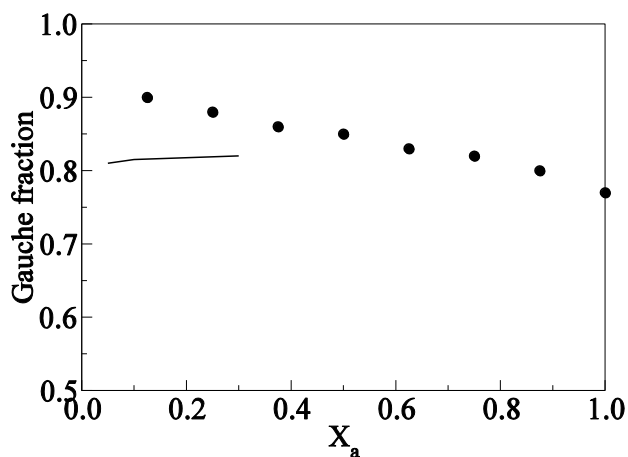
Besides the effective partial atomic charges, the dihedral parameter in aminoethanol (N-C-C-O) was also adjusted in order to reproduce the conformation trans/gauche preferences from the experimental NMR spectra.[33] The conformational characteristics of 2-aminoethanol were calculated in its aqueous mixtures. Compared with the results of Gubskaya's study with OPLS-

based potential model[40], which provided almost 100% gauche conformation in high and very low aminoethanol concentration, our model gave a smoother conformational change and was closer to the NMR spectra result at low concentrations, which suggests our model has a large improvement in the conformational preference study.

2.4.3 Acetic acid (a) + Benzene (b) / Acetic acid (a) +Methanol (m)

In order to model protonated carboxylic acid, we focused on acetic acid and mixtures of this acid in two different solvent environments. We avoided water as a solvent in order to eliminate pH effects. Binary mixtures of acetic acid with both methanol and benzene have been performed. The radial distribution functions are displayed in Figure 2.10.

Figure 2.10 Center-of-mass radial distribution functions (g) as a function of distance for acetic acid (a) with benzene (b) (left) and methanol (m) (right) at 300K and 1atm. The legends correspond to mole fractions of acetic acid.
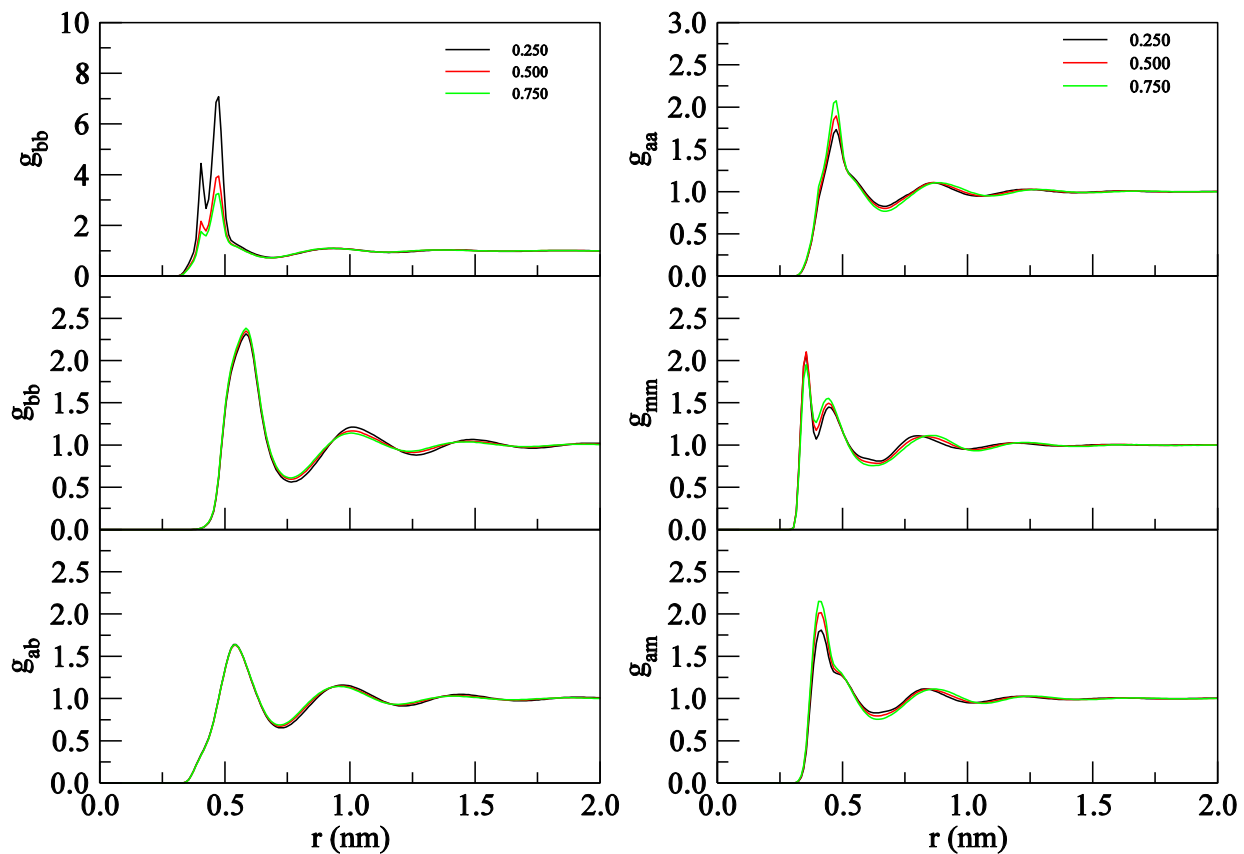
Different from before, acetic acid and benzene showed a higher second peak in the benzene-benzene COM rdfs, and the first and second peaks were decreased dramatically with increasing acetic acid composition. However, in acetic acid and methanol system, the second peaks of methanol-methanol COM rdfs were shorter and broader than the first one. Comparing these two systems, the methanol-methanol COM rdf peaks were observed in much closer distances and showed much shorter than the benzene-benzene ones, while the acetic acid-methanol COM rdf peaks were also observed in closer distances, but higher than acetic acid-benzene ones. These could be explained in terms of smaller molecular radius of methanol than that of benzene, and the strong hydrogen bonding interaction and polar-polar interaction which existed between acetic acid and methanol rather than benzene.

Figure 2.11 Excess coordination numbers (N) as a function of acetic acid mole fraction in acetic acid (a) with benzene (b) (left) and methanol (m) (right) system at 300K and 1atm. Lines represent the experimental data, dots are the simulation results and error bars show the standard deviation between multiple 5 ns block average.
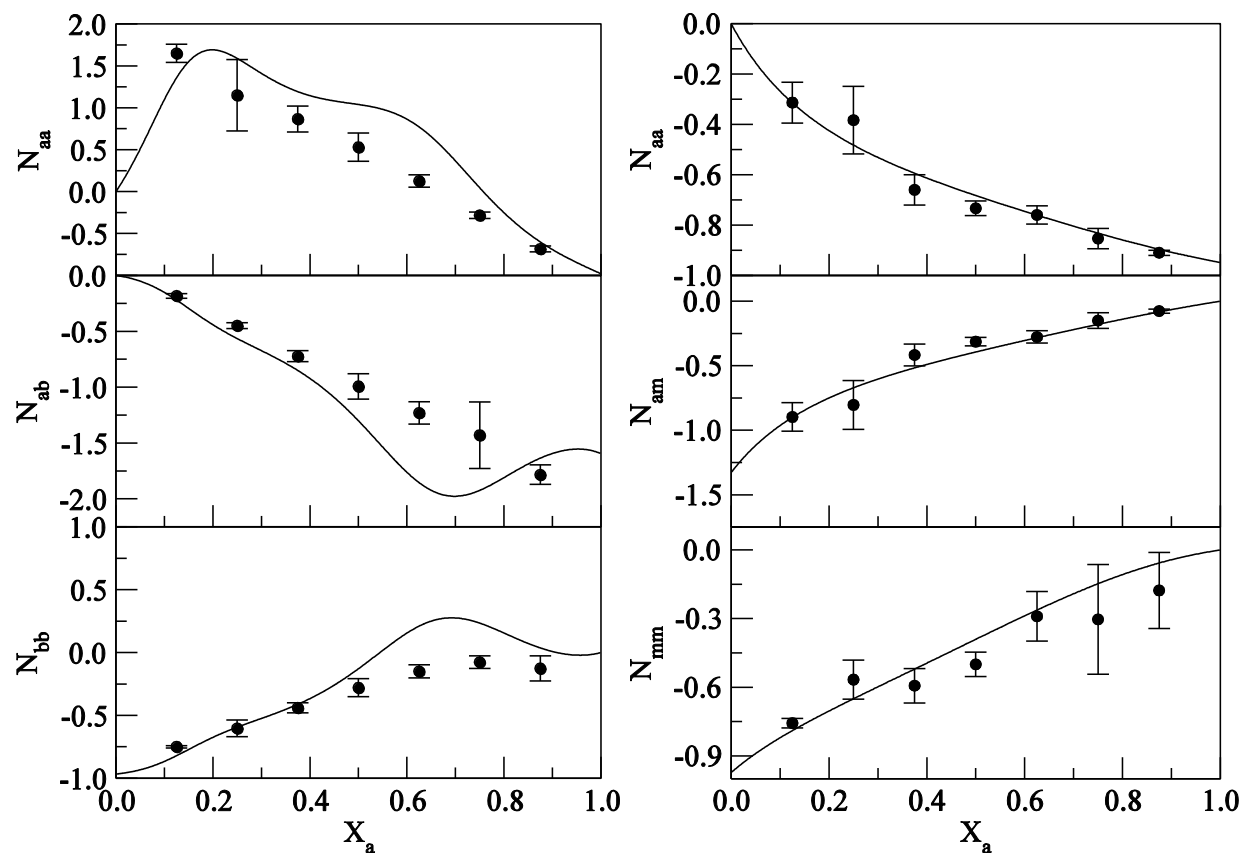
The excess coordination numbers of both the two acetic acid systems are shown in Figure 2.11. It can be seen that our model has quantitatively reproduced those experimental results, and the acetic acid-methanol model system worked well. All excess coordination numbers changed linearly with acetic acid composition within the entire concentration range. The most obvious difference between these two systems was that the acetic acid-acetic acid excess coordination number was positive in the acetic acid-rich benzene solution, but negative in methanol solution, which suggested the acetic acid self-association in the former, but not in the latter. Another difference was that the acetic acid-benzene excess coordination number was decreased, whereas

acetic acid-methanol one was increased, which indicated that acetic acid molecules prefer to interact with methanol rather than benzene.

Figure 2.12 Diffusion constants (D, $10^{-9}m^2/s$) as a function of acetic acid mole fraction in acetic acid (a) with benzene (b) (top) and methanol (m) (bottom) systems.
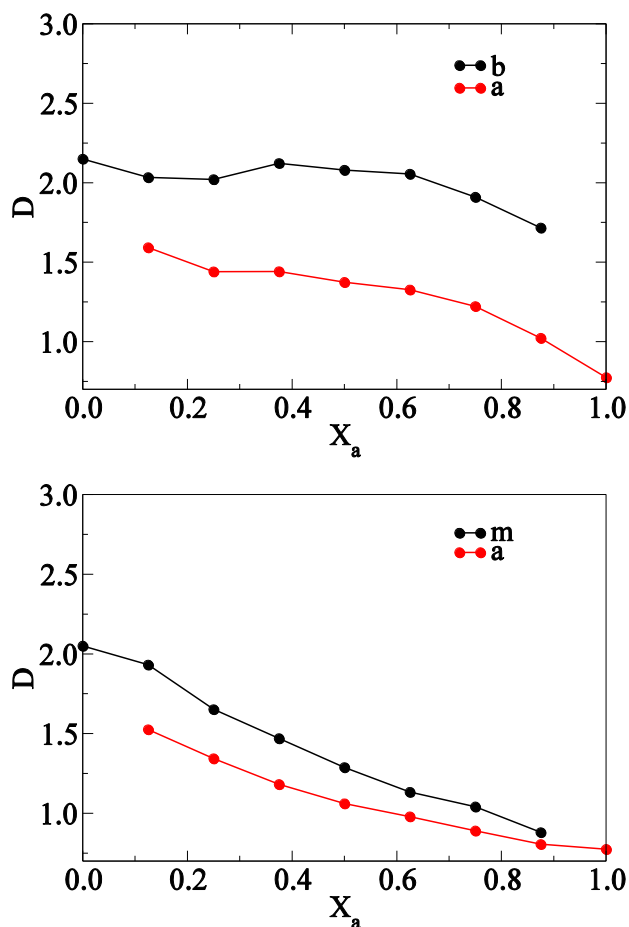
The diffusion constants for acetic acid with two different solvents are displayed in Figure 2.12. They seemed to be similar to each other, although methanol showed a larger drop than benzene with acetic acid. The relative permitivities for these two systems are displayed in Figure 2.13, and they showed opposite changes with increasing acetic acid composition. The dielectric constants for and benzene system was increased, whereas methanol system was decreased.

However, the permitivities of the system with methanol was still larger than that with benzene. These could be easily explained as the pure methanol had a much higher permittivity than benzene, and the value of acetic acid exists between them.

Figure 2.13 Relative permittivities as a function of acetic acid mole fraction in acetic acid (a) with benzene (b) (top) and methanol (m) (bottom) systems.
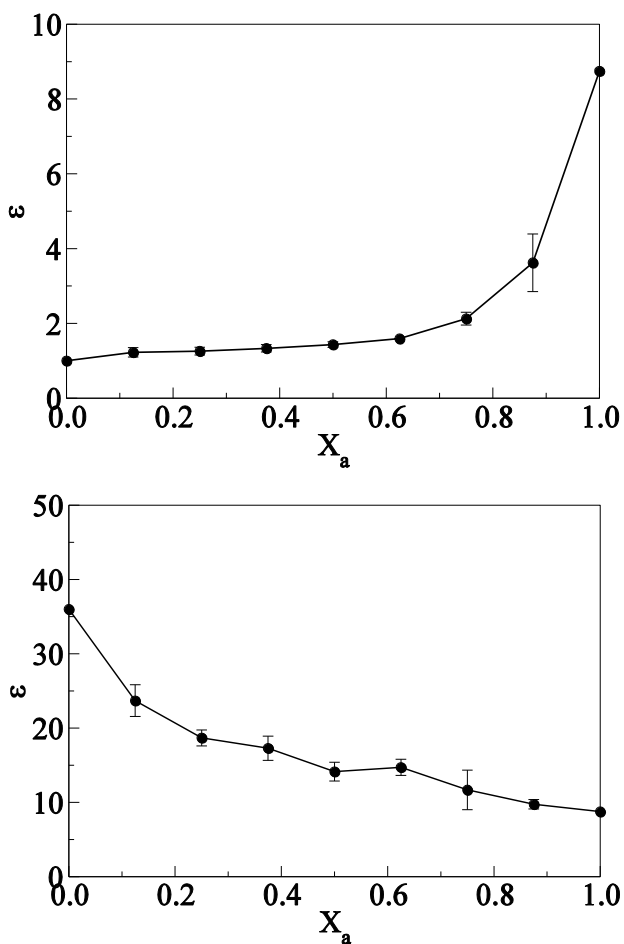
2.4.4 Pure liquid properties and enthalpy of mixing

Table 2.4 shows the experimental and simulated density, enthalpy of vaporization, isothermal compressibility, relative permitivities, and self-diffusion constant for the solutes in

these systems. All of the simulated density values were in agreement with experiments. The enthalpy of vaporization was calculated for the pure liquids according to

$$\Delta_{vap}H = \frac{-\left(E_{tot}^l - E_{intra}^g\right)}{N} + RT \qquad (2.5)$$

where E_{tot}^l and E_{intra}^g are the potential energy and intramolecular potential energy of the molecules in the liquid phase and in the gas phase, respectively, and we assumed that the intramolecular potential in the gas phase was the same as that in the liquid phase. In table 2.4, the simulated results acceptably agreed with the experiments. It is worthy to point that the enthalpy of dimerization of 2-aminoethanol was deduced from the result of equation (2.5) since the most part of 2-aminoethanol molecules exist as a dimer in gas phase. As expected, a large deviation exists between experimental and simulated enthalpy of vaporization, and it is because the induced dipole moments are changed from liquid phase to gas phase, while we only focus on the liquid phase. A polarization correction has to be applied if enthalpy of vaporization needs to be reproduced.

Table 2.4 Comparison of experimental and simulated pure liquid properties from the KBFF models of n-Butylamine, 2-Aminoethanol and Acetic acid

	$\rho(g/cm^3)$		$\Delta_{vap}H$		$K (10^5 bar^{-1})$		$D (10^{-9} m^2/s)$		ε	
	Exp	Sim	Exp	Sim	Exp	Sim	Exp	Sim	Exp	Sim
n-Butylamine	0.733[42]	0.752	35.7[43]	48.3	12.6[44]	8.8	0.96[45]	1.35	4.2[46]	9.6
2-Aminoethanol	1.018[47]	0.926	64.0[48]	88.9	3.8[49]	5.9		0.50	26.0[50]	28.8
Acetic Acid	1.044[38]	1.000	23.4[51]	20.5	7.0[52]	8.3	1.25[53]	0.77	6.6[54]	8.7

The excess enthalpies of mixing for the range of mixtures described here are calculated by

$$\Delta H_{mix}^E = \frac{H_{sol}}{N} - x_1 H_1^0 - x_2 H_2^0 \qquad (2.6)$$

and provided in Figure 2.14. As it shows, our simulating model reasonably agrees with the experimental results. The 2-aminoethanol model worked best, and acetic acid and n-butylamine

models qualitatively reproduced the same trend and the same scale as experimental data. The comparison with experimental results indicates that the mixtures with ethanol are too unfavorable, while the mixtures with n-propanol are too favorable.

Figure 2.14 Excess enthalpy of mixing (ΔH_m, kJ/mol) as a function of cosolvent (1) mole fraction in different solvent systems shown in different colors. Lines represent the experimental data, dots and error bars are from our simulation results.
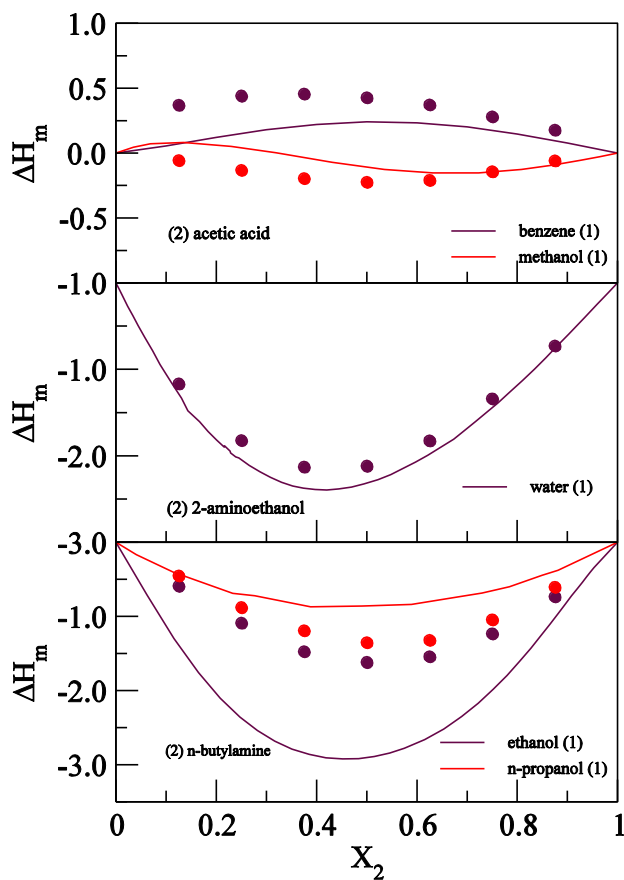
2.5 Conclusion

We have developed new models for neutral amines and protonated carboxylic acids to use in a complete protein force field. The parameters were developed using the KB theory of solution mixtures and it provides an alternative to more traditional approaches. The results obtained here appear to be reasonable in comparison to the results obtained from other solute force fields[55,56].

2.6 References

(1)   Christen, M.; Hunenberger, P. H.; Bakowies, D.; Baron, R.; Burgi, R.; Geerke, D. P.; Heinz, T. N.; Kastenholz, M. A.; Krautler, V.; Oostenbrink, C.; Peter, C.; Trzesniak, D.; Van Gunsteren, W. F. *J. Comput. Chem.* **2005**, *26*, 1719.

(2)   Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545.

(3)   Arnautova, Y. A.; Jagielska, A.; Scheraga, H. A. *J. Phys. Chem. B* **2006**, *110*, 5025.

(4)   Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. *J Chem Theory Comput* **2007**, *3*, 1851.

(5)   Chen, B.; Potoff, J. J.; Siepmann, J. I. *J. Phys. Chem. B* **2001**, *105*, 3093.

(6)   Smith, P. E.; Vangunsteren, W. F. *J. Mol. Biol.* **1994**, *236*, 629.

(7)   Shukla, D.; Shinde, C.; Trout, B. L. *J. Phys. Chem. B* **2009**, *113*, 12546.

(8)   Ma, L.; Pegram, L.; Record, M. T., Jr.; Cui, Q. *Biochemistry* **2010**, *49*, 1954.

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

(10)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 10663.

(11)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *119*, 11342.

(12)   Kukol, A. *Molecular Modeling of Proteins*; Humana Press, 2007.

(13)   Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.

(14)   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. *J. Am. Chem. Soc.* **1993**, *115*, 9620.

(15)   Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.-Us* **1993**, *97*, 10269.

(16)   Cieplak, P.; Cornell, W. D.; Bayly, C.; Kollman, P. A. *J. Comput. Chem.* **1995**, *16*, 1357.

(17)   Fox, T.; Kollman, P. A. *J. Phys. Chem. B* **1998**, *102*, 8070.

(18)   Neria, E.; Fischer, S.; Karplus, M. *J. Chem. Phys.* **1996**, *105*, 1902.

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

(20)   Weerasinghe, S.; Smith, P. E. *J. Phys. Chem. B* **2005**, *109*, 15080.

(21)   Feig, M. *Modeling Solvent Environments*; WILEY-VCH, 2010.

(22)   Ploetz, E. A.; Bentenitis, N.; Smith, P. E. *Fluid Phase Equilibr.* **2010**, *290*, 43.

(23)   Kruger, P.; Schnell, S. K.; Bedeaux, D.; Kjelstrup, S.; Vlugt, T. J. H.; Simon, J. M. *J Phys. Chem. Lett.* **2013**, *4*, 235.

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

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

(26)   Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 18154.

(27)   Chitra, R.; Smith, P. E. *J. Phys. Chem. B* **2000**, *104*, 5854.

(28)   Walser, R.; Mark, A. E.; van Gunsteren, W. F.; Lauterbach, M.; Wipff, G. *J. Chem. Phys.* **2000**, *112*, 10450.

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

(30)   Lark, B. S.; Banlpal, T. S. *J. Chem. Eng. Data* **1984**, *29*, 277.

(31)   Papaioannou, D.; Bridakis, M.; Panayiotou, C. G. *J. Chem. Eng. Data* **1993**, *38*, 370.

(32)   Touhara, H.; Okazaki, S.; Okino, F.; Tanaka, H.; Ikari, K.; Nakanishi, K. *J. Chem. Thermodyn.* **1982**, *14*, 145.

(33)   Smith, T. D.; Gerken, J. B.; Jog, P. V.; Roberts, J. D. *Org. Lett.* **2007**, *9*, 4555.

(34)   Checoni, R. F. *J. Therm. Anal. Calorim.* **2010**, *101*, 349.

(35)   Dutta Choudhury, M. K. *Ind. J. Chem.* **1976**, *14A*, 553.

(36)   Wagner, M.; Apelblat, A.; Tamir, A. *J. Chem. Thermodyn.* **1980**, *12*, 181.

(37)   Zarei, H. A. *J. Mol. Liq.* **2007**, *130*, 74.

(38)   Howard, K. S.; Hammond, L. W.; McAllister, R. A.; Pike, F. P. *J. Phys. Chem.-Us* **1958**, *62*, 1598.

(39)   Gee, M. B.; Cox, N. R.; Jiao, Y. F.; Bentenitis, N.; Weerasinghe, S.; Smith, P. E. *J. Chem. Theory Comput.* **2011**, *7*, 1369.

(40)   Gubskaya, A. V.; Kusalik, P. G. *J. Phys. Chem. A* **2004**, *108*, 7165.

(41)   Matteoli, E.; Lepori, L. *J. Chem. Phys.* **1984**, *80*, 2856.

(42)   O'Neil, M. J. *An encyclopedia of Chemicals, Drugs, and Biologicals, 13th ed*; Merck & Co., Inc., 2001.

(43)   Lide, D. R. *CRC Handbook of Chemistry and Physics, 86th ed*; CRC Press, 2006.

(44)   Oswal, S. L.; Oswal, P.; Gardas, R. L.; Patel, S. G.; Shinde, R. G. *Fluid Phase Equilibr.* **2004**, *216*, 33.

(45)   Lugg, G. A. *Anal. Chem.* **1986**, *40*, 1072.

(46)   Cezary, K. M.; Wojciech, K. J.; Dorota, C.-M. *J. Chem. Eng. Data.* **2003**, *48*, 1037.

(47)   Lide, D. R.; Milne, G. W. A. *CRC Handbook of Data on Organic Compounds, Volume I.*; CRC Press, 1994.

(48)   *American Institute for Chemical Engineering Design institute for physical properties, project 801,* ; http://www.knovel.com, 2010

(49)   In *http:/i-systems. dechema. de/detherm,* ; isothermal compressibility ed. 2011.

(50)   Austgen, D. M.; Rochelle, G. T.; Peng, X.; Chen, C. C. *Ind. Eng. Chem. Res.* **1989**, *28*, 1060.

(51)   Yarym-Agaev, N. L.; Kogan, E. A.; Rudin, V. Y.; Titova, V. A. *Rus. J. Phy. Chem. A* **1963**, 1445.

(52)   Vong, W. T.; Tsai, F. N. *J. Chem. Eng. Data* **1997**, *42*, 1116.

(53)   Bidstrup , D. E.; Geankoplis, C. J. *J. Chem. Eng. Data* **1963**, *8*, 170.

(54)   Kaatze, U.; Menzel, K.; Pottel, R. *J. Phys. Chem.-Us* **1991**, *95*, 324.

(55)   Jorgensen, W. L. *J. Phys. Chem.-Us* **1986**, *90*, 1276.

(56)   da Silva, E. F.; Kuznetsova, T.; Kvamme, B.; Merz, K. M. *J. Phys. Chem. B* **2007**, *111*, 3695.

Chapter 3 - Urea Cosolvent Effects Using Kirkwood-Buff Theory

3.1 Abstract

Adding a protein denaturant is widely used to change protein conformational stability in aqueous solutions. In this study, the cosolvent effect on peptide aggregation is investigated by molecular dynamics simulations. Urea is chosen as the cosolvent and several small solute molecules are used as simple models of amino acid side chains. Neo-pentane, benzene, methanol are selected to mimic the hydrocarbon aliphatic, aromatic and amphiphilic moieties, respectively. Moreover, glycine, which is the simplest amino acid, is used to study the urea-zwitterionic amino acid interactions. The molecular distribution changes with the concentrations of solute and cosolvent were obtained and analyzed by Kirkwood-Buff (KB) theory. The chemical potential derivatives to the cosolvent concentrations were calculated in molarity scale. Four solutes displayed large differences in solvation shells and preferential interactions. All these results indicate that KB theory could be used to describe the intermolecular interactions in the ternary systems, qualitatively and quantitatively.

3.2 Introduction

3.2.1 General information of denaturant

Protein denaturation is significant for understanding the physical nature of protein folding. As mentioned in Chapter 1, adding a cosolvent is a direct and widely used method to change the stability of protein in aqueous environment. A cosolvent is defined as small molecules which are different from the primary solvent-water in a mixed system.[1] According to their different effects on protein stability, they can be divided into two categories: protecting cosolvents and denaturing cosolvents. A protecting cosolvent, osmolyte, is known to stabilize protein native structure, e.g. polyols, some amino acids, TMAO, and sugars, while a denaturing cosolvent, denaturant, can shift the protein folded state toward the unfolded state and increase protein solubility in water, e.g. urea and guanidinium chloride (GdmCl).[2-7] Furthermore, the protein conformational change due to the addition of denaturant could have an effect on protein aggregation behavior.[8,9] Therefore, understanding the molecular interactions between the protein, cosolvent and water at molecular as well as atomic level appears to be of greater importance.

A large number of experiments have been performed to gather information about protein, cosolvent and solvent structure.[10-14] However, these experiments have displayed some molecule limitations or low resolution problems,[15-17] and they are unable to give detailed descriptions and interpretations about the preferential binding effects between different components. On the other hand, computer simulations show an obvious advantage to provide distinct insight into the protein folding/unfolding process which may occur in the cell, and the numeric calculations performed in molecular dynamics could bring us not only dynamic folding details, but also the intermolecular interaction information.

Urea is a widely used protein denaturant. Compared to GdmCl, which has the stronger protein destabilizing ability,[18] urea has its special features: it is the only organic denaturant to induce protein into random coil states[19] and it is accumulated in high concentrations in several species, for example, in mammalian renal medulla, marine elasmobranches, and amphibians,[20-22] and it is extensively used for agriculture production. Therefore, the protein denaturation mechanism caused by urea attracts worldwide attention. However, after being studied over hundred years, it still remains unclear. Two ideas are popular at present: the direct and indirect mechanisms. In the direct mechanism, urea directly interacts with the protein, destabilizing the native state consequently.[23] In the indirect mechanism, urea molecules modify the hydrogen-bonded structure of water first and then weakens the protein-stabilizing hydrophobic effect.[24] Both of these two mechanisms have a large amount of supporting data,[25-34] and neither of them is dominant or completely convincing.

3.2.2 Experimental study of urea denaturation behavior

In order to study the mechanism of urea denaturation, innumerable experiments have been performed during the recent years. For example, x-ray diffraction, nuclear magnetic resonance (NMR), and neutron scattering are used to figure out the solvent structure at the atomic level, and isothermal titration calorimetry, vapor–pressure absorption isotherm experiments are used to measure the thermodynamic properties of the solution.[10-17] Here we briefly review some of the models that some researchers have built based on experimental data to explain the cosolvent effect of urea.

3.2.2.1 Local-bulk domain model

Timasheff proposed a two-domain model in 1972.[35] In this model, the protein solution is divided into a local domain surrounding each protein molecule and a bulk domain. Generally, the composition of the local domain differs from that of the bulk domain because of the different interactions between protein-cosolvent and protein-water. Record Jr et al. completed this model in 1995.[36] They characterized these differences by a molal-scale local-bulk partition coefficient and interpreted the ratios of m-values to the protein surface area changes using an analysis on the basis of the partition coefficient which quantitatively describes the thermodynamic consequences of weak interactions between cosolvent-protein surfaces compared to interactions between water-protein surfaces. However, using current experiment techniques, it is usually difficult to measure changes in the solution composition passing through one or more solvation shells.

3.2.2.2 Binding or exchange models

In the early 1990's the binding or exchange models were built as an attempt to interpret the complicated behavior of the stability and preferential interaction of protein in mixed solvents.[23,37] In these models, the surface of a protein is divided into a denumerable set of independent and similar binding sites and these sites are bound or occupied by the cosolvent or water molecules. The number of binding sites of a protein in both folded and unfolded states are estimated from the experiment calorimetric data, and other thermodynamic properties, including free energy, enthalpy, entropy, and heat capacity effect of binding, could also be obtained through these models.[38] However, the determination of a series of independent identical binding sites and the heterogeneous nature due to the conformational freedom of proteins make these models infeasible in computer simulations.

3.2.2.3 Preferential interaction and preferential binding

Timashiff gave a complete definition of preferential interaction and summarized its application on modulating biochemical reactions and biological processes in 1998.[1] Preferential interaction means competition between cosolvent and water. When a protein and cosolvent molecules are coexisting in an aqueous solution, one could perturb the chemical potential of the other, and the perturbation, the chemical potential derivative to the concentration, is defined as the preferential interaction parameter

$$\mu_{23} = \left(\partial \mu_2 / \partial m_3 \right)_{T,P,m_2} \tag{3.1}$$

$$\mu_{32} = \left(\partial\mu_3 / \partial m_2\right)_{T,P,m_3} \tag{3.2}$$

where the subscripts are designated 1 as water, 2 as protein, and 3 as cosolvent. It should be mentioned that here m_i is the concentration of component i in molality scale. The thermodynamic equilibrium of the system would be disturbed by adding a protein or cosolvent molecules, showing as a non-zero μ_{32} or μ_{23}. In order to restore the equilibrium, the cosolvent has to be added or removed from the vicinity of the protein, and this quantity, which expresses the interactions of the protein with cosolvent, is defined as preferential binding parameter

$$\Gamma_{23} = \left(\frac{\partial m_3}{\partial m_2}\right)_{T,\mu_1,\mu_3} \tag{3.3}$$

A positive Γ_{23}, which is usually shown for protein denaturants, means a significant affinity for cosolvent molecules and protein prefers to bind with them and a negative Γ_{23}, which is usually shown for protecting osmolytes, means essentially protein has no affinity for cosolvent molecules and it predominantly contacts with water molecules.[39] These preferential interaction and preferential binding parameters could be obtained experimentally from equilibrium techniques, for example, dialysis equilibrium, isopiestic equilibrium, light scattering, X-ray scattering, or sedimentation equilibrium.[40-43] For molecular dynamics simulations, it could also be calculated through KB theory from radial distribution functions.[39]

3.2.2.4 Gibbs transfer free energy model

For more than 50 years, the Gibbs transfer free energy model has provided insight into protein folding principles.[44-48] In this model, total interactions between protein and cosolvent or water can be considered as the sum of all interactions between different parts of protein and cosolvent or water. Obviously, one of the major weaknesses of this model is the basic assumption that all of the transfer free energies of the amino acid side chains and the peptide backbone are additive. A number of studies have been performed to test the additivity, and much of the data support that the additivity occurs to a significant extent.[49-54] Besides the assumption, another approximation could affect the accuracy of this model. Instead of activities, concentrations are used to determine the transfer free energy, due to the difficulty to evaluate the activity coefficients for ternary systems.[55-57] Despite these assumptions and approximations, transfer free energy model has played a significant role in developing the concept of hydrophobic interactions and understanding the cosolvent effect on protein stability.[51,58-60] Bolen et al. used

cyclic glycylglycine, zwitterionic glycine peptides, and N-acetylglycine amide peptides as models for the peptide backbone of proteins, and tested the free energy change of transferring them from water to 1M concentrations of various denaturants, including trimethylamine-N-oxide, sarcosine, betaine, proline, glycerol, sortibol, sucrose, trehalose, and urea. Their experimental results showed that, with careful design of model compounds, proper execution of experiments and neglect of activity coefficients, a transfer free energy of the peptide backbone exhibited additivity, was independent of chemical model used, and was independent of the choice of concentration scale.[50]

3.2.3 Computational study of urea denaturation behavior

3.2.3.1 Previous calculations from free energy change

Preferential interactions between urea and protein are defined as the derivatives of the chemical potential of the solute to the urea concentration or activity.[1] Hence, the chemical potential changes and their corresponding derivatives are easily and usually chosen as the starting point to study preferential interactions. Previously, they were determined by particle insertion (PI) or thermodynamic integration (TI) techniques.

PI technique is a direct and widely used method to calculate chemical potential based on the thermodynamic identities with a large number of N particles:

$$\exp(-\beta\mu) = Q_{N+1}/Q_N = Q_N/Q_{N-1} \tag{3.4}$$

where Q_N is the partition function for N particles, and this is valid in both NVT and NpT ensembles. The excess chemical potential is defined as $\mu^{ex} = \mu - \mu^{id}$.[61] Adding a particle to the system randomly would induce a potential energy change and a volume fluctuation. Therefore,

$$\mu^{ex} = -k_B T \ln\left[\langle V \rangle^{-1} \langle V \exp(-\beta\Delta U) \rangle\right] \tag{3.5}$$

where ΔU is the change in energy upon insertion of the particle and V is the volume of the system.[62,63] However, PI calculation shows a serious limit that it only works well for small molecules at reasonable concentrations.

TI is another technique to calculate the free energy change, and it can be used for large molecules at high concentrations. The change of the system from state A to state B is described using a hybrid potential

$$U\left(r^N, \lambda\right) = \left(1-\lambda\right) U_A\left(r^N\right) + \lambda U_B\left(r^N\right) \quad (3.6)$$

where U is the potential of the system in corresponding states, r^N is the coordinates of N particles in the system, and λ is a coupling parameter, varied from 0 to 1.[64,65] The free energy change from state A to state B could be calculated from

$$\Delta G = \int_0^1 \left\langle \partial U / \partial \lambda \right\rangle_{\lambda_i} d\lambda \approx \sum_i \left\langle \Delta U \right\rangle_{\lambda_i} \Delta\lambda \quad (3.7)$$

Here the canonical average $\left\langle \partial U / \partial \lambda \right\rangle_{\lambda_i}$ is equal to $\left\langle \Delta U \right\rangle_{\lambda_i}$ which is provided the hybrid potential energy function $U\left(r^N, \lambda\right)$, linear in λ in equation (3.8). The major problem of TI calculation is the lack of precision and as a result it cannot be used to study the cosolvent effect on protein stability.

3.2.3.2 Kirkwood-Buff theory application to preferential interactions

Numbers of molecular dynamics simulations have been performed to study protein denaturation by urea.[90-93] Garcia et al. was the first group to calculate the changes in the preferential interaction coefficient of protein upon urea denaturation using the two-domain model[35]. They used a designed 20-residue Trp-cage miniprotein as a model system and examined the contributions from the backbone and side-chain groups over a wide range of urea concentration. In their results, the side-chain contribution was found to dominate in the increase of preferential interaction upon unfolding.[33,94]

Due to the difficulty of calculating the free energy change, a good method is really needed to quantify the preferential interactions. In the recent 10 years with the rapid development of computational power, analysis of preferential interactions using KB theory has shown a considerable increase in interest. In 2004, Shimizu et al. have calculated hydration changes for allosteric transitions and ligand binding through KB theory, and clarifies the relationship between osmotic and volumetric analysis.[66,67] In the next year, preferential interactions were expressed in terms of KB integrals by Schurr's group, and they were used to develop some simple models for the protein-cosolvent interactions.[68] Shulgin and co-workers have used KB theory to determine the excess or deficiency of water around several proteins with both protein protecting osmolytes and denaturants.[69] In 2006, Schellman compared the results from KB theory to the corresponding data obtained from thermodynamic binding models, and

were able to get a good agreement.[70] Our own group has studied to understand preferential interaction using KB theory for more than a decade, built a complete connection between simulated results and experimental thermodynamic data,[71] developed a model of cosolvent effects,[72] and derived the preferential interaction parameters for open, semi-open, and closed ensembles in terms of KB integrals.[73]

3.2.4 What next?

As shown above, although various methods are available to study cosolvent effects in biomolecular systems, the accurate detailed information at the atomic level is still required to describe and explain the real mechanism and thermodynamic properties. Because Kirkwood-Buff (KB) theory is an exact theory of solutions which plays a role in bridging both atomic details and the bulk thermodynamic properties, we use it as an effective tool to investigate the urea denaturation mechanism.

In this study, the cosolvent effect on peptide aggregation is investigated by molecular dynamics simulations. Urea is chosen as the cosolvent and several small solute molecules are used as simple models of amino acid side chains. Neo-pentane, benzene, and methanol are selected to mimic the hydrocarbon aliphatic, aromatic and amphiphilic moieties, respectively. Moreover, glycine, which is the simplest amino acid, is used to study the urea-zwitterionic amino acid interactions. The molecular distribution changes with the concentrations of solute and cosolvent were obtained and analyzed by Kirkwood-Buff (KB) theory. The chemical potential derivatives to the cosolvent concentrations and the coordination numbers in various solvation shells were also calculated.

3.3 Theory

3.3.1 Kirkwood-Buff theory

As described in detail at Chapter 1, KB theory has provided numerous thermodynamic properties of a solution mixture in terms of the KB integrals

$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}^{\mu VT}(r) - 1 \right] r^2 dr \qquad (3.8)$$

where G_{ij} is the KB integral between species i and j in the solution mixture, g_{ij} is the corresponding radial distribution function (rdf) in the μVT ensemble, and r is the distance from

the center of mass to another center of mass. KB integrals were calculated from simulation data in the NpT ensemble based on the assumption

$$G_{ij}(R) = 4\pi \int_0^R \left[g_{ij}^{NpT}(r) - 1 \right] r^2 dr \qquad (3.9)$$

where R defines a correlation region beyond which all rdfs are shown to be unity. This means the solution composition is the same as the bulk composition. Excess coordination numbers in KB theory, which describes local molecular distribution, can be obtained from

$$N_{ij} = \rho_j G_{ij} \qquad (3.10)$$

where a positive (or negative) N_{ij} indicates an excess (or depletion) of species j in the vicinity of species i over a random distribution.

The ternary system, which is more complicated than binary system, shows numerous interesting effects, including the cosolvent denaturation. In any system, we have

$$m_i = \frac{\rho_i}{\rho_{water}} = \frac{n_i}{n_{water}} \qquad (3.11)$$

$$\rho_i = \frac{n_i}{V} \qquad (3.12)$$

where m_i is the species molality, ρ_i is the number density, and n_i is the number of molecules of i in the volume V.

3.3.2 Transfer free energy calculations

Figure 3.1 The process of transferring a particle from a fixed position in aqueous solution to a fixed position in a cosolvent solution
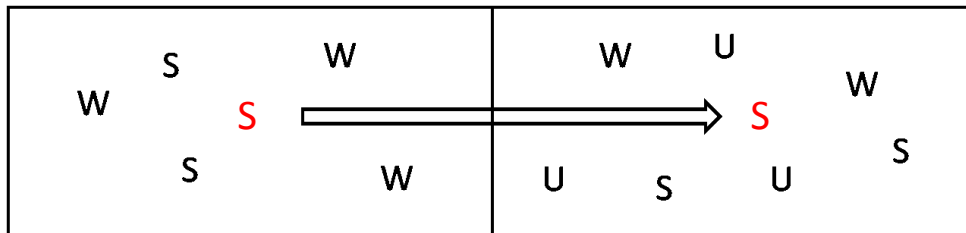
In the latter part of Chapter 3, the subscripts 1, 2, and 3 refer to water as primary solvent, several small solutes which represents different parts of proteins and urea as a cosolvent,

respectively. The transferring process used by Ben-Naim is shown in Figure 3.1.[74] A solute S is moving from a fixed position in aqueous solution to a fixed position in cosolvent solution. μ^* is referred to as pseudo-chemical potential (PCP), and the transfer free energy change at constant pressure P and temperature T is given by the difference in PCP's

$$\Delta G_{tr}^* = \mu_2^* \left(\rho_2, \rho_3 \right) - \mu_2^* \left(\rho_2, 0 \right) \tag{3.13}$$

It is worthy to mention that this is for fixed positions in each solution. Compared with experimental results, we note that

$$\beta\mu_2 = \beta\mu_{2,c}^o + \ln\left(\gamma_{2,c} c_2 / c_2^o \right)$$
$$= \beta\mu_2^* + \ln \Lambda_2^3 \rho_2 \tag{3.14}$$

where γ is activity coefficient, c is molecular concentration in molarity, and Λ refers to the thermal de Broglie wavelength of the species. Therefore,

$$\beta\mu_2^* = \beta\mu_{2,c}^o + \ln\left[\frac{\gamma_{2,c}}{\Lambda_2^3 c_2^o} \right] \tag{3.15}$$

where $\mu_{2,C}^o$ is the standard chemical potential and varies with ρ_3.

When ρ_2 is close to zero, $\gamma_{2,C}$ would be close to 1. Hence,

$$\beta\mu_2^{*,\infty} = \beta\mu_{2,c}^o - \ln\left(\Lambda_2^3 c_2^o \right) \tag{3.16}$$

which is to say $\beta\mu_{2,c}^o$ is equal to pseudo-chemical potential term at infinitely diluted condition and it is a constant. Consequently,

$$\beta\Delta G_{tr}^{*,\infty} = \beta\mu_2^{*,\infty}\left(\rho_3 \right) - \beta\mu_2^{*,\infty}\left(0 \right) = \beta\mu_{2,c}^o\left(\rho_3 \right) - \beta\mu_{2,c}^o\left(0 \right) \tag{3.17}$$

Here we should note that this is only true for $\mu_{2,C}^o$ on the molarity scale. Furthermore

$$\frac{\partial\beta\Delta G_{tr}^{*,\infty}}{\partial\rho_3} = \frac{\partial\beta\mu_2^{*,\infty}}{\partial\rho_3} = \frac{\partial\beta\mu_{2,c}^o}{\partial\rho_3} \tag{3.18}$$

3.3.3 Application of KB theory to ternary solution

From KB theory for ternary mixtures,[39,75,76] the solute molecule PCP's change could be calculated by

$$-d\beta\mu_2^* = \Gamma_{22}d\beta\mu_2 + \Gamma_{23}d\beta\mu_3 \tag{3.19}$$

where the preferential binding parameter, Γ_{23}, in terms of KB integrals can be obtained from

$$\Gamma_{23} = \rho_3 \left(G_{23} - G_{21} \right) \tag{3.20}$$

$$\Gamma_{22} = \rho_2 \left(G_{22} - G_{21} \right) \tag{3.21}$$

From equation (3.19), we are able to deduce the following equation

$$\left(\frac{\partial \beta \mu_2^*}{\partial \rho_3} \right) = -\Gamma_{22} \left(\frac{\partial \beta \mu_2}{\partial \rho_3} \right)_{T,P} - \Gamma_{23} \left(\frac{\partial \beta \mu_3}{\partial \rho_3} \right)_{T,P} = \left[-\Gamma_{22} \frac{\partial \beta \mu_2}{\partial m_3} - \Gamma_{23} \frac{\partial \beta \mu_3}{\partial m_3} \right] \frac{\partial m_3}{\partial \rho_3} \tag{3.22}$$

The KB theory for ternary system provided us

$$\frac{\partial m_3}{\partial \rho_3} = \frac{1}{\rho_1 \left(1 - \phi_3 \right)} \tag{3.23}$$

$$\frac{\partial \beta \mu_2}{\partial m_3} = \frac{-\rho_1 A_1}{\eta_{123}} \tag{3.24}$$

$$\frac{\partial \beta \mu_3}{\partial m_3} = \frac{1}{m_3} \frac{\eta_{12}}{\eta_{123}} \tag{3.25}$$

where a set of variables are defined as[77]

$$A_i = 1 + \rho_i \left(G_{ii} + G_{jk} - G_{ij} - G_{ik} \right) \tag{3.26}$$

$$\eta_{ij} = \rho_i A_j + \rho_j A_i = \rho_i \left(1 + N_{jj} - N_{ij} \right) + \rho_j \left(1 + N_{ii} - N_{ij} \right) \tag{3.27}$$

$$\eta_{123} = \rho_1 A_2 A_3 + \rho_2 A_1 A_3 + \rho_3 A_1 A_2 \tag{3.28}$$

$$1 - \phi_3 = \frac{\rho_1 A_2 \left[1 + N_{33} - N_{13} \right] + \rho_2 A_1 \left[1 + N_{33} - N_{23} \right]}{\eta_{123}} \tag{3.29}$$

Substituting these equations back, we are able to get

$$\mu_{23} = \frac{\partial \beta \mu_2}{\partial \rho_3} = \frac{\partial \beta \mu_2}{\partial m_3} \frac{\partial m_3}{\partial \rho_3} = -\frac{A_1}{\eta_{123}} \frac{1}{\left(1 - \phi_3 \right)} = -\frac{A_1}{\rho_1 A_2 \left[1 + N_{33} - N_{13} \right] + \rho_2 A_1 \left[1 + N_{33} - N_{23} \right]} \tag{3.30}$$

$$\mu_{33} = \frac{\partial \beta \mu_3}{\partial \rho_3} = \frac{1}{\rho_3} \frac{\rho_1 A_2 + \rho_2 A_1}{\rho_1 A_2 \left[1 + N_{33} - N_{13} \right] + \rho_2 A_1 \left[1 + N_{33} - N_{23} \right]} \tag{3.31}$$

And in a similar way,

$$\mu_{22} = \frac{1}{\rho_2} \frac{\rho_1 A_3 + \rho_3 A_1}{\rho_1 A_3 \left[1 + N_{22} - N_{12} \right] + \rho_3 A_1 \left[1 + N_{22} - N_{32} \right]} \tag{3.32}$$

$$\mu_{23} = -\frac{A_1}{\rho_1 A_3 \left[1 + N_{22} - N_{12}\right] + \rho_3 A_1 \left[1 + N_{22} - N_{32}\right]} \qquad (3.33)$$

In short, the derivation of transfer free energy change to the cosolvent concentration can be shown clearly as

$$\frac{\partial \beta \Delta G_{tr}^*}{\partial \rho_3} = \frac{\partial \beta \mu_2^*}{\partial \rho_3} = -\Gamma_{22}\mu_{23} - \Gamma_{23}\mu_{33} \qquad (3.34)$$

3.4 Methods

3.4.1 KBFF Models

The cosolvent urea and all of the solute molecules, including neo-pentane, benzene, glycine, and methanol were simulated using Kirkwood-Buff force field developed by the Smith group[78-80] with the SPC/E water model[81]. The urea cosolvent effects were studied at 0, 2, 4, 6, 8, 10 molalities with four different solutes. For glycine and methanol ternary systems, the solute concentrations were 0.5, 1.0, and 3.0 molalities. For neo-pentane and benzene systems, the solute concentrations were chosen to be 0.1, 0.5 and 1.0 molalities, considering their extremely low solubility in water. The size of the cubic box was 6 nm for each system.

3.4.2 Molecular dynamics simulations

All of the simulations were performed in the NpT ensemble at 300K and 1 atm using the GROMACS program.[82] The Berendsen weak coupling technique was adopted to modulate the temperature and pressure with relaxation times of 0.1 and 0.5 ps, respectively.[83] The time-step was 2 fs, and bond length was constrained using Lincs algorithm for non-water and SETTLE algorithm for water molecules.[84,85] In order to evaluate electrostatic interactions, the Particle-Mesh-Ewald technique (PME) was used.[86] A real space convergence parameter of 3.5 nm^{-1} was used in combination with the cutoff distances of 1.0 nm for Coulomb interaction and of 1.5 nm for Van der Waals interaction. The nonbonded update frequency was 10 steps. The initial cubic boxes of different solutions were generated by randomly inserting water, solute and cosolvent molecules to boxes until the expected concentration was obtained, and were followed by energy minimization using the steepest descent method. Extensive equilibrations were performed until all intermolecular potential energy contributions and molecular distributions displayed no change

with time. Configurations were saved every 1.0 ps for analysis. All systems were simulated for 100 ns, and the last 40 ns were used to calculate ensemble averages.

3.5 Result and Discussion

3.5.1 KB analysis

Figure 3.2 Simulated center-of-mass radial distribution functions (g_{ij}) as a function of distance (r) for the neo-pentane-urea systems.

PEN-urea
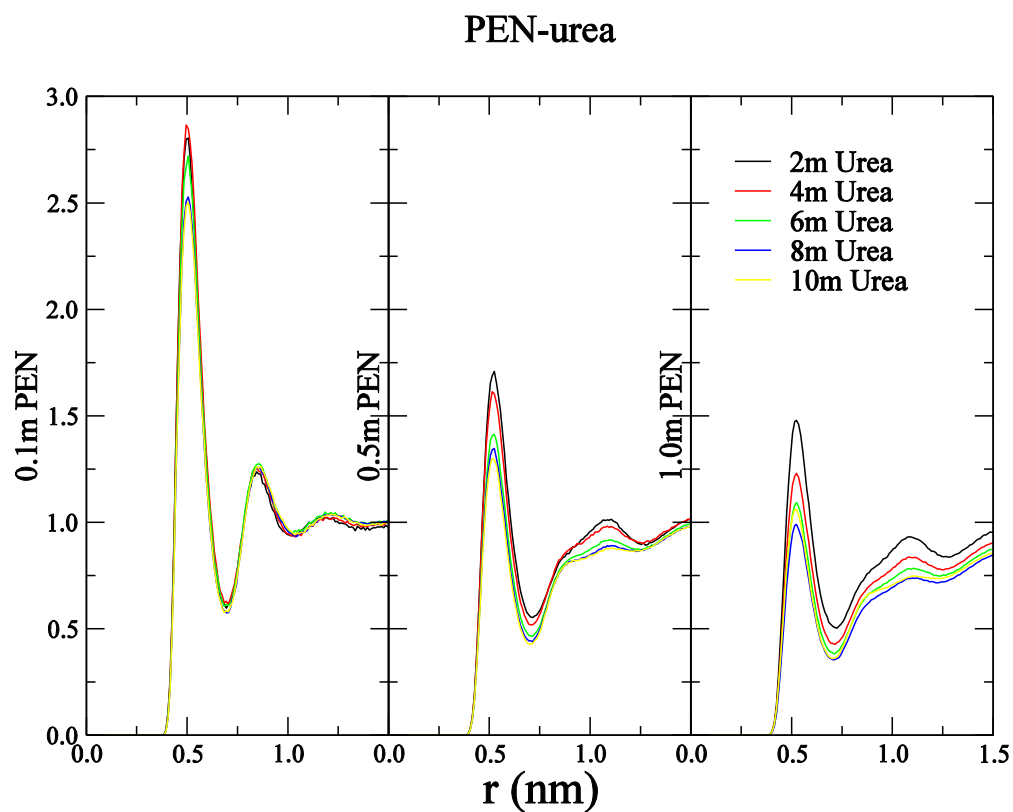
H$_2$O-urea
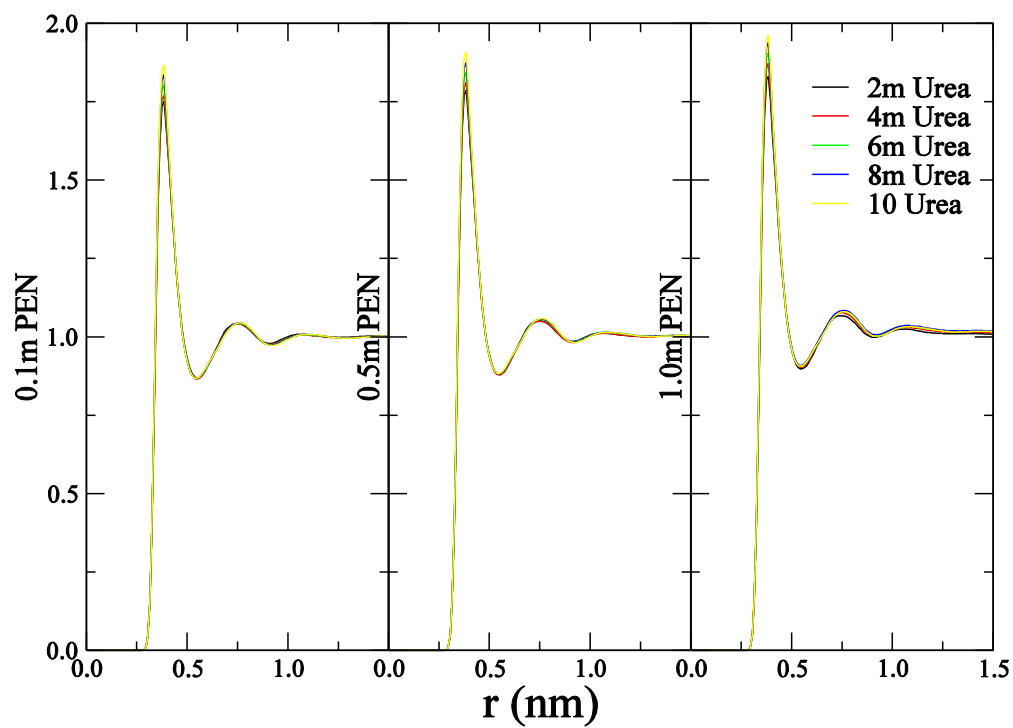
PEN-H$_2$O
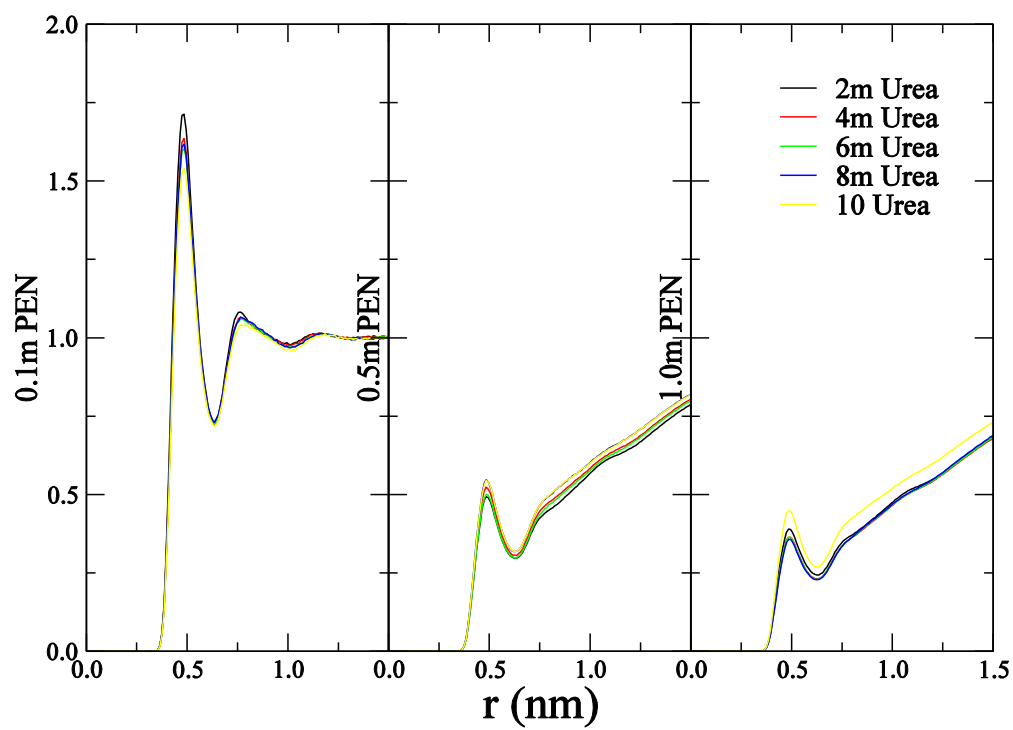
Figure 3.2 shows the center of molecular mass radial distribution functions (rdf) with a series of neo-pentane and urea molal concentrations. With the increase of neo-pentane concentration, the first peaks, which indicate the interaction within the first shells, for neo-pentane—urea and neo-pentane—water decrease dramatically, especially the neo-pentane—water peaks. In contrast, the first peaks for urea—water do not exhibit any visible change with different concentrations. Meanwhile, at the same concentration of neo-pentane and urea, the peak for neo-pentane—urea is significantly higher than that for neo-pentane—water, which means in the vicinity of neo-pentane, urea molecules are higher in number than water molecules. In solutions of 0.1 m neo-pentane, rdfs of neo-pentane—water converge to unity immediately beyond the solvation shells, which indicates that the systems were in metastable state and no phase separation arises. This could be attributed to the cosolvent effect of urea because the concentration of 0.1 m neo-pentane is highly above its solubility (about 0.0012 M in water at 25 ^{o}C). In the solutions of 0.5 m and 1.0 m neo-pentane, the rdfs of neo-pentane—water show continuous and dramatic increase beyond the first shells, which means the phase separations appear in these concentrations of neo-pentane.

The center of mass rdfs for benzene ternary solutions are displayed in Figure 3.3. Compared with neo-pentane systems, the urea-water rdf peaks are almost identical as concentration changed, and benzene—urea and benzene—water peaks decrease with increasing benzene molality. However, these decreases are only noticeable at highest benzene concentration, and more gently than those of neo-pentane systems. Like in neo-pentane systems, the benzene—urea peaks are much higher than benzene-water ones. The peaks of benzene—water interactions grow up with the increasing urea concentrations, which do not happen in neo-pentane systems. The rapid converges in benzene—water rdfs with the concentration of 0.1 m and 0.5 m suggest that these systems are in metastable states (the solubility of benzene in water at 25 ^{o}C is 0.023 M), and the cosolvent effect of urea works in a wider concentration range of benzene than that of neo-pentane. The phase separation appears at very high benzene concentrations (1.0 m). Considering the phase separation problem, further analysis, including preferential binding and transfer free energy derivatives, will not be shown for 0.5 m and 1.0 m neo-pentane and 1.0 m benzene systems.

Figure 3.3 Simulated center-of-mass radial distribution functions (g_{ij}) as a function of distance (r) for the benzene-urea systems

BEN-urea

H$_2$O-urea
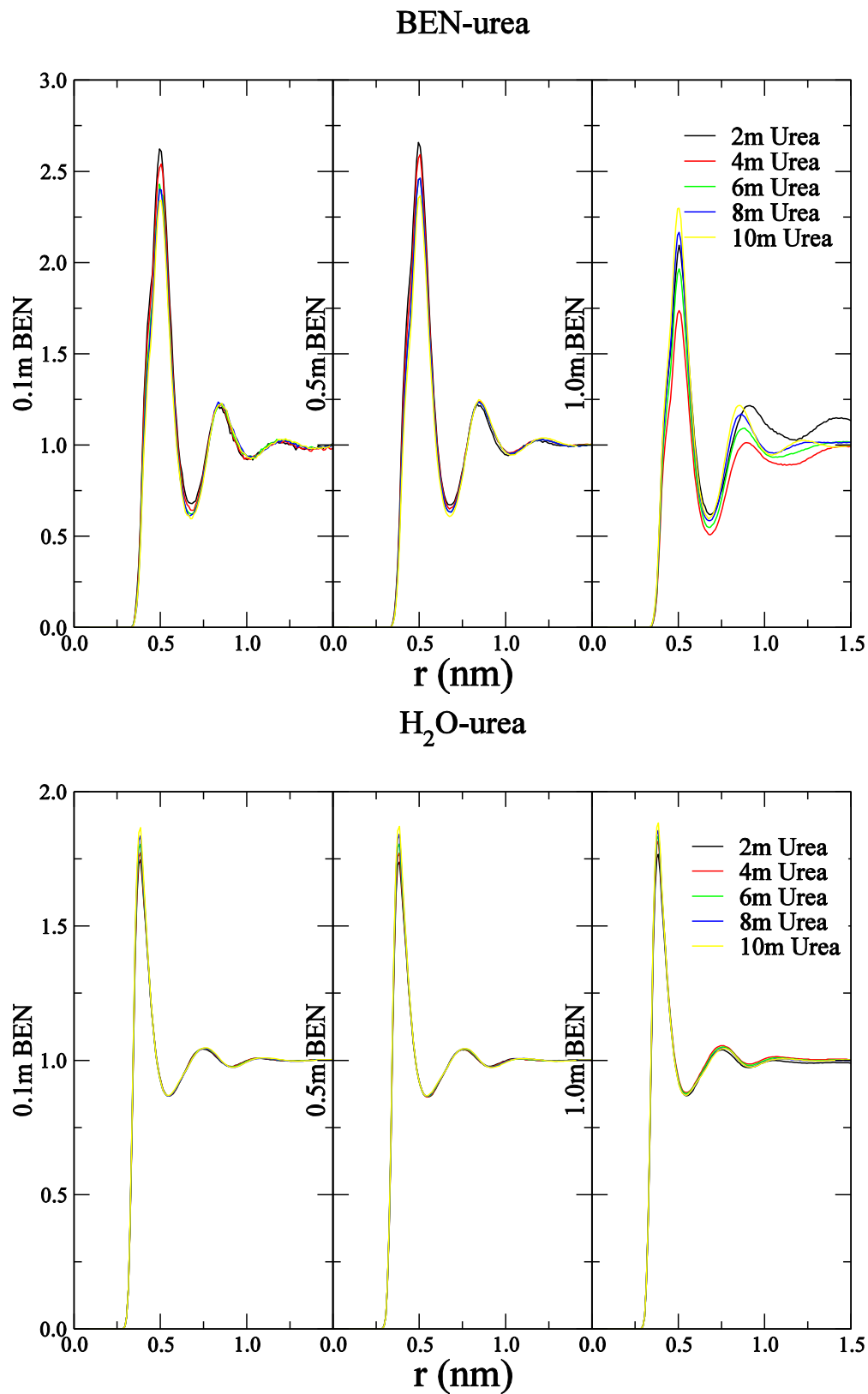
BEN-H₂O
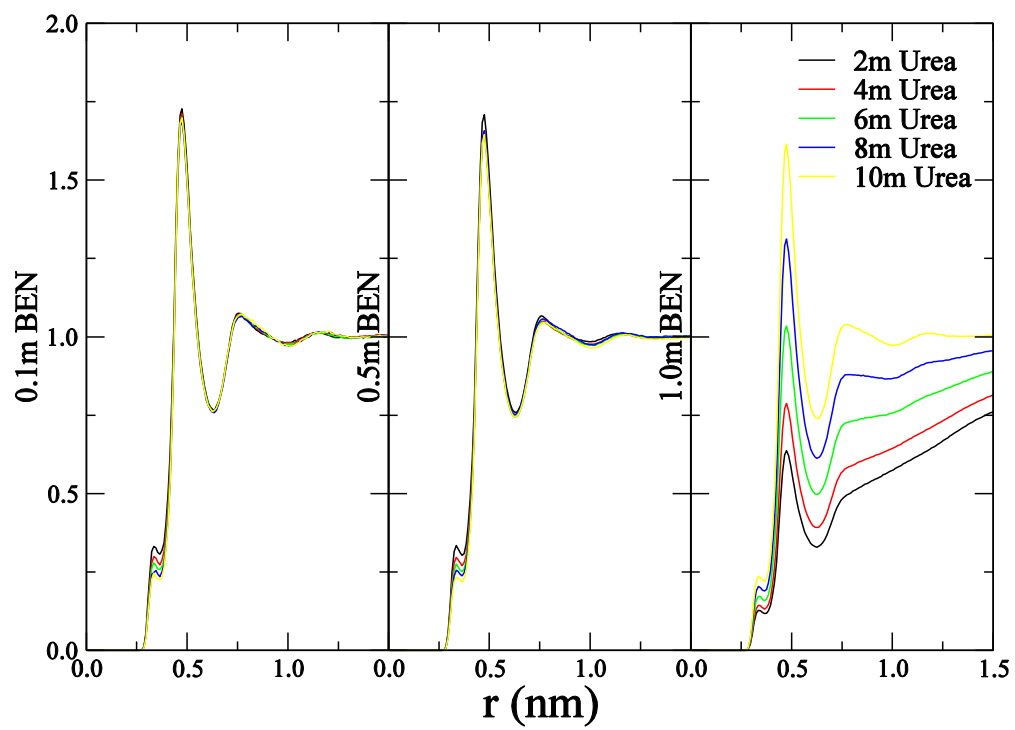
Figure 3.4 Simulated center-of-mass radial distribution functions (g_{ij}) as a function of distance (r) for the glycine-urea systems

Gly-urea

H$_2$O-urea
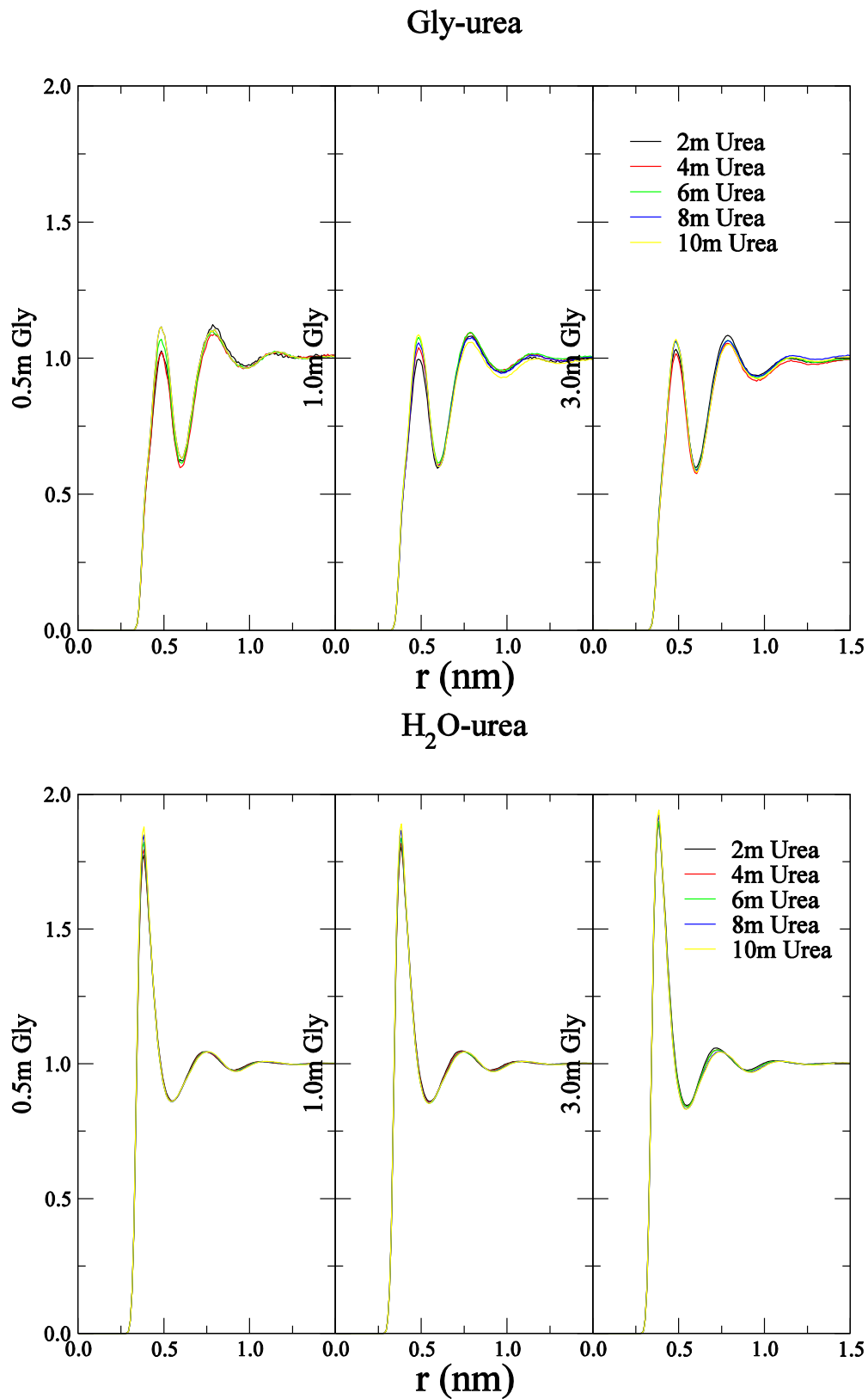
Gly-H$_2$O
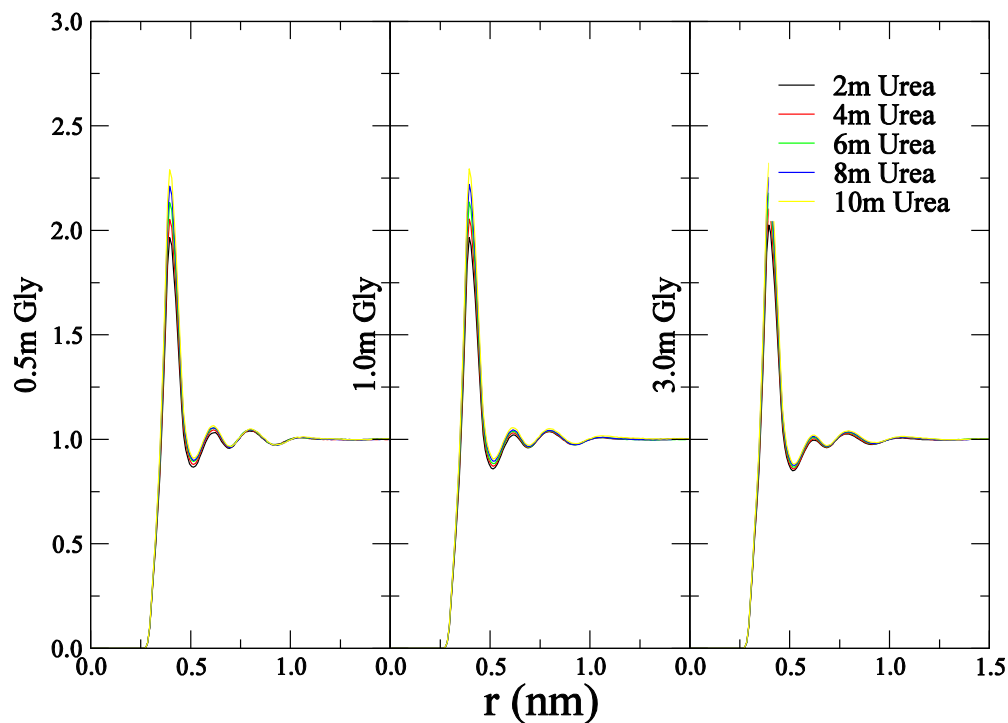
Figure 3.4 shows the center of molecular mass rdfs obtained for glycine mixtures. It is very easy to notice that the first peaks of glycine—urea interactions are much lower than those of glycine—water interactions, and much lower than those in the hydrophobic solute systems as well. It appears that a lot of water molecules contact with glycine, but urea molecules do not. Different from before, the concentration of urea or glycine does not seem to have big effect on any of center of mass radial distribution functions.

Figure 3.5 Simulated center-of-mass radial distribution functions (g_{ij}) as a function of distance (r) for the methanol-urea systems

MOH-urea

H$_2$O-urea
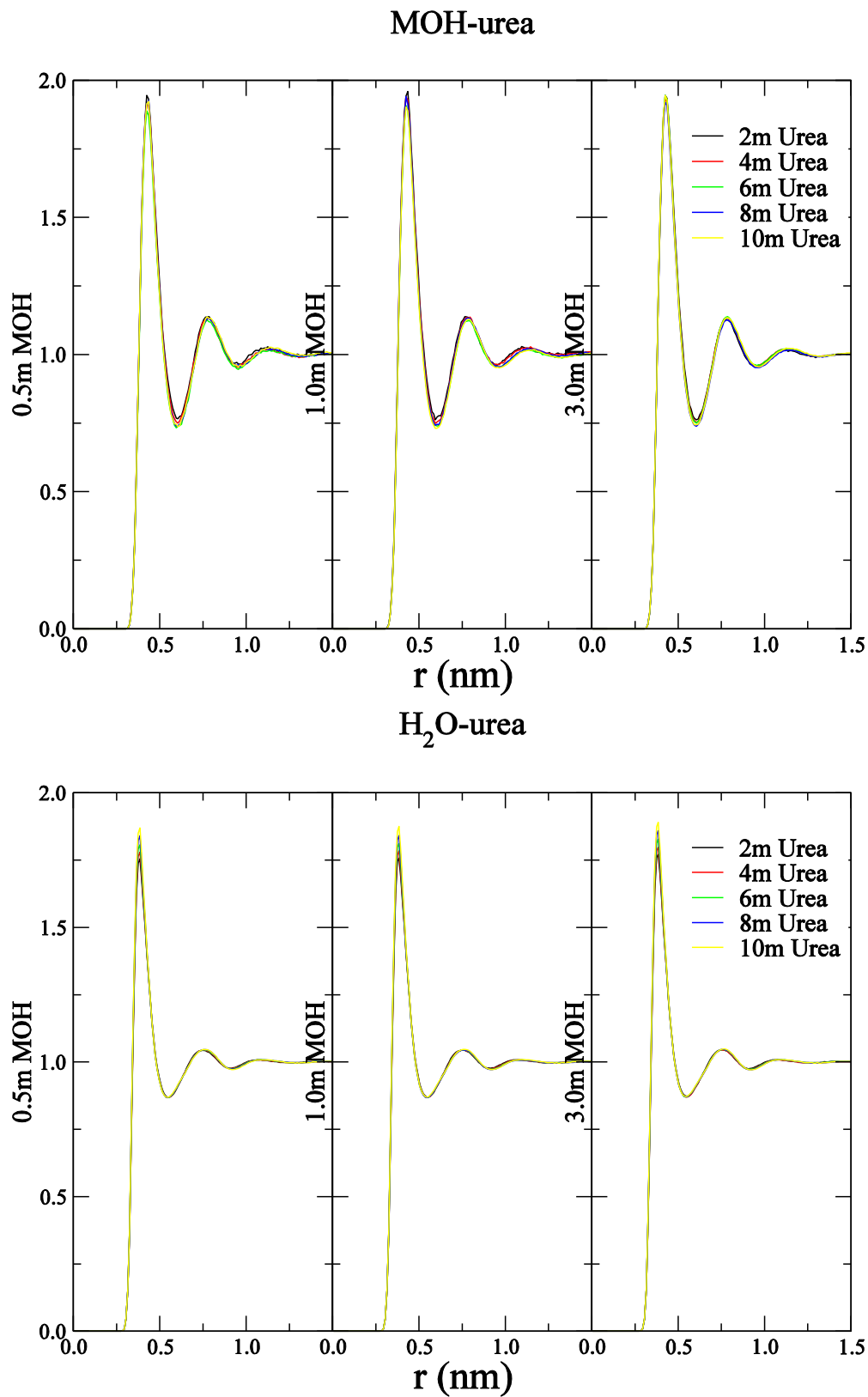
MOH-H$_2$O
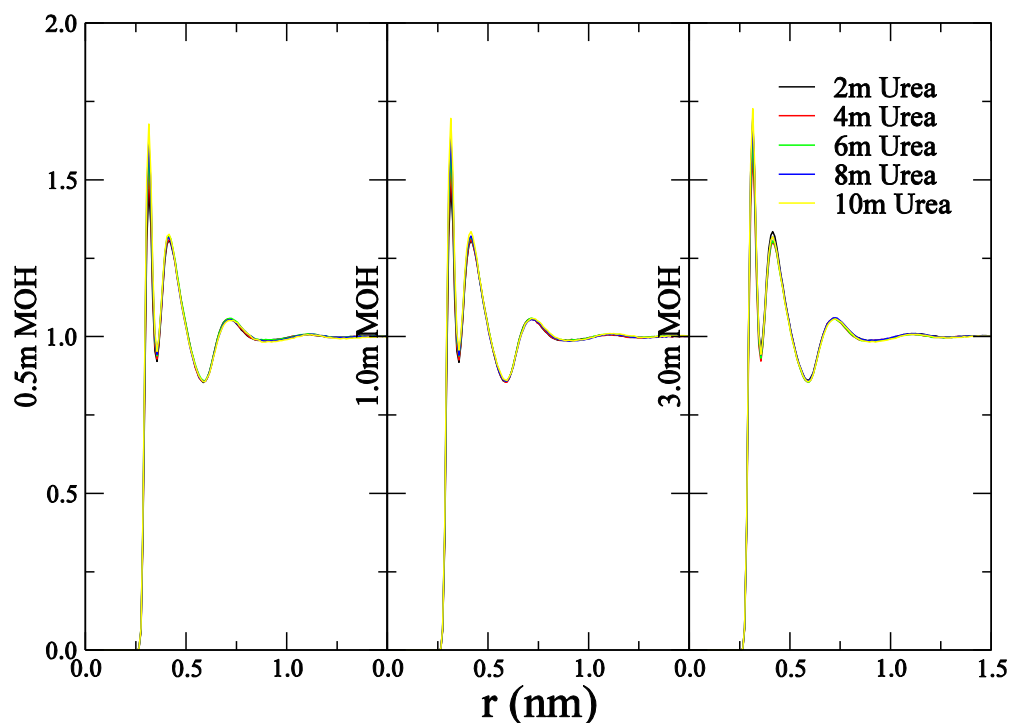
In Figure 3.5 for methanol ternary systems, the first peaks for methanol—urea interactions are slightly higher than those for methanol—water interactions, which means urea molecules exist more, but not dominant around methanol molecules. Similar to zwitterionic glycine systems, the rdfs of the amphiphilic methanol do not show any visible changes with concentration of urea or methanol.

All the rdfs shown above converged to unity at 3 nm, and displayed serious fluctuations before 1.5 nm. Therefore, the region between 0.9 nm and 1.2 nm was chosen to gather the rdf data and obtain KB integrals. As mentioned in Chapter 1, a positive integral generally means a net attraction between the two species, and negative integral means repulsion between them.

Figure 3.6 Simulated KB integrals (G_{ij}, cm^3/mol) as a function of urea molality for neo-pentane-urea systems
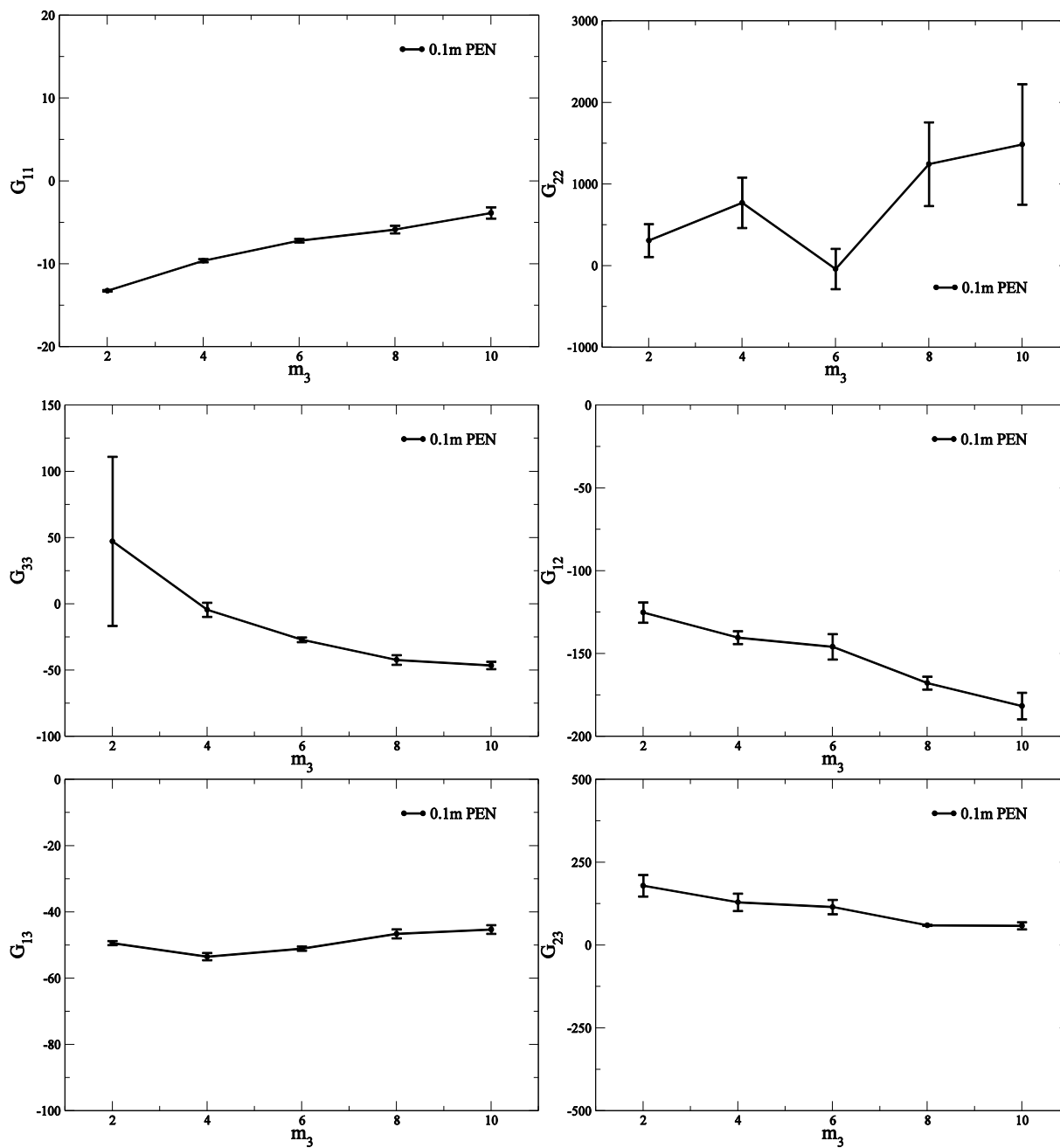
Figure 3.7 Simulated KB integrals (G_{ij}, cm^3/mol) as a function of urea molality for benzene-urea systems.
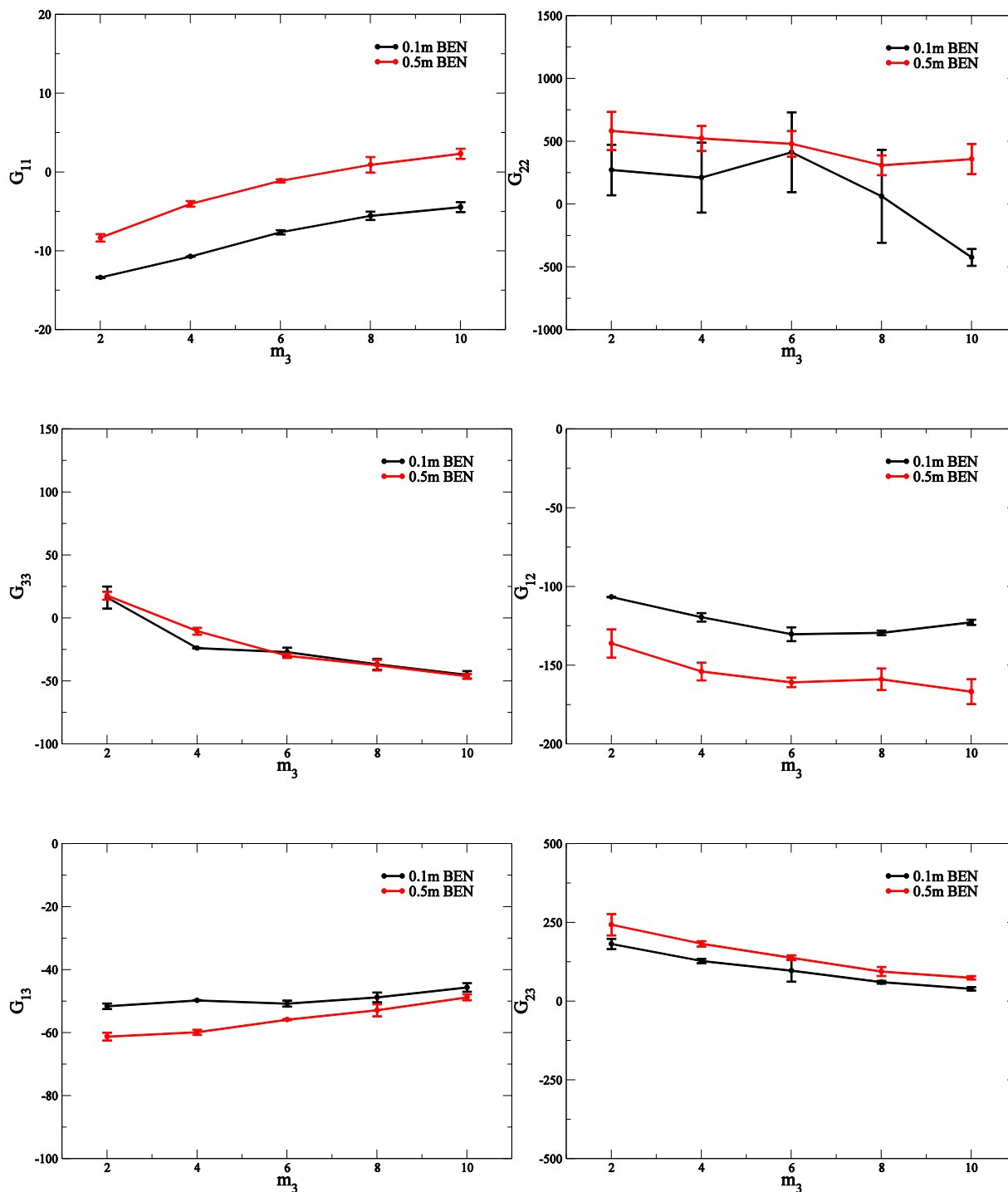
Figure 3.8 Simulated KB integrals (G_{ij}, cm^3/mol) as a function of urea molality for glycine-urea systems
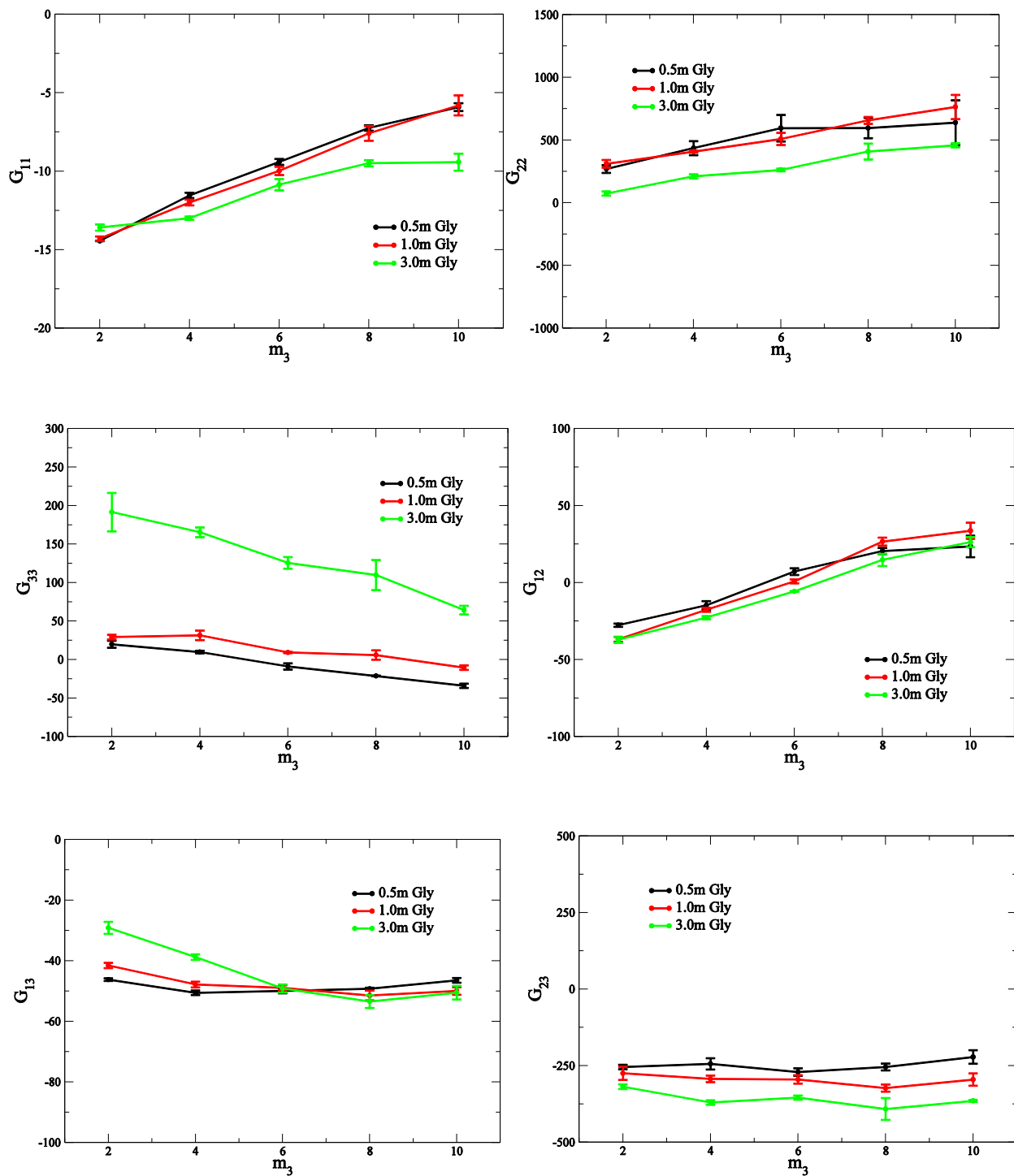
Figure 3.9 Simulated KB integrals (G_{ij}, cm^3/mol) as a function of urea molality for methanol-urea systems
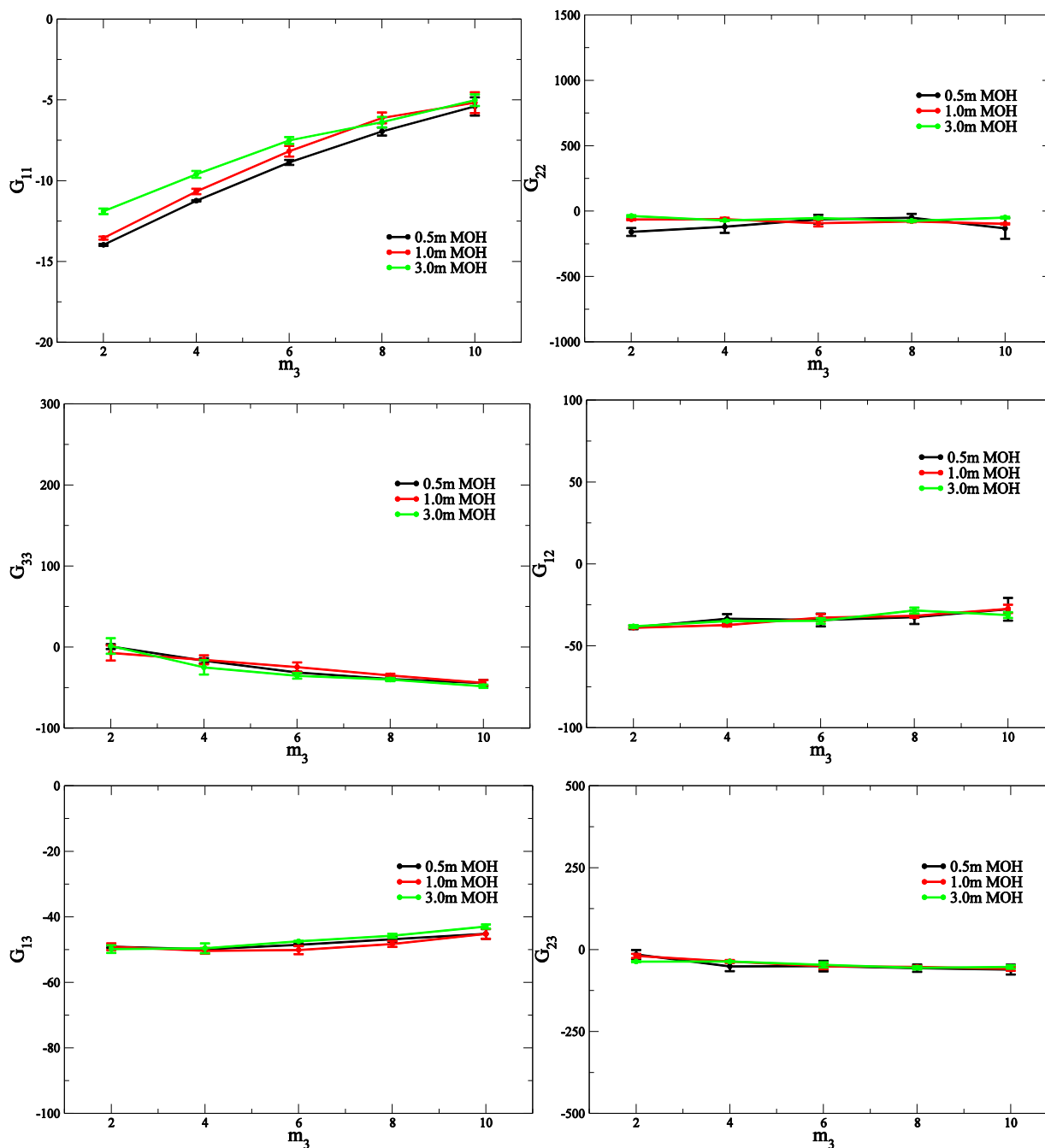
The KB integrals between various species in different systems are shown in Figure 3.6–Figure 3.9. Here we only focus on the concentrations of 0.1 m neo-pentane, 0.1 m and 0.5 m benzene solutions, which are in the metastable state, KB integrals of neo-pentane—neo-pentane

or benzene—benzene show large positive numbers with big error bars. In glycine systems, the solute-solute integrals change from positive numbers a negative numbers as the urea concentration increases, which means urea has a positive effect on the interaction between glycine and water. On the other hand, the methanol-urea KB integrals are nearly zero at all concentrations of urea and methanol.

The urea-urea KB integrals decrease when the urea concentration increases and solute concentration decreases. These changes are very obvious for hydrophobic and zwitterionic solute systems, but negligible for the methanol system.

The solute—water interactions in neo-pentane systems and benzene systems show huge unfavorable repulsions and the unfavorable repulsion become more obvious with increasing urea concentration. In contrast, the solute-water interaction is small for glycine and methanol systems and displays a slight increase as glycine concentration increases.

The water-urea KB integrals are similar and close to zero for all these systems. Generally, the solute-urea KB integrals decrease with increasing solute concentrations and the trend is very noticeable for benzene and glycine solutions, but invisible for methanol solutions.

3.5.2 Chemical potentials and preferential binding

In general, the transfer free energy could be expressed using any concentration scale with quantitative and qualitative changes. The mole fraction is often used theoretically because it was the basis of the original development of Raoult's law, and it is convenient for the calculation of symmetrical ideal solutions in the whole composition range. The molality scale is widely used in experiments because it is easy to measure weight of each component experimentally.[78,87] Moreover, the molality indicates the ratio of a solute to a solvent, or of a cosolvent to a solvent, clearly. Therefore, the KB integrals shown above are presented as a function of molal concentration. However, the molarity (or number density) scale has been proved to be the most useful and meaningful way to calculate standard quantities for transferring processes. [88] The molar concentration implicitly includes temperature and pressure effects, which could affect the chemical potential and consequently, the density. Hence, we calculated and exhibited the derivatives of chemical potential as functions of molar concentration.

All of the simulated preferential interaction parameters shown in Figure 3.10—Figure 3.13 are calculated as a combination of the above KB integrals using equations (3.30)-(3.33).

According to their mathematical definitions, a positive preferential interaction parameter μ_{ij} indicates the chemical potential of component i is increasing with the concentration of component j and component j destabilizes component i, and a negative μ_{ij} indicates chemical potential is decreasing and component j stabilizes component i.

Figure 3.10 Simulated chemical potential derivatives (μ_{ij}, cm^3/mol) as a function of neopentane and urea molarity
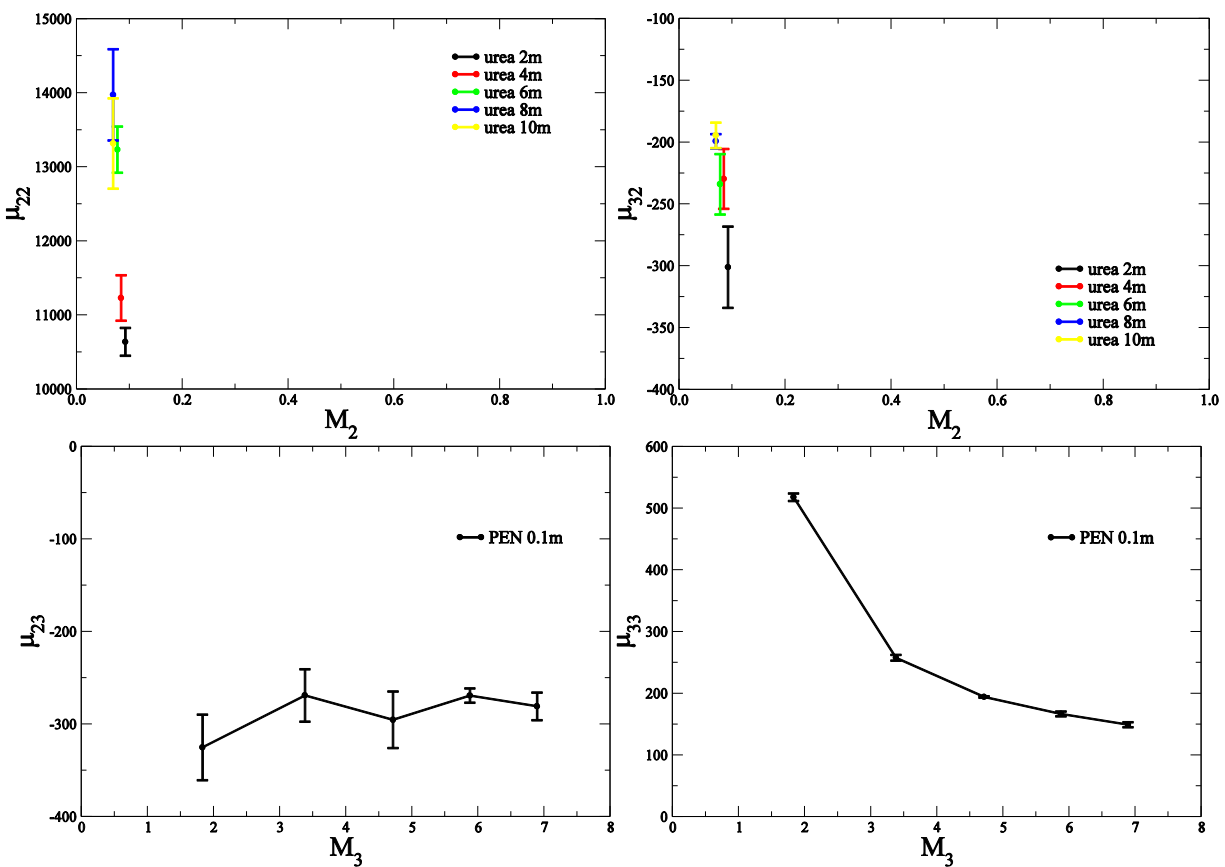
Figure 3.11 Simulated chemical potential derivatives (μ_{ij} cm^3/mol) as a function of benzene and urea molarity
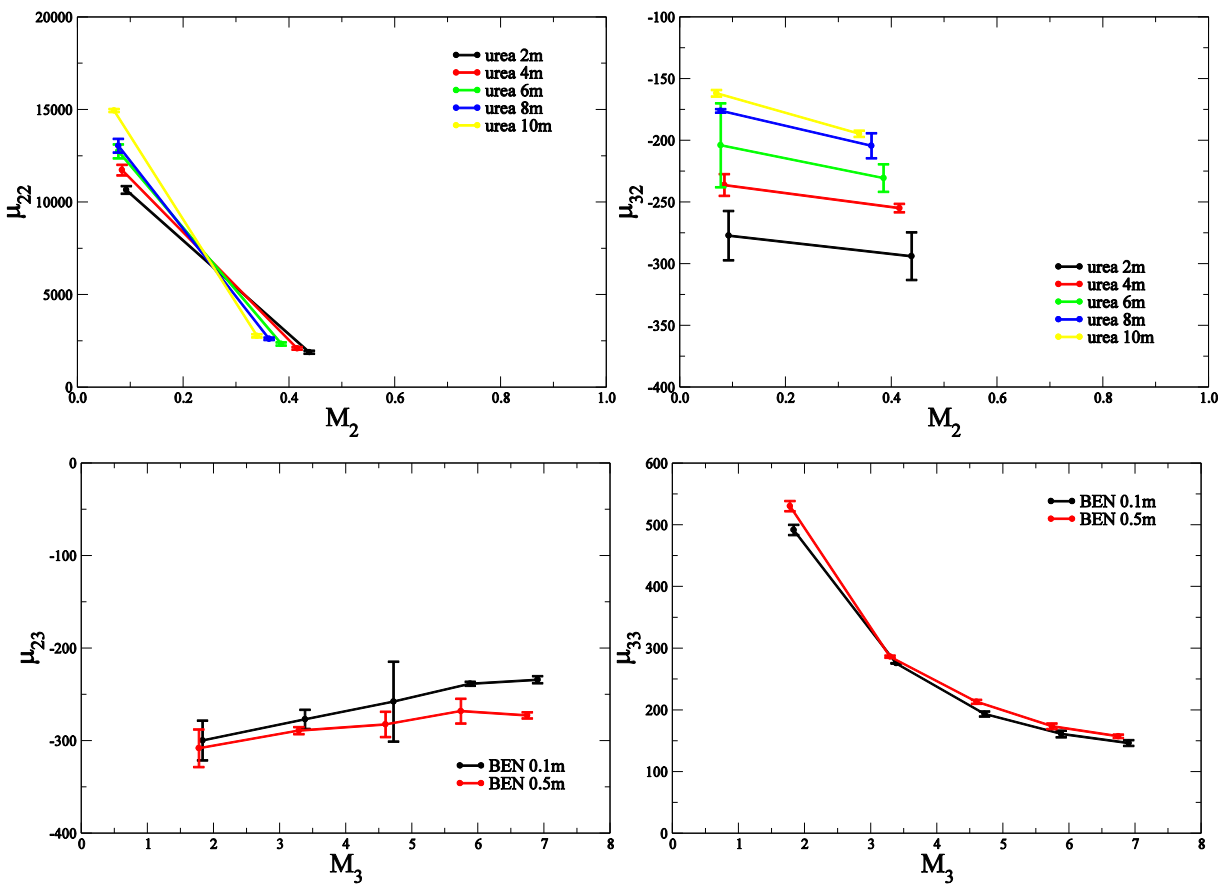
Figure 3.12 Simulated chemical potential derivatives (μ_{ij}, cm^3/mol) as a function of glycine and urea molarity
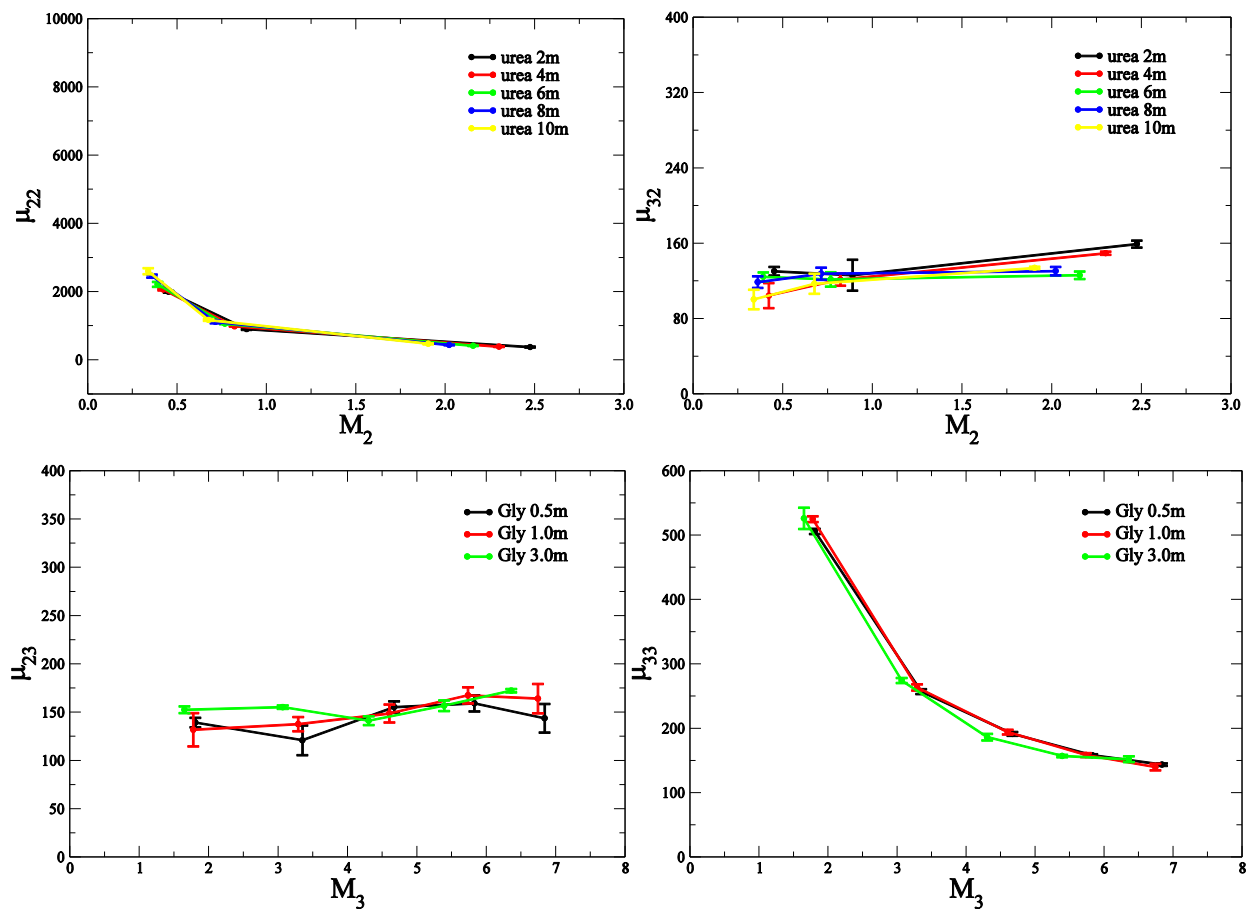
Figure 3.13 Simulated chemical potential derivatives (μ_{ij}, cm^3/mol) as a function of methanol and urea molarity
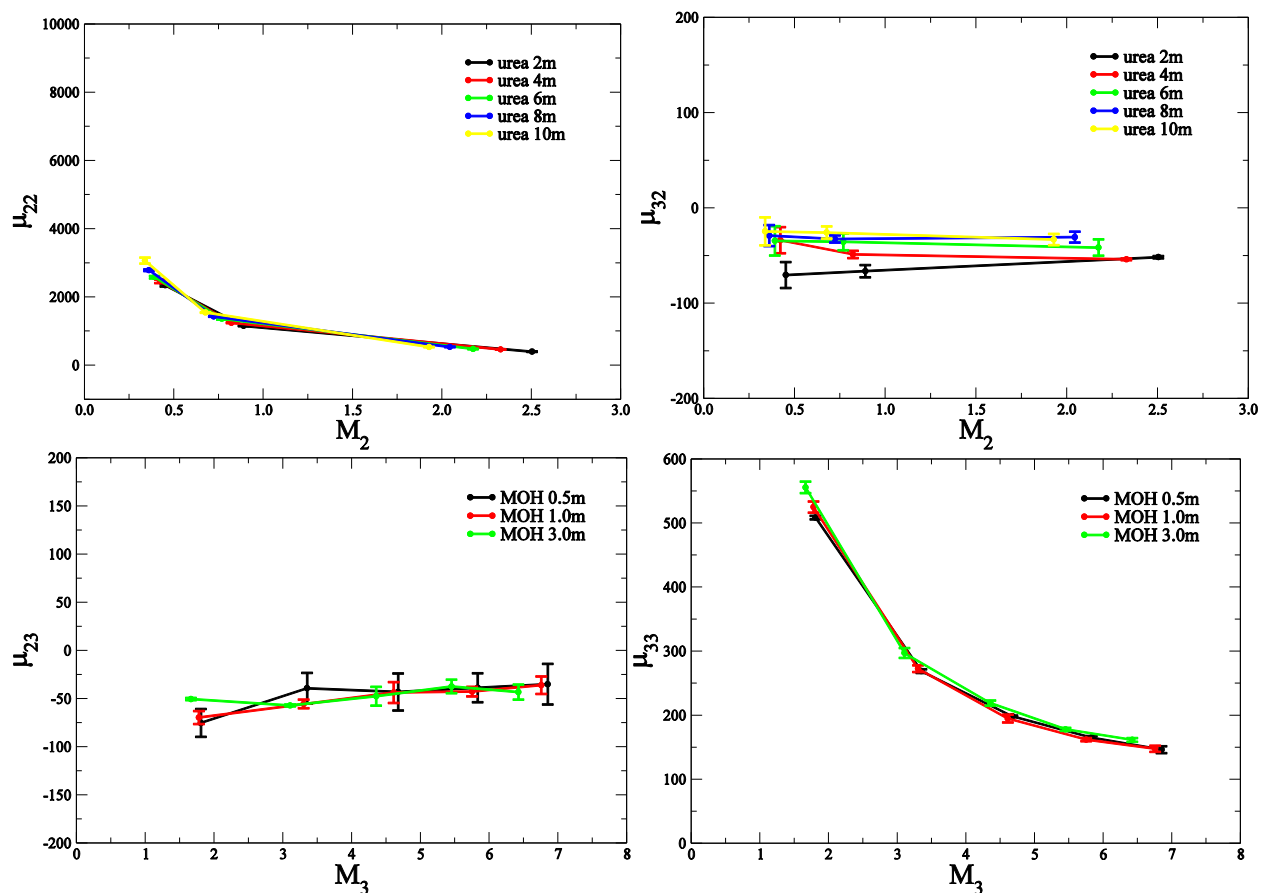
All of the four systems show positive μ_{22}, and the values of μ_{22} in benzene, glycine and methanol decrease as the solute concentrations increases. The decreases appear seriously in benzene systems, while gently in glycine and methanol systems. For neo-pentane solutions and benzene solutions, μ_{32} shows negative numbers and grow up with the increasing urea concentration. In benzene systems, adding benzene decreases the negative μ_{32}. This means adding benzene molecules decreases the chemical potential of urea. μ_{32} is a small positive and a small negative number for glycine and methanol solutions, respectively, and does not change much with solute or urea concentrations.

The neo-pentane systems and benzene systems have negative μ_{23}, which means urea stabilizes these hydrophobic molecules in an aqueous environment. For benzene systems, μ_{23} shows more negative with the increasing benzene concentrations, which indicates that the

stabilization effect increases as the solute gets concentrated. μ_{23} for glycine and methanol solutions does not have any consistent changes with solute or urea concentrations. All these systems show positive μ_{33}, and they decrease as urea concentration increases. This means urea is distablized when it is added to any of these solutions, especially at low urea concentrations.

The preferential binding parameters Γ in Figure 3.14—Figure 3.17 are calculated by equations (3.20) and (3.21). As mentioned before, the cosolvent, which is urea here, acts as a denaturant and the solute prefers to be surrounded by it if Γ_{23} is positive, while the cosolvent acts as protecting osmolytes, and solute prefers to be surrounded by water if Γ_{23} is negative.

Figure 3.14 The simulated preferential binding parameter (Γ_{ij}) and preferential structure parameter (Γ_{ij}/ρ_j, L/mol) for urea and neo-pentane as a function of urea molarity
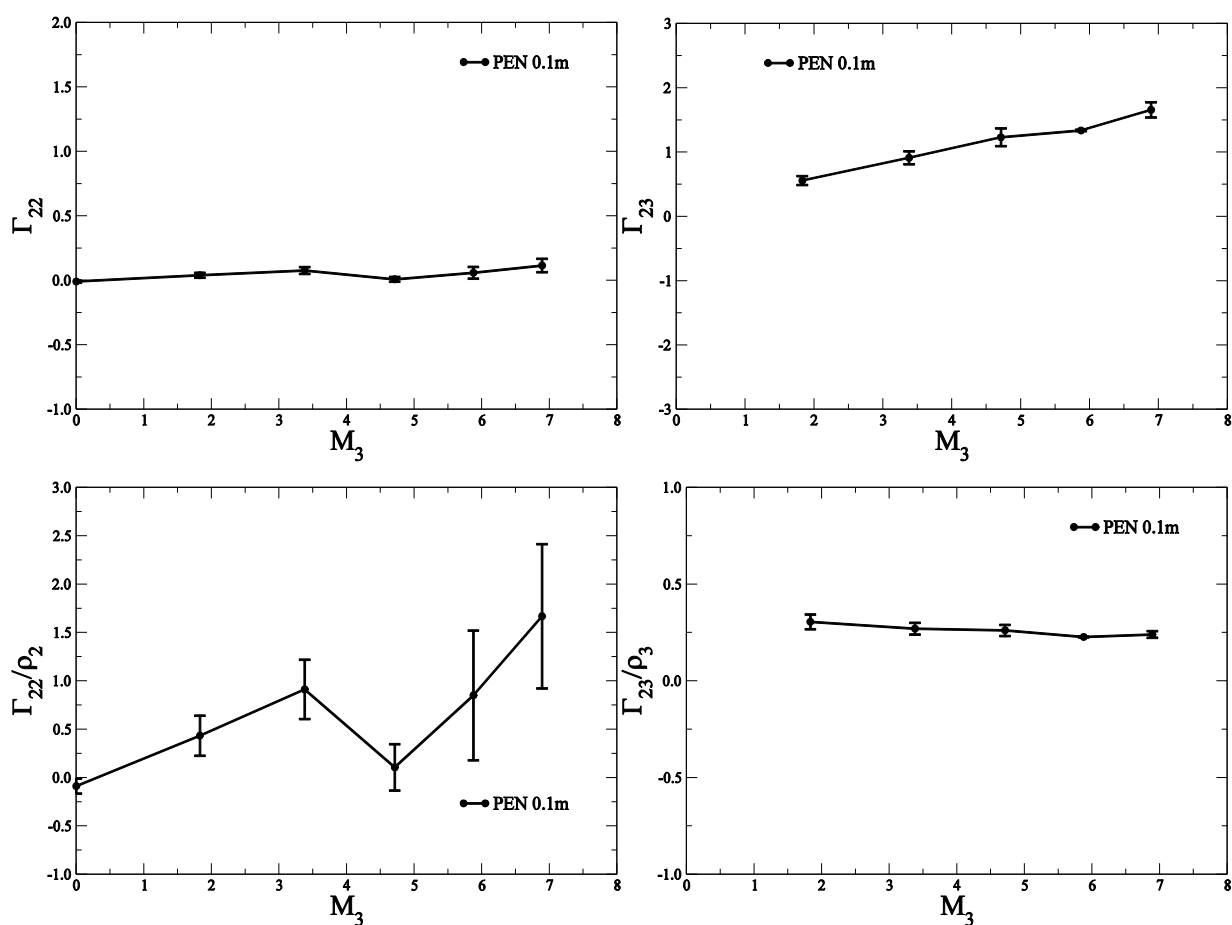
Hydrophobic solute solutions show positive Γ_{22} and Γ_{22} increases with increasing benzene concentration, which means they prefer to bind with themselves rather than water. The small

positive Γ_{22} for glycine solutions indicate that the glycine molecules have a tendency to bind with themselves and the trend becomes obvious with the increasing glycine and urea concentrations. However, methanol Γ_{22} do not display any visible change with the concentration of either methanol or urea.

Figure 3.15 The simulated preferential binding parameter (Γ_{ij}) and preferential structure parameter (Γ_{ij}/ρ_j, L/mol) for urea and benzene as a function of urea molarity
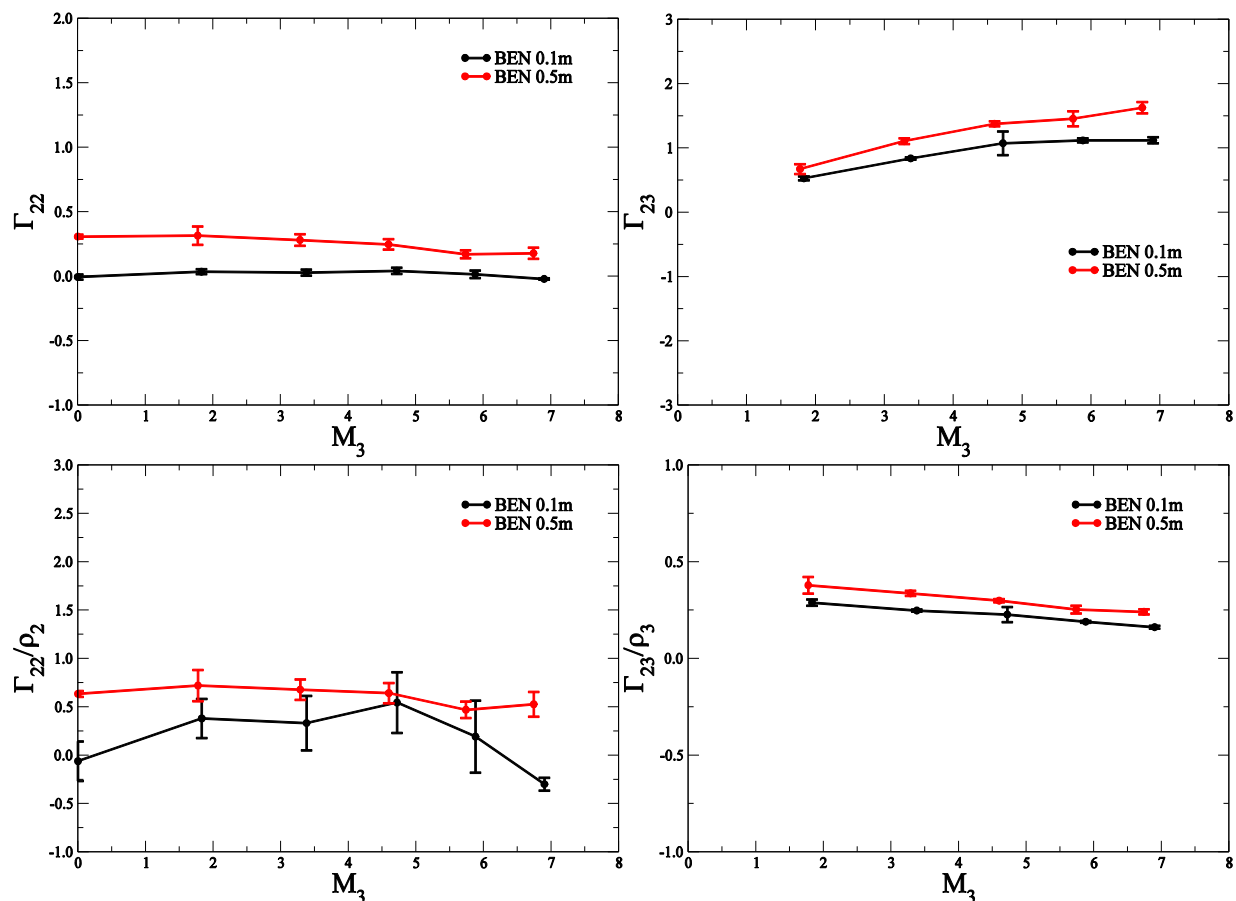
Γ_{23} for neo-pentane and benzene are positive and increase with increasing urea concentration. However, it shows small negative numbers and decreases with increasing urea concentration for glycine solutions. Γ_{23} for methanol is close to zero with a very slight downward trend, which indicates that urea does not have any obvious cosolvent effect on methanol.

According to equations (3.20) and (3.21), the preferential binding parameter includes the contributions from both the number densities and the relative distributions. In order to investigate

the real reason for the preferential binding and avoid the contributions from the number densities, we defined the preferential structure parameter as $\Gamma_{ij}/\rho_j = G_{ij} - G_{i1}$ to focus on the relative distributions around the solute molecules. When Γ and Γ/ρ are compared, these figures do not display any dramatic changes. As expected, the hydrophobic molecules show larger preferential structure parameters Γ_{23}/ρ_3 than glycine or methanol. Therefore, these different preferential binding could be attributed to different interactions between species, rather than the concentration of the components.

Figure 3.16 The simulated preferential binding parameter (Γ_{ij}) and preferential structure parameter (Γ_{ij}/ρ_j, L/mol) for urea and glycine as a function of urea molarity
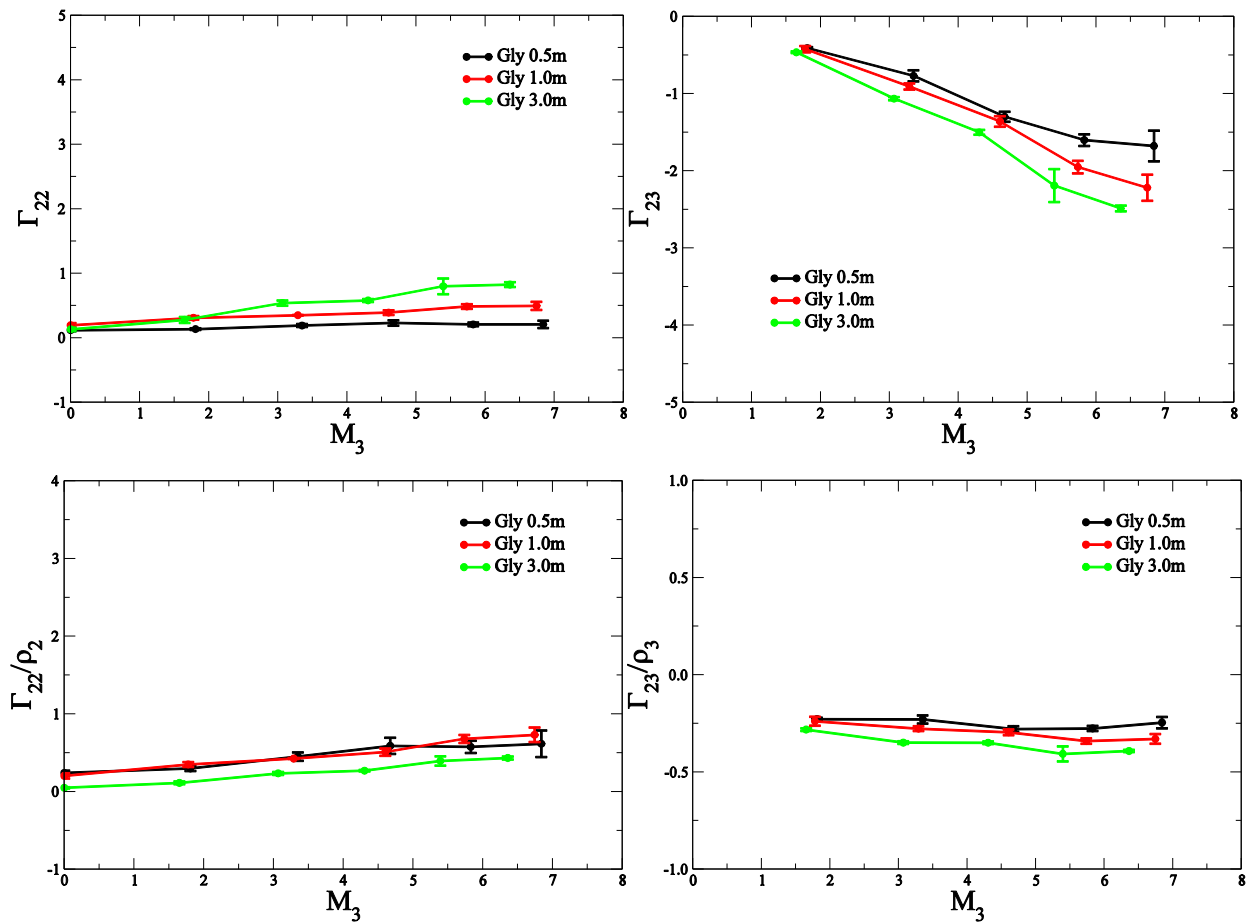
Figure 3.17 The simulated preferential binding parameter (Γ_{ij}) and preferential structure parameter (Γ_{ij}/ρ_j, L/mol) for urea and methanol as a function of urea molarity
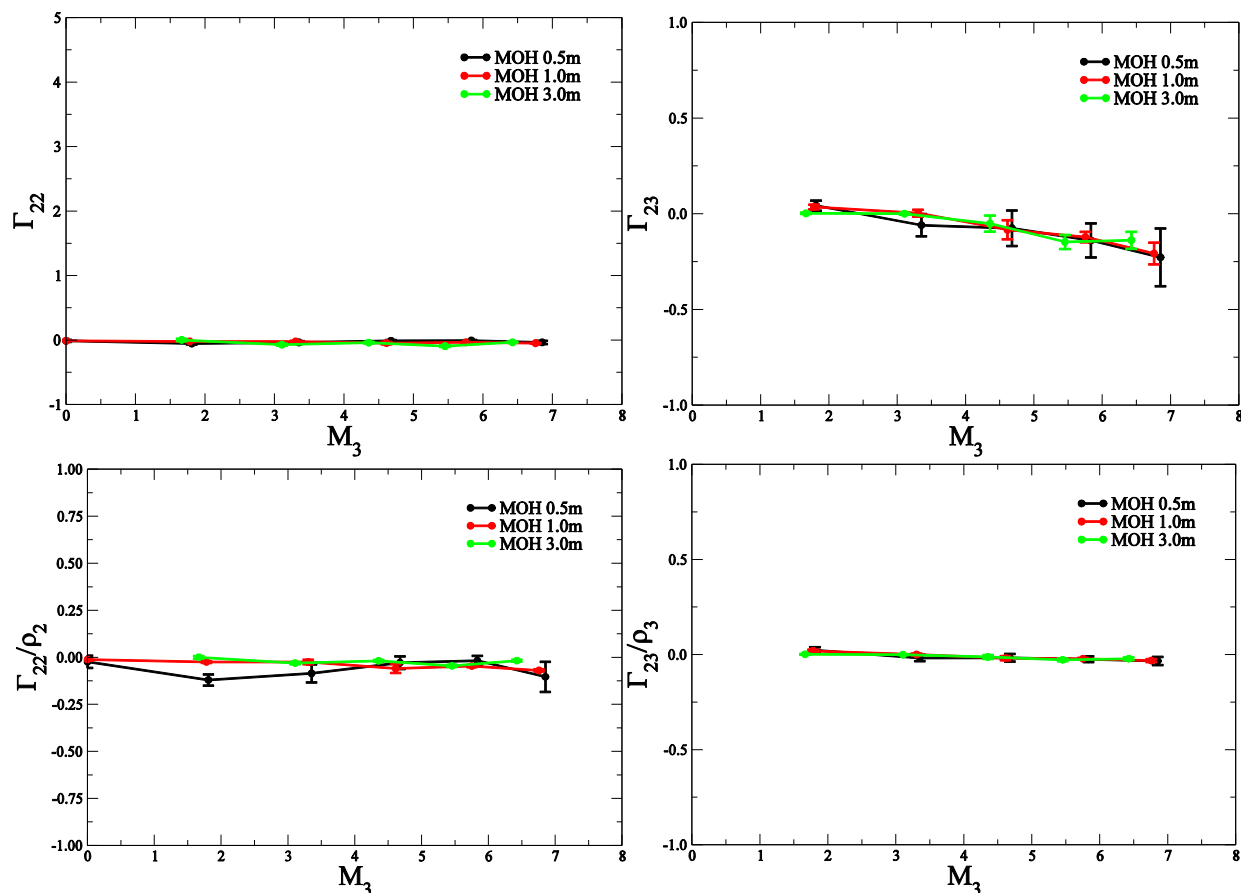
3.5.3 Contributions to the preferential binding

The systems with the lowest, middle, and highest urea concentrations are selected to compare the rdf and preferential binding to urea for all solute concentrations, and the results are shown in Figure 3.18-3.21.

For neo-pentane and benzene solutions in Figure 3.18 and Figure 3.19, Γ_{23} is almost zero within the excluded region, and only shows small peak in the first shell. In the low concentrations without phase separation, the preferential binding parameters are small positive numbers and increase as the urea concentration increases, which indicate that diluted neo-pentane slightly prefers to stay with urea than water in various urea concentrations.

84

Figure 3.18 Simulated center of mass rdf g_{23} (solid) and g_{21} (dash) at different pentane-urea molality composition corresponding to respective Γ_{23} as a function of distance
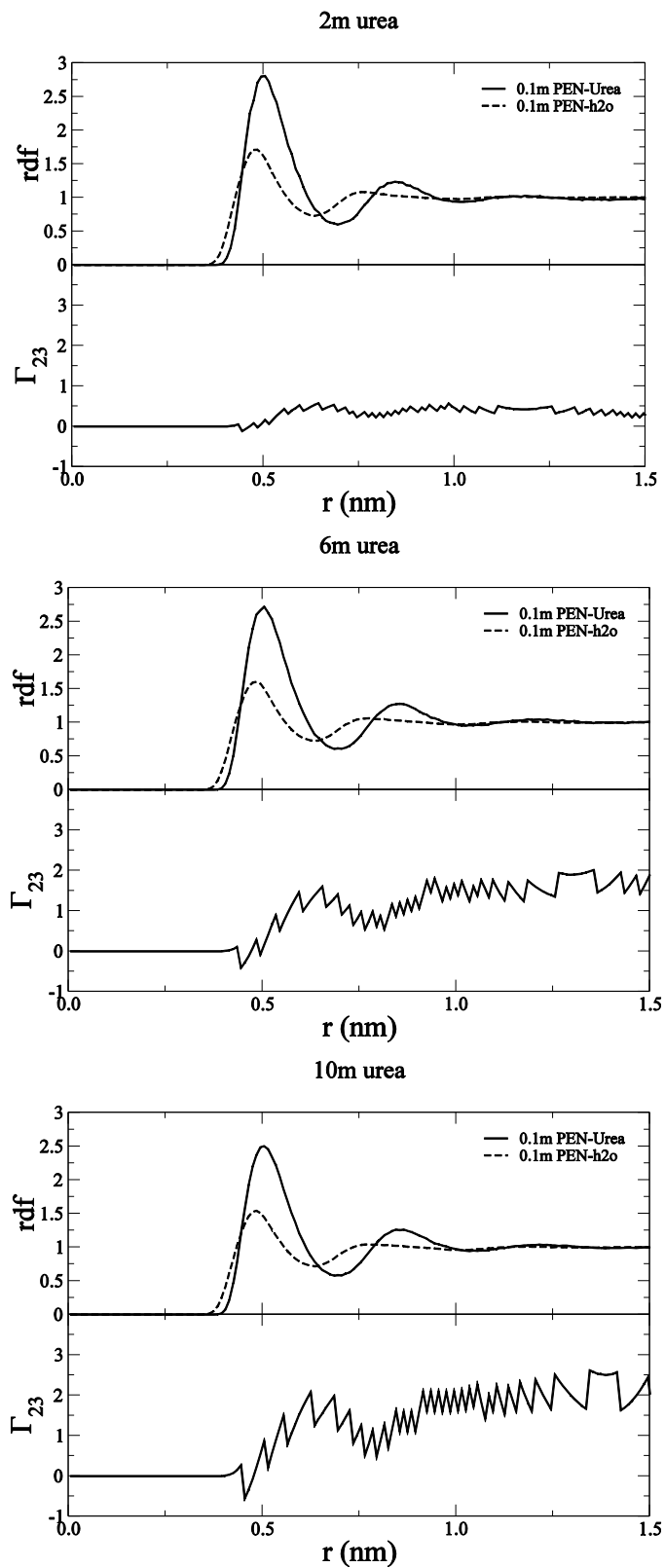
Figure 3.19 Simulated center of mass rdf g_{23} (solid) and g_{21} (dash) at different benzene-urea molality composition corresponding to respective Γ_{23} as a function of distance r.
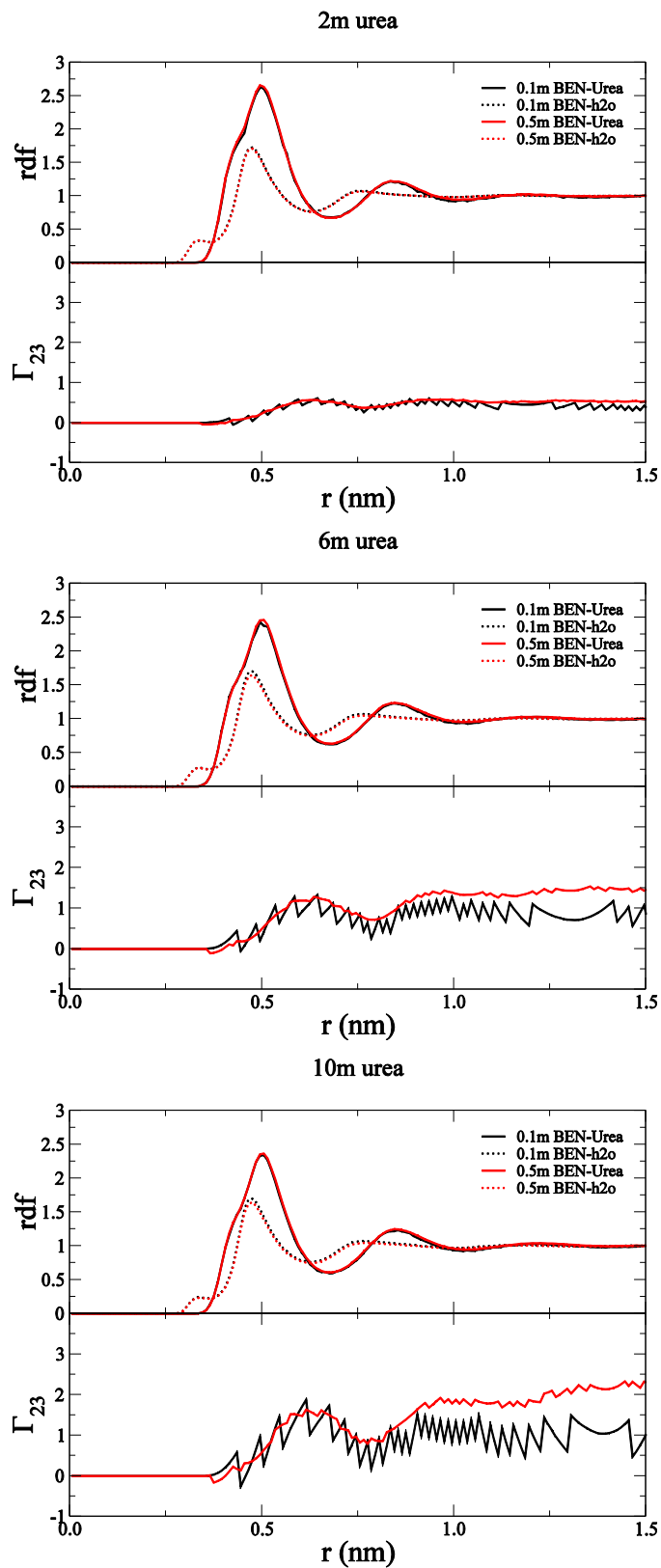
Figure 3.20 Simulated center of mass rdf g$_{23}$ (solid) and g$_{21}$ (dash) at different glycine-urea molality composition corresponding to respective Γ_{23} as a function of distance r.
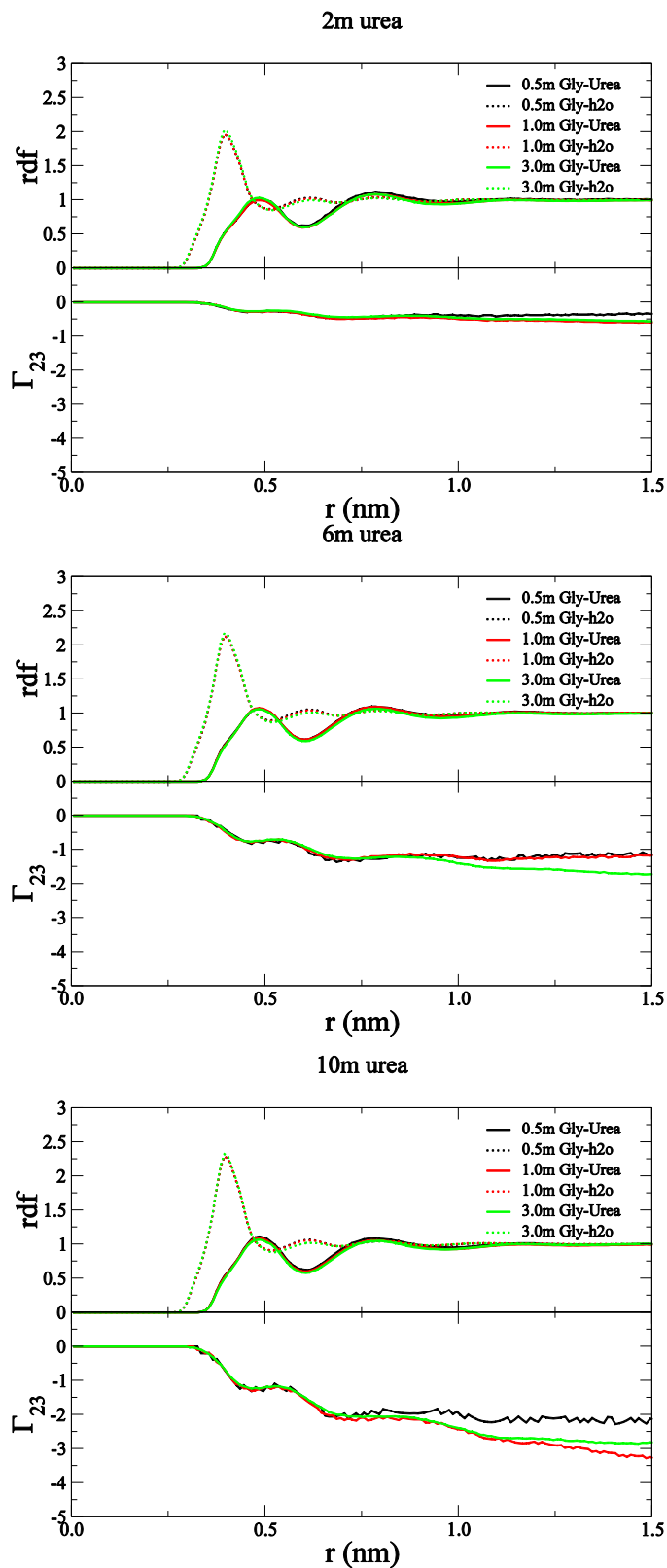
Figure 3.21 Simulated center of mass rdf g$_{23}$ (solid) and g$_{21}$ (dash) at different methanol-urea molality composition corresponding to respective Γ_{23} as a function of distance r.
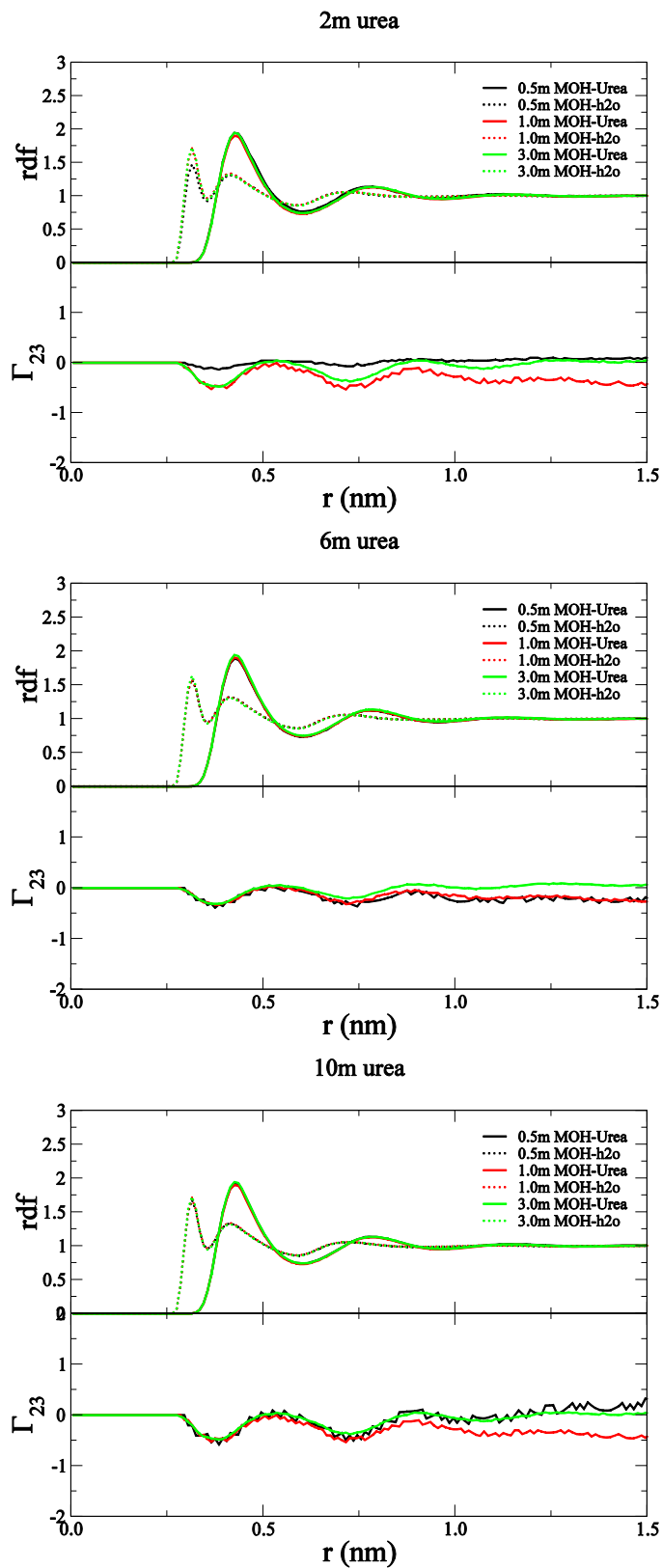
Opposite to previous hydrophobic solute systems, Γ_{23} of glycine systems shown in Figure 3.20 start to decrease in the first solvation shell. As the rdfs go through the second and third solvation shells, Γ_{23} decrease stage by stage, and those stages become more distinct with the increasing urea concentration. This indicates that at high urea concentration, glycine prefers to bind with water molecules. Different from zwitterionic glycine systems, Γ_{23} for methanol in Figure 3.21 display two clear valleys, and the values of Γ_{23} for methanol are small negative numbers.

The preferential structure parameter voiding the concentration contribution is decomposed into each solvation shell. According to equation (3.20), Γ_{23}/ρ_3 could be expressed as $G_{23} - G_{21}$, and $G_{23}-G_{21}>0$ indicates solute prefers to stay with urea, while $G_{23}-G_{21}>0$ means it prefers to stay with water. These values are displayed as functions of urea concentration in Figure 3.22—Figure 3.25.

Figure 3.22 Separation of preferential structure parameters (G_{23}-G_{21}, cm^3/mol) from pentane-urea aqueous system into four terms as function of urea molality according to the relative radial distribution functions. A) r =0~0.385 nm when g_{23} keeps at zero. B) r =0.385~0.705 nm when g_{23} is in 1st shell. C) r =0.705~1.275 nm is in 2nd shell. D) r =1.275~3.000 nm when g_{23} is in the extra part.
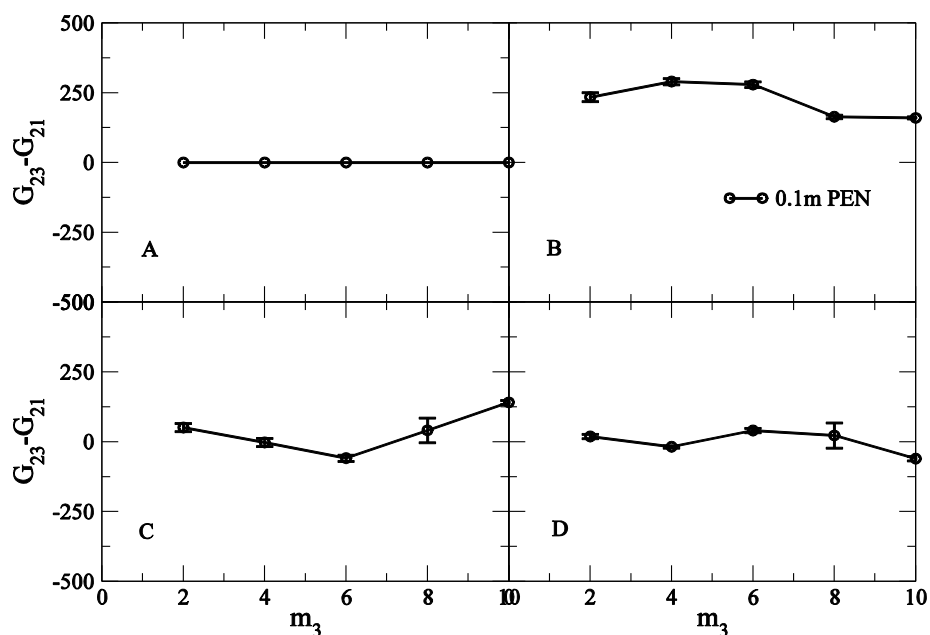
Figure 3.23 Separation of preferential structure parameters (G_{23}-G_{21}, cm^3/mol) from benzene-urea aqueous system into four terms as function of urea molality according to the relative radial distribution functions.   A) r =0~0.335 nm when g_{23} keeps at zero. B) r =0.335~0.675 nm when g_{23} is in 1st shell. C) r =0.675~1.045 nm is in 2nd shell. D) r =1.045~3.000 nm when g_{23} is in the extra part.
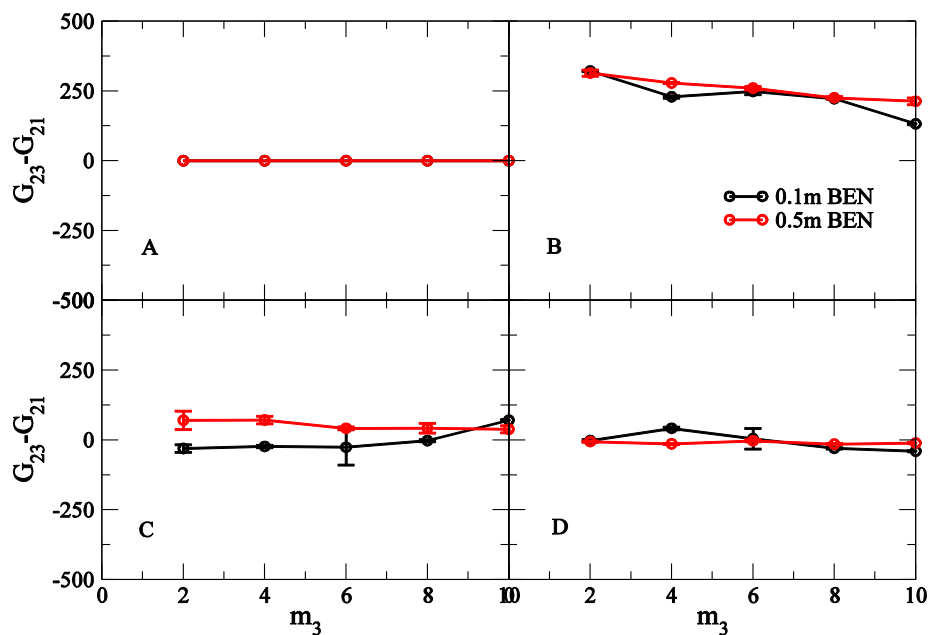
Figure 3.24 Separation of preferential structure parameters (G_{23}-G_{21}, cm^3/mol) from glycine-urea aqueous system into four terms as function of urea molality according to the relative radial distribution functions.   A) r =0~0.325 nm when g_{23} keeps at zero. B) r =0.325~0.605 nm when g_{23} is in 1st shell. C) r =0.605~0.965   nm is in 2nd shell. D) r =0.965~3.000 nm when g_{23} is in the extra part.
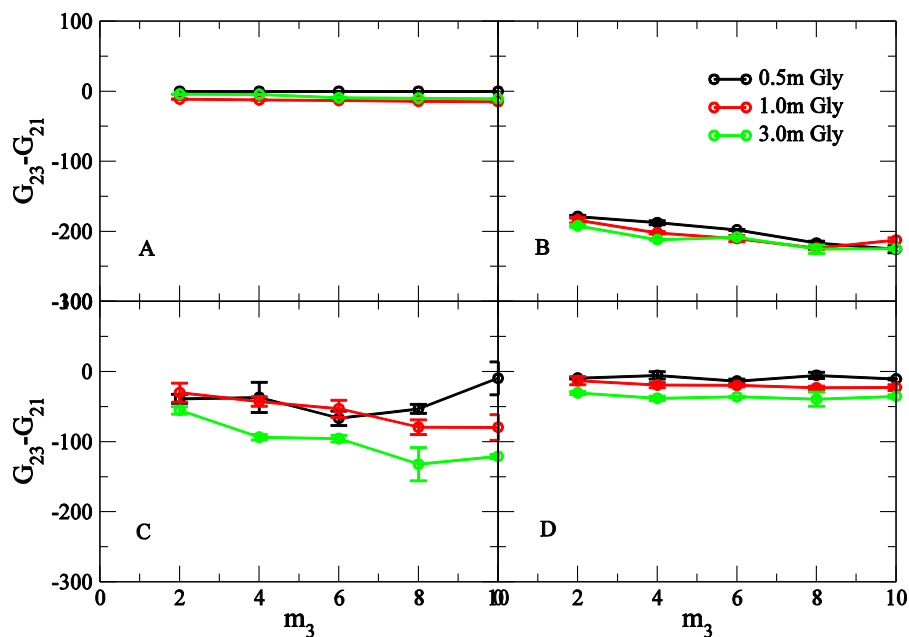
Figure 3.25 Separation of preferential structure parameters (G_{23}-G_{21}, cm^3/mol) from methanol-urea aqueous system into four terms as function of urea molality according to the relative radial distribution functions.   A) r =0~0.315 nm when g_{23} keeps at zero. B) r =0.315~0.605 nm when g_{23} is in 1st shell. C) r =0.605~0.955 nm is in 2nd shell. D) r =0.955~3.000 nm when g_{23} is in the extra part.
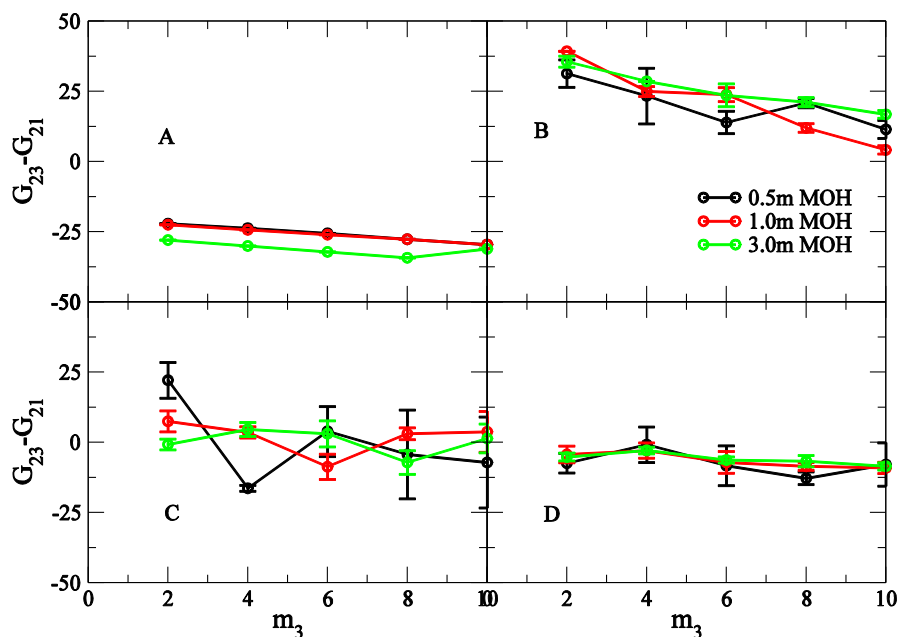
Different types of solute molecules show quite various preferential structure results. In principle, due to the smaller volume of water molecule, solute should prefer to bind water rather than urea in the excluded region (A). However, it only happens for glycine and methanol and the preference of glycine is very small. For neo-pentane and benzene, the values of G_{23}-G_{21} are almost zero. As expected, those nonpolar molecules would keep away from water despite of the small volume of water molecule. In neo-pentane solutions and benzene solutions, solutes displayed a strong preference with urea in the first shell (B) and less strong preference with urea in the second shell (C). More water molecules show up beyond the second shell in neo-pentane solutions, but not much in benzene. Glycine appears to bind with water within the first and second solvation shells, especially first, and does not show obvious preference when it goes further. Considering the scales of G_{23}-G_{21} in methanol systems, urea and water do not reveal any big preference difference.

3.5.4 Transfer free energy derivatives of solutes to urea solutions

According to equation (3.34), the derivative of the free energy change of transferring a solute molecule from pure water to urea aqueous environment is composed of two terms: $-\Gamma_{22}\mu_{23}$ and $-\Gamma_{23}\mu_{33}$. The first term accounts for the effect of urea on the interactions between solutes themselves, while the second term accounts for the effect on the interactions between solute and urea. Both two terms and the sum of them are displayed in Figure 3.26- Figure 3.29 to analyze which term gives a larger contribution to the derivatives of free energy changes.

Figure 3.26 Simulated transfer free energy derivatives (cm^3/mol) of urea in pentane-urea aqueous systems and detailed comparison of two parts contributing to transfer free energy.
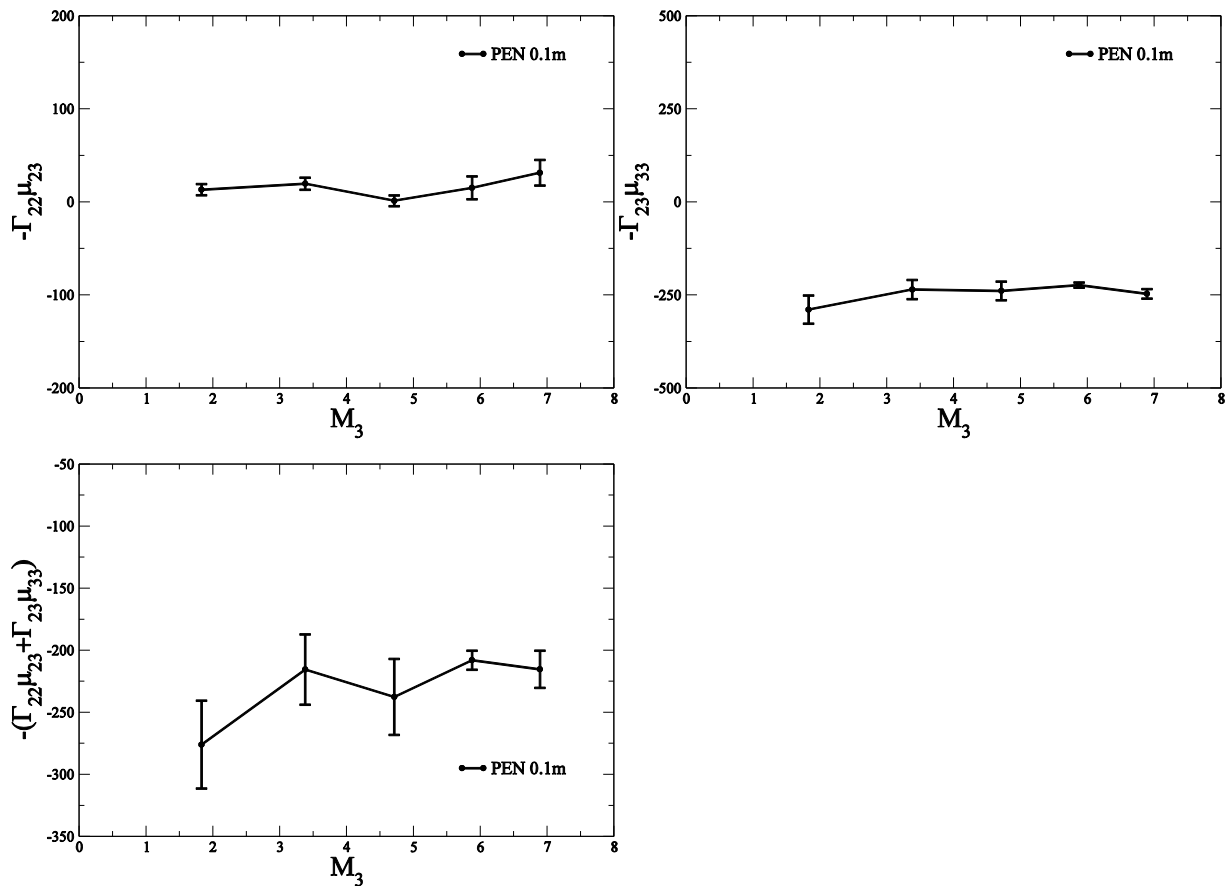
In neo-pentane solutions shown in Figure 3.26, the first term is shown as positive, , while the second term is shown as a large negative number. The second term is dominant in the derivative of free energy change, so the derivative of free energy change for diluted neo-pentane solutions is a large negative number, which means adding urea could stabilize the system.

Figure 3.27 Simulated transfer free energy derivatives (cm^3/mol) of urea in benzene-urea aqueous systems and detailed comparison of two parts contributing to transfer free energy.
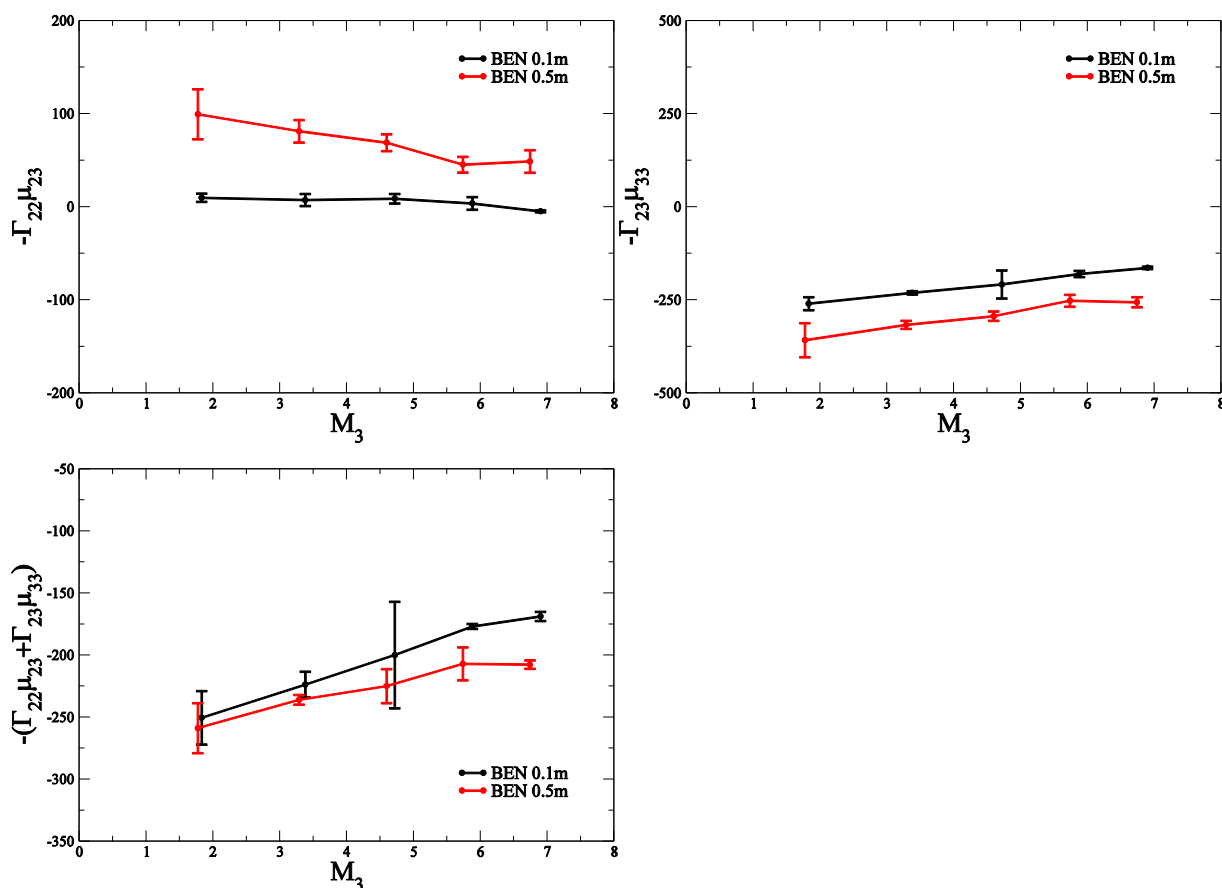
Similar to neo-pentane, the first term is positive, and the second term is a larger negative number in benzene system. When comparing these two terms, the second one gives dominant contribution to the derivative of free energy change, which leads to the conclusion that urea increases the benzene stability, especially for concentrated benzene solutions.

Figure 3.28 Simulated transfer free energy derivatives (cm³/mol) of urea in glycine-urea aqueous systems and detailed comparison of two parts contributing to transfer free energy.
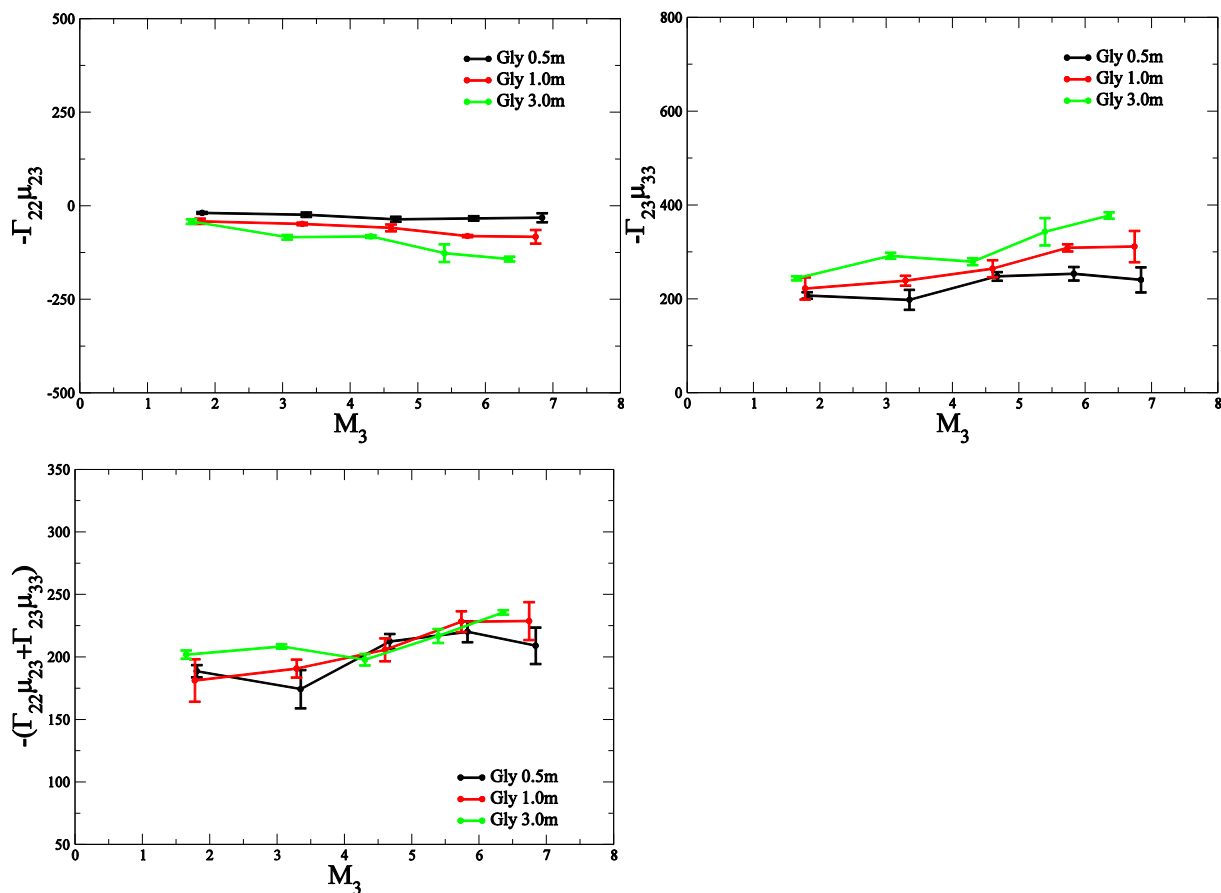
Figure 3.29 Simulated transfer free energy derivatives (cm^3/mol) of urea in methanol-urea aqueous systems and detailed comparison of two parts contributing to transfer free energy.
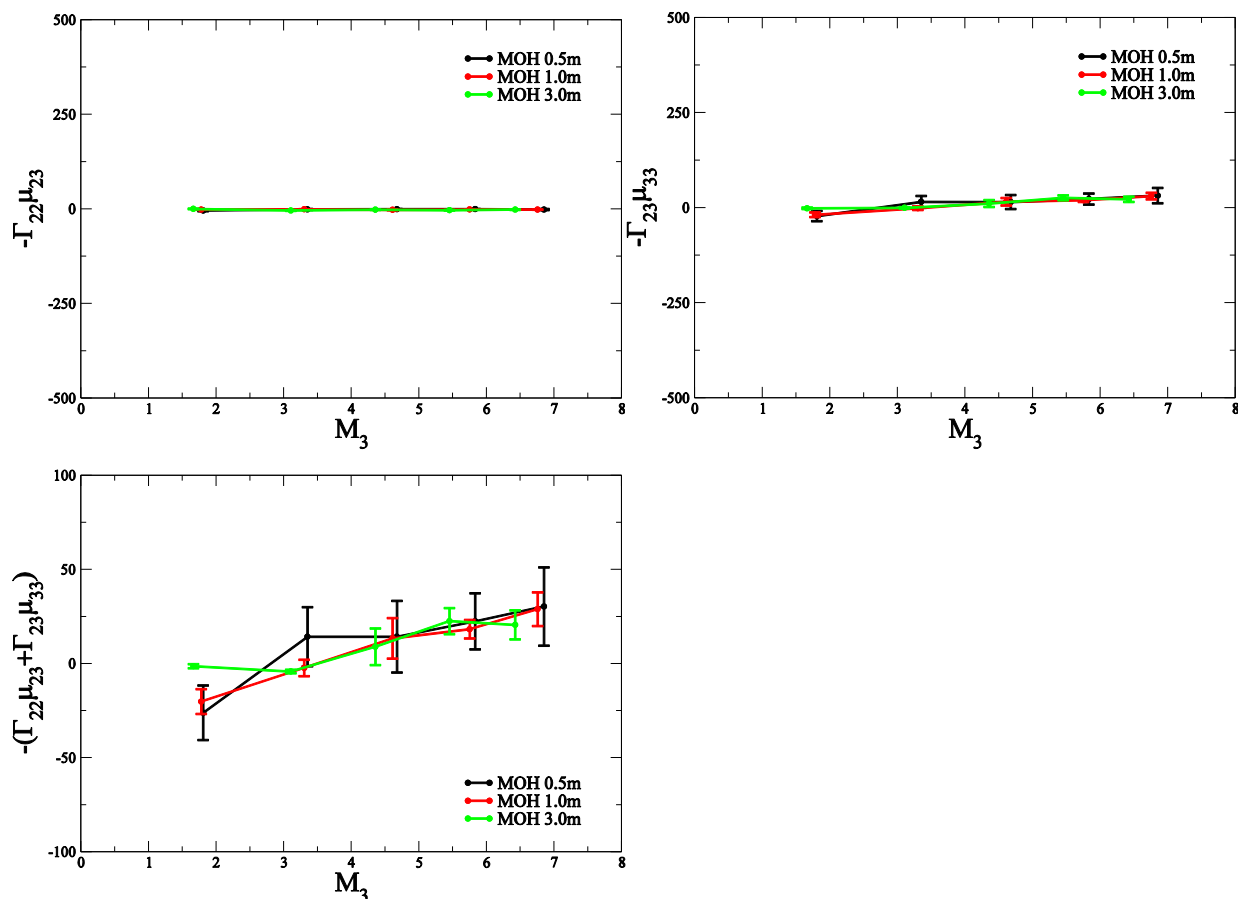
Compared with previous values, the terms for glycine and methanol systems are much smaller. For glycine, the first term is almost zero. Hence, the positive second term is dominant in the derivatives, which indicates that adding urea increases the free energy of the system. When the scales are considered, the terms for methanol are negligible. Urea does not change the stability of methanol aqueous solutions.

The Bolen's[60] and Tanford's[45] experimental results of transfer free energies were obtained from infinite diluted solutions and provided in Table 3.1-3.4. Infinite diluted concentration is normally used in experiments because it is able to avoid the interactions between solute molecules. However, in order to obtained statistical meaningful results, we used finite but low concentrations in simulations, and compared them with the experimental data to evaluate the application of KB theory on the calculation of preferential interactions and the accuracy of our

KB force fields. It is obvious that the KB model we used here gave us the correct signs of transfer free energy derivatives for all four systems and the correct magnitudes for pentane, benzene and methanol systems. For glycine systems, our calculated results are an order of magnitude larger than these very small experimental values. Considering all of them, we can generally conclude that the KB model and our KB force fields did a reasonably good job in the calculation of urea preferential interactions.

Table 3.1 Simulated neo-pentane derivatives of transfer free energy (kJ·L/mol^2) with respect to urea concentration in 0.1 m neo-pentane solutions with different urea concentrations compared to the experimental results of infinite concentrated solutions of similar solute molecules

Experiment	2M	4M	6M	7M	8M
Isobutane[45]	-0.23	-0.09	-0.15		-0.15
Neo-pentane[53]				-0.1	
Isobutane[60]	-0.23				
Simulated	**1.8M**	**3.4M**	**4.7M**	**5.9M**	**6.9M**
Neo-pentane (0.1 m)	-0.58	-0.53	-0.59	-0.52	-0.53

Table 3.2 Simulated benzene derivatives of transfer free energy (kJ·L/mol^2) with respect to urea concentration in 0.1 m benzene solutions with different urea concentrations compared to the experimental results of infinite concentrated solutions of similar solute molecules

Experiment	2M	4M		6M	8M
Toluene[45]	-0.38	-0.31		-0.29	-0.40
Toluene[60]	-0.35				
Simulated	**1.8M**	**3.4M**	**4.7M**	**5.9M**	**6.9M**
Benzene (0.1 m)	-0.62	-0.56	-0.50	-0.44	-0.42

Table 3.3 Simulated glycine derivatives of transfer free energy (kJ·L/mol^2) with respect to urea concentration in 0.5 m glycine solutions with different urea concentrations compared to the experimental results of infinite concentrated solutions

Experiment	2M	4M		6M	8M
Glycine[45]	0.004	0.008		0.057	0.050
Glycine[60]	0.078				
Simulated	**1.8M**	**3.3M**	**4.6M**	**5.7M**	**6.7M**
Glycine (0.5 m)	0.45	0.47	0.51	0.57	0.57

Table 3.4 Simulated methanol derivatives of transfer free energy (kJ·L/mol^2) with respect to urea concentration in 0.5 m methanol solutions with different urea concentrations compared to the experimental results of infinite concentrated solutions of similar solute molecules

Experiment	2M	4M		6M	8M
Ethanol[45]	-0.08	-0.04		-0.06	-0.05
Methanol[60]	-0.08				
Simulated	**1.8M**	**3.3M**	**4.6M**	**5.7M**	**6.7M**
Methanol (0.5 m)	-0.05	-0.01	-0.03	0.05	0.07

3.5.5 Atom-atom radial distribution function

The coordination numbers of first and second shells from atom-atom radial distribution functions are calculated in order to study the change in molecular distribution more quantitatively.

Figure 3.30 Simulated 1st shell coordination numbers in glycine-urea aqueous system atom-atom radial distribution functions (g_{ij}) as a function of urea molality
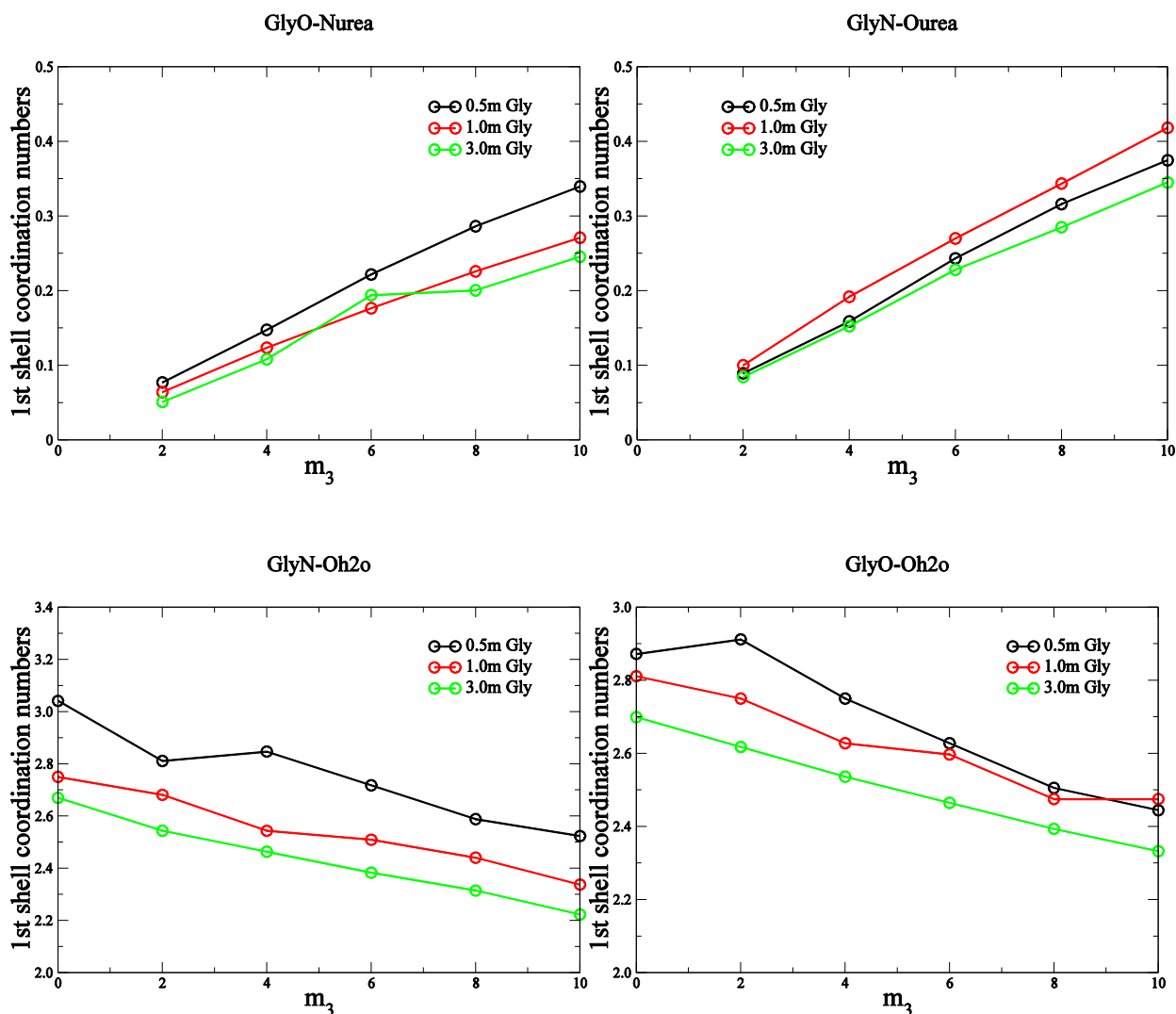
As shown in Figure 3.30, the first shell coordination numbers between glycine and urea are increased with the increasing urea concentration, which indicates more hydrogen bonds form between glycine and urea. On the other hand, as glycine or urea concentration increases, the coordination numbers between glycine and water decrease to nearly the same extent, suggesting that the hydrogen bonds between glycine and water are replaced by urea when urea is added.

The second shell coordination numbers for glycine mixtures are displayed in Figure 3.31, and they show the exact same trends as the first shell coordination numbers. And these results

give the conclusion that adding glycine or urea would facilitate the clustering of water, and the number of urea in glycine vicinity is proportional to urea bulk concentration.

Figure 3.31 Simulated 2nd shell coordination numbers from glycine-urea aqueous system atom-atom radial distribution function (g_{ij})
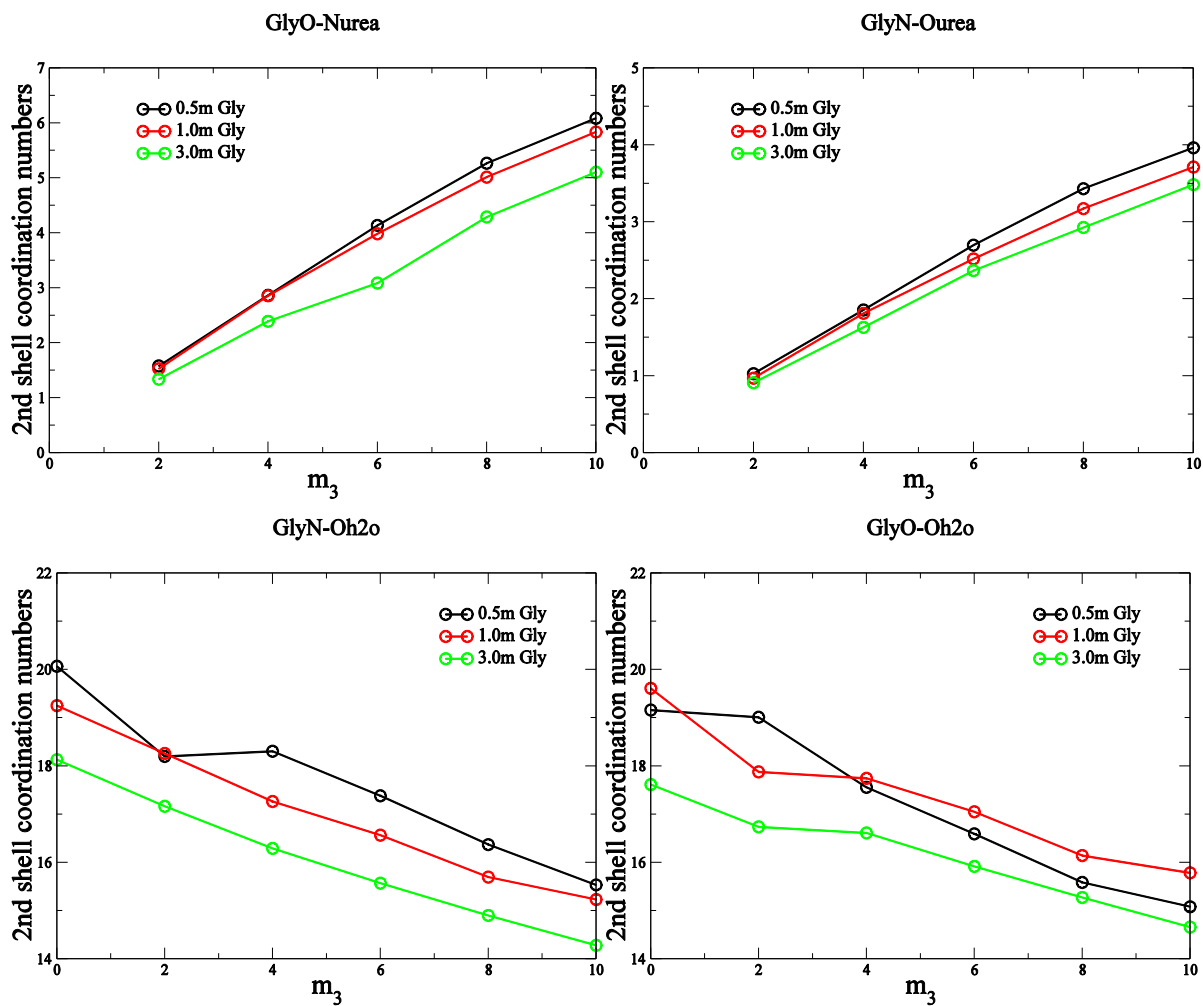
Figure 3.32 Simulated 1st shell coordination numbers in methanol-urea aqueous system atom-atom radial distribution functions (g$_{ij}$) as a function of urea molality
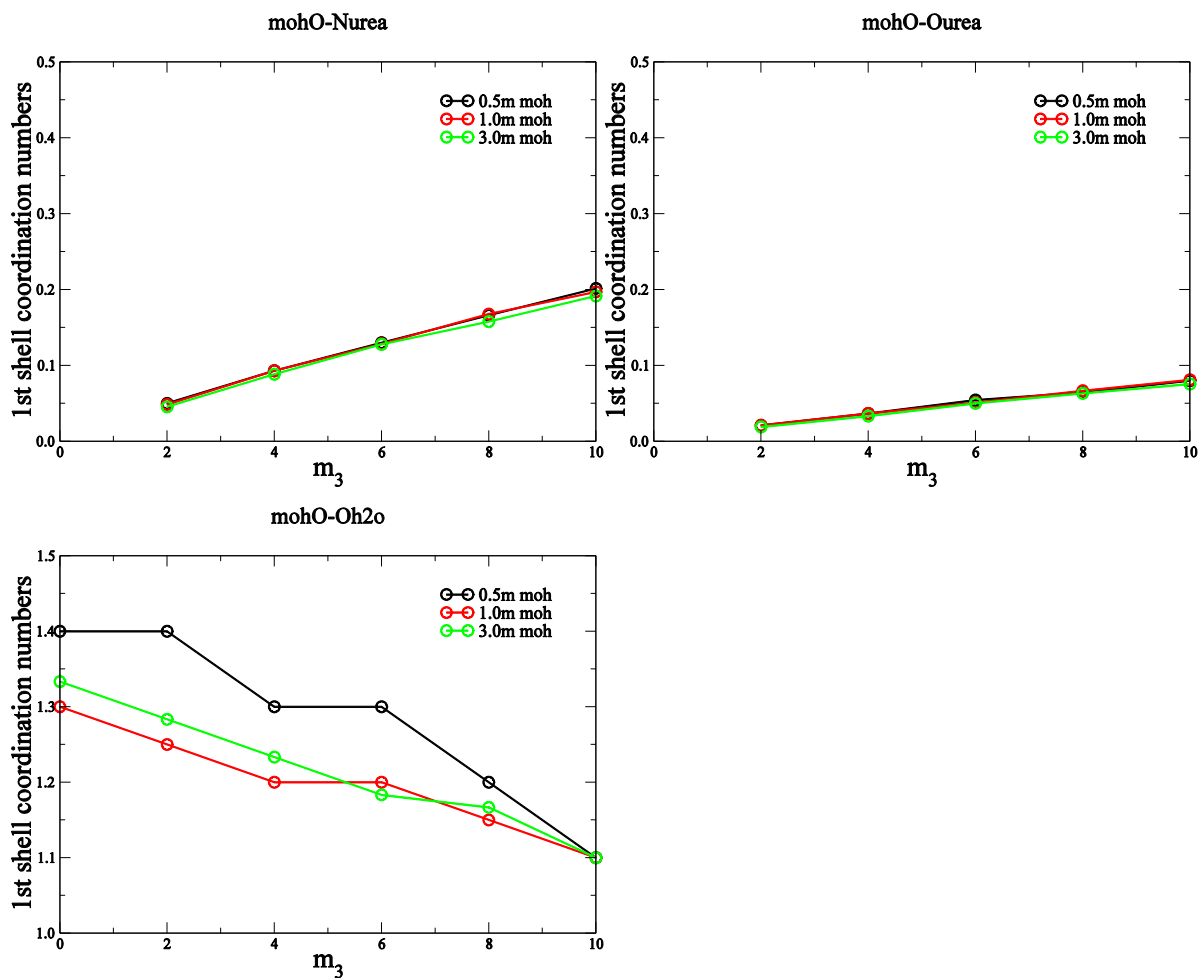
Figure 3.32 and Figure 3.33 provide the coordination numbers in the first and second shells for methanol ternary solutions. The trends of the changes for methanol systems are the same as those for glycine systems. However, the coordination numbers are much smaller than those in glycine solutions. It appears that glycine have more interactions with water and urea than methanol, which could probably be attributed to its local charges and one more hydrogen bonding site.

Figure 3.33 Simulated 2nd shell coordination numbers from methanol-urea aqueous system atom-atom radial distribution function (g_{ij})
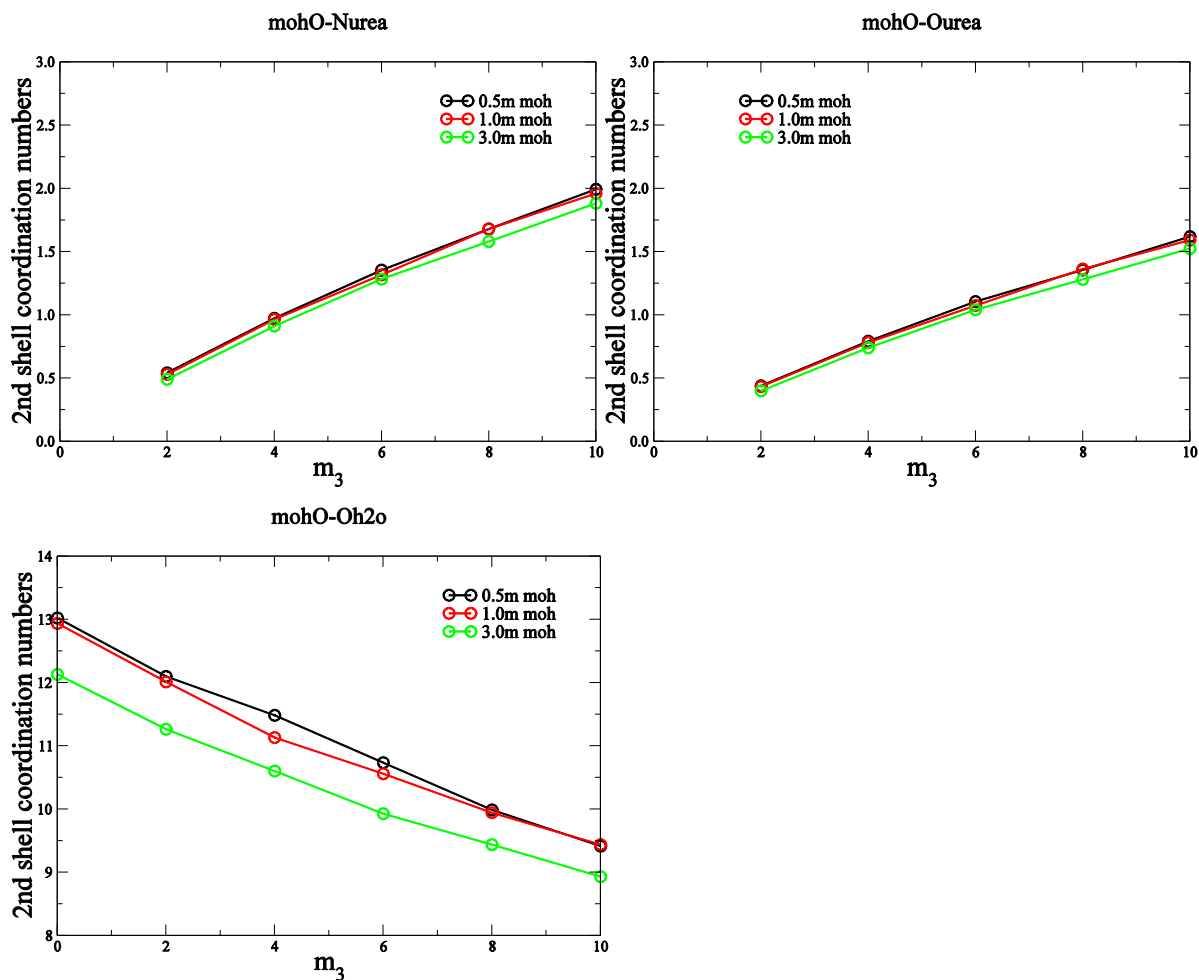
3.6 Conclusion

In summary, urea molecules have different cosolvent effects on various types of molecules. For the hydrophobic solutes, aliphatic neo-pentane and aromatic benzene, urea works as a denaturant, and decreases the chemical potential and makes large negative transfer free energy. These properties changes with the increase of urea concentration. Furthermore, the solute-cosolvent interactions, which are shown as the second term in the free energy change derivative equation, give bigger contributions to the transfer free energy. Urea plays a role of protecting osmolytes in glycine solution. Adding urea into the solutions increases the chemical potential, and gives a positive transfer free energy. Glycine molecules prefer to bind with water

rather than urea. Compared with the previous systems, it seems that urea does not exhibit any cosolvent effect on methanol, because most of the thermodynamic properties calculated for methanol solutions are shown as small and negligible fluctuations, not systematic trends. Briefly, urea shows favorable interactions with neo-pentane and benzene, and unfavorable interactions with glycine, while it does not show an obvious preference for methanol. A short comparison table is shown below.

		Neo-pentane	Benzene	Glycine	Methanol
μ_{23}		$-$	$-$	$+$	$-$
μ_{32}		$-$	$-$	$+$	$-$
Γ_{23}		$+$	$+$	$-$	~ 0
$\partial\beta\mu_2^{*}/\partial\rho_3$	1st term	$+$	$+$	~ 0	~ 0
	2nd term	$-$	$-$	$+$	~ 0
	Total	$-$	$-$	$+$	~ 0

The atom-atom coordination numbers in first and second shells are analyzed for glycine mixtures and methanol mixtures. As the urea concentration increases, the urea accumulated in the first shell of solute while the water solvation goes down. The same changes happen in the second shell, which indicates that urea induces not only the change of hydrogen bonding, but also the packing effects. This observation support the direct mechanism that urea directly interact with protein and destabilize the native structure.

Although the simulations of neo-pentane solutions and benzene solutions are far beyond their realistic solubility, the analyzed results bring reasonable explanations for the cosolvent effect of urea and agree with the conclusion given by Stumpe that charged and polar amino acids have slight preferences for contact with water molecules, and urea has strong preference for contact with aromatic and nonpolar side-chains.[89]

3.7 References

(1)   Timasheff, S. N. *Adv. Prot. Chem.* **1998**, *51*, 355.
(2)   Anderson, C. F.; Record, M. T., Jr. *Annu. Rev. Biophys. Biophysical chem.* **1990**, *19*, 423.
(3)   Timasheff, S. N. *Biochemistry* **1992**, *31*, 9857.
(4)   Franks, F.; Eagland, D. *CRC Crit. Rev. Biochem.* **1975**, *3*, 165.

(5)   Brown, A. D.; Simpson, J. R. *J. Gen. Microbiol.* **1972**, *72*, 589.

(6)   Yancey, P. H.; Somero, G. N. *Biochem. J.* **1979**, *183*, 317.

(7)   Forster, R. P.; Goldstein, L. *Am. J. Phys.* **1976**, *230*, 925.

(8)   Hebda, J. A.; Miranker, A. D. *Annu. Rev. Biophys.* **2009**, *38*, 125.

(9)   Westermark, P.; Andersson, A.; Westermark, G. T. *Phys. Rev.* **2011**, *91*, 795.

(10)   Throop, G. J.; Bearman, R. *J. Chem. Phys.* **1966**, *44*, 1423.

(11)   Kuntz, I. D., Jr.; Kauzmann, W. *Adv. Prot. Chem.* **1974**, *28*, 239.

(12)   Pessen, H.; Kumosinski, T. F. *Meth. Enzym.* **1985**, *117*, 219.

(13)   Rupley, J. A.; Careri, G. *Adv. Prot. Chem.* **1991**, *41*, 37.

(14)   Svergun, D. I.; Richard, S.; Koch, M. H.; Sayers, Z.; Kuprin, S.; Zaccai, G. *Proc. Nat. Acad. Sci.* **1998**, *95*, 2267.

(15)   Levitt, M.; Park, B. H. *Structure* **1993**, *1*, 223.

(16)   Liepinsh, E.; Otting, G. *Nat. Biotech.* **1997**, *15*, 264.

(17)   Liepinsh, E.; Sodano, P.; Tassin, S.; Marion, D.; Vovelle, F.; Otting, G. *J. Biomol. NMR* **1999**, *15*, 213.

(18)   Shimizu, A.; Fumino, K.; Yukiyasu, K.; Taniguchi, Y. *J. Mol. Liq.* **2000**, *85*, 269.

(19)   Feig, M. *Modeling Solvent Environments*; WILEY-VCH, 2010.

(20)   Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214.

(21)   Lin, T. Y.; Timasheff, S. N. *Biochemistry* **1994**, *33*, 12695.

(22)   Hochachka, P. W.; Somero, G. N. *Biochemical Adaptation. Mechanism and Process in Physiological Evolution*; Oxford University Press, 2002.

(23)   Makhatadze, G. I.; Privalov, P. L. *J. Mol. Biol.* **1992**, *226*, 491.

(24)   Rezus, Y. L.; Bakker, H. J. *Proc. Nat. Acad. Sci* **2006**, *103*, 18417.

(25)   Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. *Proc. Nat. Acad. Sci.* **2008**, *105*, 16928.

(26)   O'Brien, E. P.; Ziv, G.; Haran, G.; Brooks, B. R.; Thirumalai, D. *Proceedings of the National Academy of Sciences of the United States of America* **2008**, *105*, 13403.

(27)   Rossky, P. J. *Proc. Nat. Acad. Sci.* **2008**, *105*, 16825.

(28)   Auton, M.; Bolen, D. W. *Proc. Natl. Acad. Sci.* **2005**, *102*, 15065.

(29)   Auton, M.; Holthauzen, L. M.; Bolen, D. W. *Proc. Natl. Acad. Sci.* **2007**, *104*, 15317.

(30)   Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci.* **2003**, *100*, 5142.

(31)   Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. *Biophys. J* **2005**, *89*, 842.

(32)   Sharp, K. A.; Madan, B.; Manas, E.; Vanderkooi, J. M. *J. Chem. Phys.* **2001**, *114*, 1791.

(33)   Canchi, D. R.; Paschek, D.; Garcia, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338.

(34)   Zhang, Y. J.; Cremer, P. S. *Annu. Rev. Phys. Chem.* **2010**, *61*, 63.

(35)   Inouye, H.; Timasheff, S. N. *Biopolymers* **1972**, *11*, 737.

(36)   Record, M. T., Jr.; Anderson, C. F. *Biophys. J.* **1995**, *68*, 786.

(37)   Schellman, J. A. *Biopolymers* **1994**, *34*, 1015.

(38)   Schellman, J. A.; Gassner, N. C. *Biophys. Chem.* **1996**, *59*, 259.

(39)   Pierce, V.; Kang, M.; Aburi, M.; Weerasinghe, S.; Smith, P. E. *Cell Bioch. Biophys.* **2008**, *50*, 1.

(40)   Scatchard, G. *J. Am. Chem. Soc.* **1946**, *68*, 2315.

(41)   Hade, E. P.; Tanford, C. *J. Am. Chem. Soc.* **1967**, *89*, 5034.

(42)   Kerker, M. *Electromagnetic Scattering*; Pergamon, **1963**.

(43)   Kielley, W. W.; Harrington, W. F. *Bioch. Biophy. Acta.* **1960**, *41*, 401.

(44)   Murphy, K. P. *Protein Structure, Stability, and Folding*; Humana Press, 2001.

(45)   Nozaki, Y.; Tanford, C. *J. Biol. Chem..* **1963**, *238*, 4074.

(46)   Nozaki, Y.; Tanford, C. *J. Biol. Chem.* **1970**, *245*, 1648.

(47)   Nozaki, Y.; Tanford, C. *J. biol. chem.* **1965**, *240*, 3568.

(48)   Nozaki, Y.; Tanford, C. *J. biol. chem.* **1971**, *246*, 2211.

(49)   Bolen, D. W.; Baskakov, I. V. *J. mol. biol.* **2001**, *310*, 955.

(50)   Auton, M.; Bolen, D. W. *Biochemistry-Us* **2004**, *43*, 1329.

(51)   Nandi, P. K.; Robinson, D. R. *Biochemistry* **1984**, *23*, 6661.

(52)   Schrier, M. Y.; Schrier, E. E. *Biochemistry* **1976**, *15*, 2607.

(53)   Wetlaufer, D. B.; Lovrien, R. *J. Biol. Chem.* **1964**, *239*, 596.

(54)   Tanford, C. *J. Am. Soc.* **1964**, *86*, 2050.

(55)   Uedaira, H. *Bul. Chem. Soc. Jpn.* **1972**, *45*, 3068.

(56)   Uedaira, H. *Bul. Chem. Soc. Jpn.* **1977**, *50*, 1298.

(57)   Schonert, H.; Stroth, L. *Biopolymers* **1981**, *20*, 817.

(58)   Liu, Y.; Bolen, D. W. *Biochemistry* **1995**, *34*, 12884.

(59)   Qu, Y.; Bolen, C. L.; Bolen, D. W. *Proc. Nat. Acad. Sci.* **1998**, *95*, 9268.

(60)   Wang, A.; Bolen, D. W. *Biochemistry* **1997**, *36*, 9101.

(61)   Allen, M. P. a. T., D. J. *Computer Simulation of Liquids*; Clarendon Press, 1987.

(62)   Widom, B. *J. Chem. Phys.* **1963**, *39*, 2808.

(63)   Widom, B. *J. Phys. Chem.-Us* **1982**, *86*, 869.

(64)   Gao, J.; Kuczera, K.; Tidor, B.; Karplus, M. *Science* **1989**, *244*, 1069.

(65)   Goodfellow, J. M. *Computer Modelling in Molecular Biology*; VCH, 1995.

(66)   Shimizu, S. *J. Chem. Phys.* **2004**, *120*, 4989.

(67)   Shimizu, S. *Proc. Nat. Acad. Sci.* **2004**, *101*, 1195.

(68)   Schurr, J. M.; Rangel, D. P.; Aragon, S. R. *Biophys. J.* **2005**, *89*, 2258.

(69)   Shulgin, I. L.; Ruckenstein, E. *Biophys. J.* **2006**, *90*, 704.

(70)   Schellman, J. A. *Quart. Rev. Biophys.* **2005**, *38*, 351.

(71)   Smith, P. E. *J. Phys. Chem. B* **2004**, *120*, 18716.

(72)   Smith, P. E. *J. Phys. Chem. B* **2004**, *108*, 16271.

(73)   Smith, P. E. *Biophys. J.* **2006**, *91*, 849.

(74)   BEN-Naim, A. *Molecular Theory of Solutions*; OXFORD, 2006.

(75)   Kang, M.; Smith, P. E. *J.Chem. Phys.* **2008**, *128*, 244511.

(76)   Kang, M.; Smith, P. E. *Fluid Phase Equilibr.* **2007**, *256*, 14.

(77)   Ploetz, E. A.; Smith, P. E. *Adv. Chem. Phys.* **2013**, *153*, 311.

(78)   Ploetz, E. A.; Smith, P. E. *Phys. Chem. Chem. Phys. : PCCP* **2011**, *13*, 18154.

(79)   Weerasinghe, S.; Smith, P. E. *J. Phys.Chem. B* **2005**, *109*, 15080.

(80)   Kang, M.; Smith, P. E. *J. Comput. Chem.* **2006**, *27*, 1477.

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

(82)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435.

(83)   Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

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

(85)   Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952.

(86)   York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345.

(87)   Kokubo, H.; Rosgen, J.; Bolen, D. W.; Pettitt, B. M. *Biophys. J.* **2007**, *93*, 3392.

(88)   Ben-Naim, A. *J. Chem. Phys.* **1978**, *82*, 792.

(89)   Stumpe, M. C.; Grubmuller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126.

(90)   Samiotalis, A.; Cheung, M. S. *J. Chem. Phys.* **2011**, 135, 175101

(91)   Meloni, R.; Camilloni, C.; Tiana, G.; *J. Chem. Theor. Comp.* **2014**, 10, 846

(92)   Stumpe, M. C.; Grubmuller, H., *PLoS Comp. Biol.*, **2008**, 4, 1

(93)   Das, A.; Mukhopadhyay, C. *J. Phys.Chem. B*, **2009**, 113, 12816

(94)   Canchi, D. R.; Garcia, A. E. *Biophys. J.* **2011**, 100, 1526.

Chapter 4 - Contribution from the Vibrational Partition Function to Free Energy and Enthalpy Changes

4.1 Abstract

Contributions from the vibrational partition function (Q_{vib}) are investigated for several systems and processes in solution including: the enthalpy of evaporation, the free energy of solvation, protein folding, the activity of a solute in solution, and the enthalpy of mixing. The infrared frequencies obtained from experimental results, quantum calculations, and molecular dynamic simulations are used to analyze the Q_{vib}. It is determined that the contributions of Q_{vib} can be significant but usually difficult to evaluate especially for biomolecular systems. It is argued that in classical force fields these contributions can be, and have been, implicitly included by the use of effective intermolecular interactions. However, this requires the use of experimental and simulated free energy (or enthalpy) changes which have not been corrected for changes in the vibrational partition during the parameterization process. Even then, an effective interaction approach is not expected to fully reproduce the temperature dependent properties of real systems.

4.2 Introduction

Computer simulations are broadly used to investigate new phenomena and to expand our understanding of existing experimental data. The quality of a simulation is highly dependent on the accuracy of the force field and the degree of sampling of the process under consideration. Force fields are now developing to a stage where they can be expected to quantitatively reproduce a wide range of experimental physical and thermodynamic data. However, a large majority of computer simulations, especially those of biomolecular systems, use classical dynamics and this leads to the neglect of some significant quantum mechanical effects. One effect, which cannot be reproduced in classical simulations, involves the accurate contribution of intramolecular and intermolecular vibrations to the vibrational partition function. The importance of the difference between the classical and quantum mechanical vibrational partition functions was recognized some time ago.[1] The effects result in corrections to either the classical simulations results via the simulated density of vibrational states, or corrections to the experimental target data obtained using additional experimental data (typically intramolecular

and intermolecular vibrational frequencies). Most previous studies have focused on correcting the experimental or simulated enthalpies of vaporization and free energies of solvation to determine the accuracy of newly developed force fields for liquids[2]. However, except for during force field development, it is rare for a simulation to be corrected for quantum mechanical vibrational effects. Even though quantum effects usually dominate the vibrational partition function, classical simulations have still been widely used because they provide considerable advantages in computational efficiency which is essential for the study of slow processes such as protein folding.

In the first part of this study, using the available experimental spectroscopy data, we investigate the possible contributions of the vibrational partition function to several processes, including enthalpy of vaporization, the free energy of solvation, the variation in chemical potential (activity) of a solute in a solute plus solvent mixture, protein folding, and enthalpy of mixing. In order to obtain more frequencies of vibrations which are missing or unreadable from experimental results, quantum calculated and classical simulated frequencies are used to analyze the contribution of the vibrational partition function to the free energy changes or enthalpy changes in the second part. It is argued that the quantum contributions are significant and yet, in most cases, it is unfeasible to correct the results of a classical simulation. The implication of the results for the philosophy of force field development and modification is then discussed.

4.3 Background

4.3.1 Harmonic oscillator and anharmonic oscillator

Harmonic oscillator can be described by the Hooke's law, in which the restoring force is proportional to the distance from the equilibrium position. Moreover, for a harmonic oscillator, the potential energy, E, can be written in the quadratic form[3]

$$E = \frac{1}{2}kX^2 \tag{4.1}$$

where the k represents the force constants and X is the displacement from the equilibrium position. In general, the quadratic-term potential energy is a good approximation for most intramolecular vibrations and high frequency intermolecular vibrations.[1]

If the quadratic terms were accurate enough to express the potential energy, infinite energy would be required to break a valence bond. Therefore indeed, actual molecules belongs to

anharmonic oscillator.[3,4] Fortunately, when considering the effect of vibrational frequency shifts on free energy changes in molecular dynamics simulations, the harmonic oscillator models can be used because of two suppositions[1]: (1) anharmonicities have main effect on the low frequency motions which are nearly classical, and should be reasonably well reproduced by an accurate force field that could reproduce the experimental vibrational frequencies; (2) the high frequency motions, where the quantum effects are more significant, are almost harmonic. Based on these two suppositions, the vibrational partition functions could be simply evaluated using the density of vibrational states represented in the spectrum.

4.3.2 Classical and quantum harmonic oscillator and vibrational partition function

In a classical system,[5] the vibrational frequency of a harmonic oscillator could be calculated by

$$\nu = \frac{1}{2\pi}\sqrt{\frac{k}{\mu}} \tag{4.2}$$

where k is the force constant and μ is the reduced mass, which is given by

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \tag{4.3}$$

The oscillator undergoes sinusoidal oscillation of amplitude

$$A = \sqrt{\frac{2E}{k}} \tag{4.4}$$

Different from the continuous potential energy calculated in equation (4.1) for classical harmonic oscillators, in the quantum systems, the energy of the harmonic oscillator is obtained by solving the Shrödinger's equation. The discrete energy is expressed as

$$E = \left(n + \frac{1}{2}\right)h\nu \tag{4.5}$$

in which n is the quantum number taking on integer values from zero to infinity. The zero point energy is defined in the ground state with n=0, and populated as the temperature close to zero.

For the quantum systems, the vibrational partition function q_{vib} of each individual mode is shown as[6]

$$q_{vib} = \sum_n e^{-\beta E_n} = e^{-\beta h\nu/2} \sum_{n=0}^{\infty} e^{-\beta h\nu n}$$

$$= \frac{e^{-\beta h\nu/2}}{1-e^{-\beta h\nu}} = \frac{e^{-\frac{u}{2}}}{1-e^{-u}}$$

(4.6)

where $u=h\nu/k_B T=4.8 \times 10^{-3}\nu$ (cm^{-1}) at 298 K, and $\beta=1/ k_B T$. When $\beta h\nu$ is smaller than 1, which means the temperature is high enough and is usually satisfied in the classical dynamics simulated systems, the sum in equation (4.6) can be replaced by an integral to get

$$q_{vib} = e^{-\beta h\nu/2} \int_0^{\infty} e^{-\beta h\nu n} dn$$

$$= \frac{1}{\beta h\nu} = \frac{1}{u}$$

(4.7)

The total partition function Q_{vib} for the system can be expressed in terms of partition functions q_i for the individual modes as

$$Q_{vib} = \prod_i q_i$$

(4.8)

or

$$\ln Q_{vib} = \sum_i \ln q_i$$

(4.9)

Substituting equation (4.6) and (4.7) into (4.9), respectively, we can obtain

$$\ln Q_{vib} = -\sum \left[\tfrac{1}{2} u_i + \ln\left(1-e^{-u_i}\right) \right]$$

(4.10)

for real systems using quantum calculations and

$$\ln Q_{vib} = -\sum_i \ln u_i$$

(4.11)

for systems using classical dynamics. All the vibrational frequencies and how they change during a process are needed to evaluate the above $\ln Q_{vib}$. The intramolecular frequency changes are often known for small molecules, but this is unlikely for large molecules. Furthermore, even the quantum contribution of the low frequency modes for pure water, a highly studied system, is the subject of some debate[7-9].

4.3.3 Contribution of classical and quantum vibrational partition function to the free energy change

The free energy changes studied here involve the determination of the chemical potential of a solute in water or in gas phase. In statistical mechanics, the chemical potential μ of a species at a number density (molar concentration) of $\rho = N/V$ can be expressed as[10]

$$\beta\mu = \beta W + \ln\left[\Lambda^3\rho / Q_{int}\right] = \beta\mu * + \ln\left[\Lambda^3\rho\right] \qquad (4.12)$$

in which $\beta = 1/RT$ and Λ is the thermal de Broglie wavelength of the species. The chemical potential of a species depends on the pressure (P), temperature (T), and composition of the phase. The first term (W) quantifies the changes in the solute-solute, solute-solvent and solvent-solvent interactions on introduction of the particle. This includes cavity formation and all nonbonded interactions (van der Waals and Coulomb interactions) typically used in common classical force fields. The pseudo chemical potential $\mu* = W - RT \ln Q_{int}$ is discussed by Ben-Naim.[11] Under the assumption that the internal (intramolecular) and configurational (intermolecular) contributions to the total partition function can be separated, the internal partition function Q_{int} is a product of the corresponding rotational, vibrational, and electronic partition functions $Q_{int} = Q_{rot} Q_{vib} Q_{el}$. The aim of this study is to determine the contribution of Q_{vib} to several common process of interest. It is assumed that Λ does not change at a constant temperature, that Q_{rot} can be treated classically,[12] and that Q_{el} is a constant when molecules remain in their electronic ground states, which is the process considered here. It is known that the magnitude of Q_{vib} can be large compared to W.[1] However, very little is known concerning changes in the relative contributions of W and Q_{vib} for different processes.

Substituting equation (4.10) and (4.11) into equation (4.12), respectively, we can obtain

$$\beta\mu = \beta W + \ln\left[\Lambda^3\rho\right] + \sum_i [\tfrac{1}{2}u_i + \ln(1 - e^{-u_i})] \qquad (4.13)$$

for real systems and,

$$\beta\mu = \beta W + \ln\left[\Lambda^3\rho\right] + \sum_{i'} \ln u_i \qquad (4.14)$$

for classical systems. As a result, a red shifted vibrational frequency leads to a smaller contribution to Q_{vib}, decreasing the chemical potential of the molecule. In many cases, especially

intramolecular bond stretching and angle bending, the zero point energy terms, $\frac{1}{2}u_i$, dominate the value of Q_{vib}.

Therefore, the quantum correction from a certain vibrational frequency u_i is obtained by subtracting the classical representations from the quantum mechanical representations:

$$W_A^A(u_i) = \ln \frac{1 - \exp(-u_i)}{\exp\left(-\dfrac{u_i}{2}\right)} - \ln u_i \tag{4.15}$$

4.3.4 The isotope effect

The contributions to the quantum vibrational partition function can be evaluated using deuteration analysis.[3] When a hydrogen in a molecule is substituted by a deuterium, it is assumed that the potential energy and configuration of the molecule are altered by negligible amounts,[13,14] as the atoms are chemically equivalent. But the frequency of any vibrations involving the hydrogen will be changed and thereby affect the zero point energy and the vibrational partition function. Furthermore, it is also assumed that these frequencies are then simply shifted according to the change in the reduced mass μ. According to equation (4.3), when H atoms are replaced by D atoms, the corresponding vibrational frequency should be decreased by a factor of $\sqrt{\mu}$.

If we consider the symmetric vibration of methane as an example, frequencies are 2914 cm^{-1} for CH_4, and 2085 cm^{-1} for CD_4. The experimental ratio of them is 0.715, close to the theoretically expected ratio of 0.734. The discrepancy is due to the anharmonicity, and it is small enough to be neglected. The corresponding frequency ratio of heavy water to normal water is nearly equal to the theoretically expected ratio $1/\sqrt{\mu} = 1/1.37$. If the change in zero point energies dominates the quantum mechanical value of Q_{vib} (which is true for $\nu > 1000$ cm^{-1}), then using the fact that $\Delta\nu = \nu_D - \nu_H = (1/1.37-1)\nu_H = -0.27\nu_H$, the change in $\ln Q_{vib}$ on deuteration is found to be -0.27 times the magnitude of $\ln Q_{vib}$ for the vibrational modes involving hydrogen atoms in the undeuterated system. This is true for both red and blue shifted resonances. While the heavier deuterium atoms will also change the value of classical vibrational partition function, these changes will be cancelled for all processes.

4.3.5 Spectroscopy studies of molecular vibration

The absorption or emission spectrum for molecular vibrations are located in the infrared (IR) region ($0.1 \sim 13000$ cm^{-1}). Therefore, IR spectra have been commonly used to study the nature of the forces between the atoms of a molecule.

4.3.5.1 Understanding of IR spectra

In principle, the vibrational and rotational motion of a molecule could give rise to radiation. However, in IR spectroscopy measurement, the bands arising from rotation are often lost, especially in the case of low-resolution spectra of polyatomic molecules. Therefore, the bands or peaks in IR spectrum are considered as fundamental vibrational transitions. This works well for heavy molecules. With the high-resolution spectrographs, the bands or peaks in some light molecule spectra are resolved into a series of closely spaced lines. Quantum theory provides a good explanation that the molecule has discrete energy levels of rotation and vibration, and radiation occurs only when a molecule undergoes a transition from one energy level to another energy level.[3,4,15]

In the view of quantum theory, energy levels of vibration are much larger than those of rotation. The lowest frequency bands in the IR spectrum correspond to the transitions between two different rotational energy levels in the molecule. The other higher frequencies correspond to transitions between both rotational and vibrational energy levels. The widely separated bands in high frequencies are considered as vibrational bands.

4.3.5.2 Simulated vibrational frequency

A) Classical simulation of IR spectrum[16]

A major advantage of molecular dynamics over the Monte Carlo method is able to calculate time-dependent properties. One of the time-dependent properties, the velocity auto-correlation coefficient, can be used to analyze molecular vibrational frequencies.

The velocity autocorrelation coefficient of a system with N atoms is shown as normalized

$$C_v(t) = \frac{1}{N} \sum_{i=1}^{N} \frac{\langle v_i(0) \bullet}{\langle v_i(0) \bullet} \tag{4.16}$$

where $v_i(t)$ is the velocity of atom i at time t. The value of velocity autocorrelation coefficient indicates how closely the velocity at a time t is correlated with the initial velocity, and generally as shown in Figure 4.1, it has initial value of 1 and converges to 0 after a certain time.

Figure 4.1 Velocity autocorrelation function for liquid water simulated at 300K.
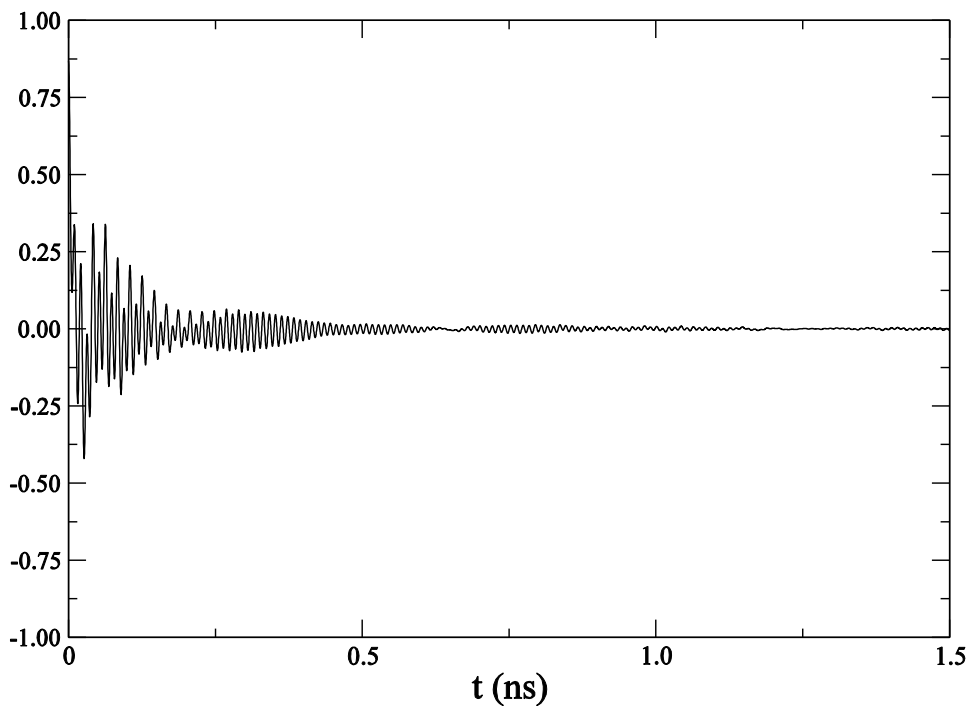
Compared with experimental results, the vibrational spectrum can be obtained from the velocity auto-correlation coefficient through Fourier transform

$$\hat{S}(\omega) = 2\int_0^{t_{max}} \cos(\omega t)C_v(t)dt \qquad (4.17)$$

where t is the simulation time. The spectrum $\hat{S}$ could be converted to a function of frequency ν with $\nu=\omega/(2\pi)$. An example of simulated water and heavy water IR spectra is shown in Figure 4.2.

Figure 4.2 Water and heavy water IR spectra from molecular dynamics simulations
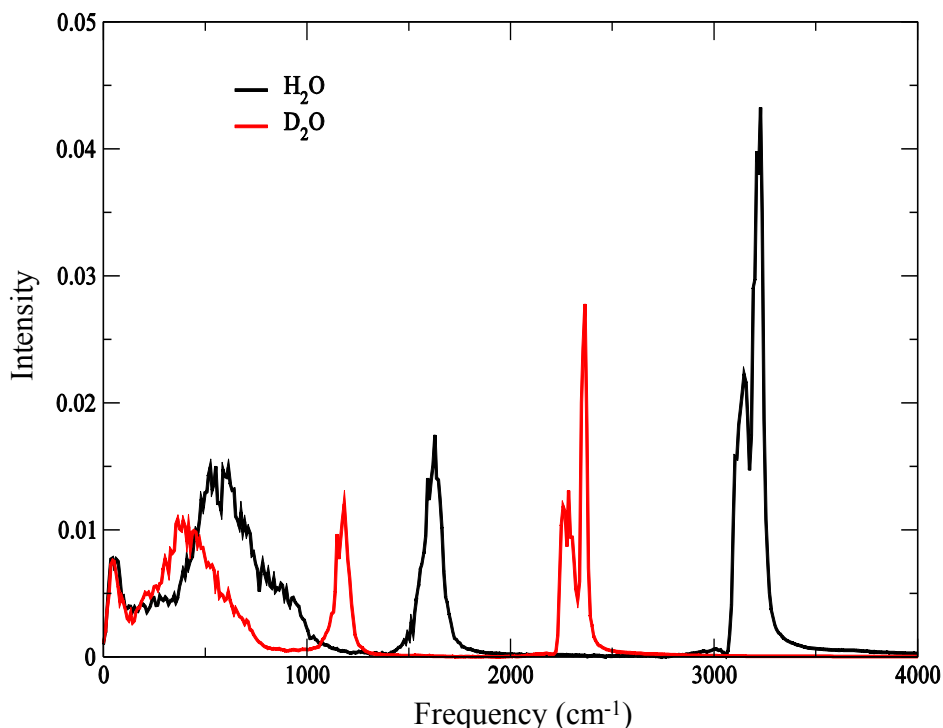
B) Quantum calculation of IR spectrum[5]

In quantum calculations, the number of degrees of freedom available to vibrate needs to be determined to analyze the vibrations of a molecule. An N-atom molecule has 3N degrees of freedom, while three degrees of freedom are in terms of translational motion and two degrees of rotational freedom for linear molecules or three degrees of rotational freedom for nonlinear molecules. This leaves 3N-6 (or 3N-5) degrees of freedom corresponding to the internal coordinates of a molecule of bond lengths and angles, and generates 3N-6 (or 3N-5) vibrational frequencies. Therefore, linear molecules have a total of 3N-5 degrees of vibrational freedom and nonlinear molecules have a total of 3N-6 degrees of vibrational freedom.

For a nonlinear molecule, the 3N-6 vibrational frequencies could be calculated as follows. In order to find the equilibrium geometry of the molecule, the electronic Schrödinger equation has to be solved. Based on the fact that the mass of nuclei are much larger then electrons, the nuclei could be considered as fixed and the nuclear kinetic-energy terms are negligible. Consequently, the Schrödinger equation is given by

$$\left(\hat{H}_{el} + V_{NN}\right)\psi_{el} = U\psi_{el} \tag{4.18}$$

where the energy U is the electronic energy including inter-nuclear repulsion and the first operator is the purely electronic Hamiltonian and could be expressed as

$$\hat{H}_{el} = -\frac{\hbar}{2m_e}\sum_i \quad_i -\sum_\alpha\sum_i \frac{Z_a e'^2}{r_{i\alpha}} + \sum_j\sum_{i>j}\frac{e'^2}{r_{ij}} \tag{4.19}$$

The nuclear-repulsion term V_{NN} could be expressed as

$$V_{NN} = \sum_\alpha\sum_{\beta>\alpha}\frac{Z_\alpha Z_\beta e'^2}{r_{\alpha\beta}} \tag{4.20}$$

here $r_{\alpha\beta}$ is the distance between nuclei α and β, and Z is the number of charges.

C) Normal mode analysis

Using the molecular electronic energy U, the set of second derivatives ($\partial^2 U/\partial X_i\partial X_j$) of U is calculated with respect to the 3N nuclear coordinates of a coordinate system. These derivatives are evaluated at the equilibrium geometry and can be computed numerically or analytically.

A mass-weighted force-constant matrix is built based on these derivatives and its elements displayed as

$$F_{ij} \equiv \frac{1}{\left(m_i m_j\right)^{1/2}}\left(\frac{\partial^2 U}{\partial X_i\partial X_j}\right)_e \tag{4.21}$$

here i and j change from 1 to 3N and m_i is the mass of atom located in coordinate X_i. A set of 3N linear equations with 3N unknowns is formed

$$\sum_j^{3N}\left(F_{ij} - \delta_{ij}\lambda_k\right)l_{jk} = 0 \tag{4.22}$$

and δ_{ij} is one if i equals j or zero if i is different from j. In order to have a nontrivial solution, which means a set of nonzero l_{jk}, for this set of homogenous equations, the coefficient determinant has to be

$$\det\left(F_{ij} - \delta_{ij}\lambda_k\right) = 0 \tag{4.23}$$

This 3N order determinant will give 3N roots for λ_k, and the molecular harmonic frequencies are calculated by

$$\nu_k = \lambda_k^{\frac{1}{2}} \big/ 2\pi \qquad\qquad (4.24)$$

Six of the 3N vibrational frequencies, which correspond to the three translational and three rotational degrees of freedom of the nonlinear molecule, will be zero. The molecular harmonic vibrational frequencies are the remaining 3N-6 vibrational frequencies.

4.4 Analysis of Experimental Frequencies

The contribution of changes in Q_{vib} to the total free energy change and the enthalpy change will be investigated for five different types of processes involving aqueous solutions. All calculations are performed for processes occurring at 298K and 1atm in the Gibbs ensemble. In order to simplify the following analysis to compare with results from a simulation using a classical force field, the corresponding force field is assumed to reproduce the experimentally observed vibrational frequencies and the intermolecular interaction (W) exactly. Therefore, the only difference between the experiment and simulation involves the quantum effects contained in the vibrational partition function. The contribution to the corresponding enthalpy and entropy changes could be estimated by a similar approach.

4.4.1 Enthalpy of vaporization

The enthalpy of vaporization is a physical property commonly used as a target to be reproduced in the force field parameterization. Because significant red shifts were observed for water passing from liquid to gas, the quantum correction for intermolecular and intramolecular degrees of freedom has to be taken into account

$$W_E^{\Delta}\left(u_i\right) = \frac{u_i}{2} + \frac{u_i}{\exp u_i - 1} - 1 \qquad\qquad (4.25)$$

Estimates of the corrections are included by

$$\Delta H_{vap} = -E_{pot} + P\Delta V + RT \sum W_E^{\Delta} \qquad\qquad (4.26)$$

here $-E_{pot}$ is the classical potential energy computed by molecular dynamics, the second term is for the work of expanding the gas. Using the three intramolecular frequencies for the gas phase (3756, 3657, and 1595 cm^{-1})[17-19] and liquid phases (3490, 3280, 1645 cm^{-1})[20], and the three higher intermolecular vibrational frequencies quoted by Postma[21] for the solution phase (800, 500, and 200 cm^{-1}, which are absent in gas phase because no intermolecular interaction) and equation (4.25), the intramolecular and intermolecular quantum corrections to ΔH_{vap} are 3.54 and

-3.77 kJ/mol, respectively. Since they have opposite signs, compared with the experimental ΔH_{vap} ~44 kJ/mol at 298K, the quantum corrections from intramolecular and intermolecular vibrations are nearly cancelled. Jorgensen and Madura[2] compared the heat of vaporization from the experimental data and the simulated results of TIP4P water model with the quantum corrections and an extra term, which was the enthalpy difference between the ideal and real gases and small enough to be neglected. With the quantum corrections, the simulated ΔH_{vap} could reproduce experimental results better than without the quantum corrections. However, the quantum corrections calculated here are not perfectly strict and have great reservations. The intermolecular frequencies are specific to the condensed phase and have more problems: 1) the experimental frequencies need to be assigned, which is not trivial for the low frequency modes. Even frequencies corresponding to restricted rotational or translational modes can make the assignment with more difficulty. 2) The condensed phase of water may be better approximated as a collection of water molecules in different environments, and each environment giving rise to different vibrational frequencies. Hence, the three rotational frequencies corresponding to the gas phase can become a mixture of rotation and vibration modes according to the environment of water molecules. Furthermore, exactly how many different species of water molecule environment there are, and their corresponding populations, are unknown.

4.4.2 Free energy of solvation

The transfer of a molecule from the gas phase to an aqueous solution corresponds to the free energy of hydration, or more generally solvation. The simplest case is that the solvation of water in pure water. In the Ben-Naim transfer process, the free energy for transfer of a water molecule from a fixed position in the gas phase to a fixed position in solution, similar to the transferring in (3.3.2), is given by equating the chemical potentials of water in the gas (g) and liquid (l) phases and rearranging to give the experimental free energy of transfer from gas to liquid

$$\Delta G^* = \mu_l^* - \mu_g^* = -\ln\left(\rho^l / \rho^g\right) \tag{4.27}$$

The free energy of transfer is expressed for a quantum (real) system as

$$\beta \Delta G^* = \beta \Delta W + \sum_i \left[\tfrac{1}{2}(u_i^l - u_i^g) + \ln\left[\frac{(1 - e^{-u_i^l})}{(1 - e^{-u_i^g})}\right] \right] = \beta \Delta W + \beta \Delta Q_{qm} \tag{4.28}$$

and for systems in classical simulations as

$$\beta\Delta G^* = \beta\Delta W + \sum_{i'} \ln\left[\frac{u_i^l}{u_i^g}\right] = \beta\Delta W + \beta\Delta Q_{cl} \qquad (4.29)$$

in which the ΔW term corresponds to the change in the non internal (intermolecular) work and the second term to the vibrational contribution to total free energy change. A red shift in a vibrational frequency on moving to the solution phase leads to a decrease in the solvation free energy. The gas and solution phase intramolecular vibration frequencies for water are known, although significant line broadening due to environmental effects is obvious in the solution phase.

The appropriate corrections can be evaluated with the reservations mentioned at the end of section 4.4.1. Using the frequencies listed above and equation (4.28), results in a contribution of $\Delta Q_{qm} = 3.9$ kJ/mol for the solvation process. The experimental value of ΔG^* is -26.47 kJ/mol. The contribution is very important (15%), and therefore has to be considered for accurate work. The calculation treats the intermolecular frequencies as vibrations. However, this can cause problems as it brings into question the separation of the internal and configurational partition functions.

Based on the data above, the contribution of the intramolecular and intermolecular modes to the values of ΔQ_{qm} are -3.5 and 7.4 kJ/mol calculated by equation (4.28), respectively. In most classical water force fields, the intramolecular modes are constrained to be rigid, so that they do not contribute to the equivalent simulated free energy change. Even when flexible bonds are used, according to equation (4.29), the intramolecular contribution is small, ΔQ_{cl} is -0.4 kJ/mol. The classical contribution to the intermolecular modes is 5.4 kJ/mol. The difference for the intermolecular modes of 2.0 kJ/mol does not cancel the intramolecular contribution -3.5 kJ/mol. This is the reason for the corrections that are applied to the simulated or experimental data. In principle, these corrections should be applied to all simulations before comparing with experiments. However, this procedure is not universally performed, except during actual parameterization state, and then only for pure liquids.

4.4.3 Solute activities in solution

The variation in the chemical potential of a solute in solution is characterized by the activity coefficient as

$$\beta\mu = \beta\mu^o + \ln[y\rho] \qquad (4.30)$$

in the scale of molarity and y is the molar activity coefficient. Using the derivatives of the chemical potential with respect to solute concentration provides a useful method to study the contribution of the vibrational partition functions since it avoids the dependence of the chemical potential on the choice of standard state. This can be achieved in several ways. From equations (4.12) and (4.30), the equation for a given solvent (1) and solute (2) could be written as

$$\frac{\partial \ln y_2}{\partial \ln \rho_2} = \frac{\partial \beta\mu_2^*}{\partial \ln \rho_2} = \frac{\partial \beta W_2}{\partial \ln \rho_2} - \frac{\partial \ln Q_2}{\partial \ln \rho_2} \qquad (4.31)$$

This approach involves a direct approximation to the last term in equation (4.31), and requires knowledge of the composition dependent vibrational frequencies for a solute-solvent system. It is easy to understand that results will vary between different solute and solvent systems. Our main concern is for solutes in an aqueous environment. For example, we will consider the data corresponding to hydrogen bonding solute and solvent systems where the vibrational frequencies have been studied as a function of solution composition. Even then, it is common for only some of the vibrations to be determined.

Take N-methylacetamide (NMA) with water at 313K as a good example. In this system the NH vibrational frequency changes sharply from 3410 cm^{-1} at x_{NMA} = 0.2 (6.84 M) to 3310 cm^{-1} at x_{NMA} =0.5 (10.69 M), and results in the change in zero point energy. The molar activity coefficient (y) of NMA changes from 0.78 to 0.88 for the corresponding changes[22]. Therefore, a simple finite difference calculation suggests that the left side of equation (4.31) is around 0.27, whereas the NMA vibrational shift contributes a value of -0.54 to the derivative in second term on the right side of equation (4.31), $-\dfrac{\partial \ln q_2}{\partial \ln \rho_2}$, from only a single vibrational mode. Obviously, the contributions from the vibrational partition function can rival that of the changes in the work performed against the environment (0.81). This particularly large effect is based on finite difference calculation instead of derivatives, and there are exceptions to this. For instance, studies of methanol-water and acetone-water mixtures suggest little changes in the methanol or acetone frequencies, although considerable shifts in the water frequencies are observed.[23,24] Consequently, the contribution from changes in vibrational partition function to changes in the

activity coefficient varies quite widely but can be significant. The large shifts observed for NMA are especially important for accurate simulations of peptide systems.

An alternative way to study the activity of solutes is using normal and heavy water. The difference observed should, to a good approximation, be directly related to the difference in Q_{vib} for normal and heavy water. Sodium chloride solutions provide an interesting system for which the molal activity coefficients have been determined in normal and heavy water.[25] In particular, the salt itself has no internal degrees of freedom and so the changes in vibration partition function could be attributed to water, which then affects the salt activity via the Gibbs-Duhem equation, $\rho_1 d\mu_1 + \rho_2 d\mu_2 = 0$ at constant T and P. The molar volumes of normal and heavy water are almost the same to within 0.3% and therefore all partial molar volumes are assumed to be identical in both solvents.[26] Differences in molal activities coefficients[25], although small, are noticeable and can be roughly calculated by the expressions,

$$\ln \gamma_2^D - \ln \gamma_2^H = avm_2 + b\left(vm_2\right)^2$$

$$\Delta\gamma = \frac{\partial \ln \gamma_2^D}{\partial \ln m_2} - \frac{\partial \ln \gamma_2^H}{\partial \ln m_2} = avm_2 + 2b\left(vm_2\right)^2$$

(4.32)

where γ_2 is the molal activity coefficient, m_2 is the salt molality, and v is the number of ions in the salt (equal to 2 for NaCl). The fitting curves compared with experimental results are shown in Figure 4.3.

Figure 4.3 Differences in solute activity coefficients between heavy and normal water after fitting to the equation (4.32). The dots are experimental results and the lines are from the fitting curve using the fitting parameters are listed in Table 4.1.
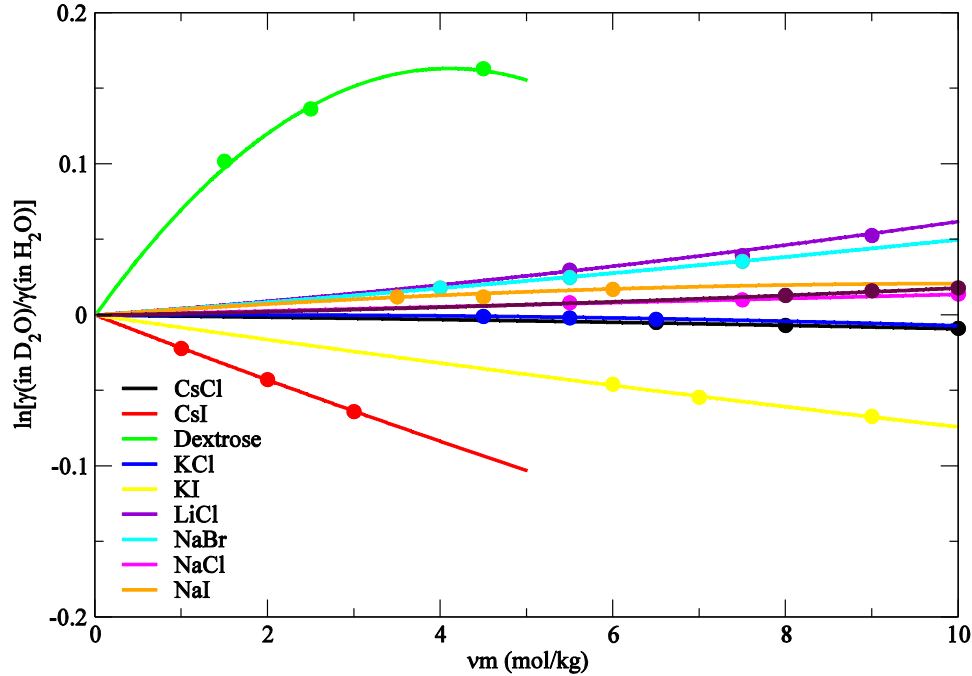
In order to investigate further, the expressions provided by Kirkwood-Buff (KB) theory[27] is used to relate changes in the chemical potential of a solute in terms of the corresponding intermolecular distribution functions in solution. This approach will prove particularly useful for the later discussion concerning the development of classical force fields. KB theory relates changes in activity to KB integrals which quantify the distribution of species i around species j according to

$$G_{ij} = 4\pi \int_0^\infty \left[g_{ij}(r) - 1 \right] r^2 dr \qquad (4.33)$$

The radial distribution function, g_{ij}, is usually considered to be a direct result of the intermolecular interactions. KB theory provides

$$\frac{\partial \beta \mu_2}{\partial \ln \rho_2} = \frac{\partial \beta a_2}{\partial \ln \rho_2} = \frac{1}{1 + \rho(G_{22} - G_{12})} = \frac{\phi_1}{1 + \rho_2(G_{22} - RT\kappa_T)} \approx \frac{\phi_1}{1 + \rho_2 G_{22}} \qquad (4.34)$$

where ρ_2 is the total ion concentration in the scale of molarity and the approximation is very good due to the low compressibility (κ_T) of water. The difference between water and heavy water can be shown as

$$\frac{1}{1+\rho_2 G_{22}^D} - \frac{1}{1+\rho_2 G_{22}^H} = \frac{\Delta\gamma}{\phi_1^2} \tag{4.35}$$

in which we have used the fact that $\left(\partial\ln\rho_2/\partial\ln m_2\right)=\phi_1$. This can be rearranged to give an expression for the KB integral in heavy water in terms of the KB integral in normal water and the activity derivative in equation (4.31). The result is

$$G_{22}^D = \frac{G_{22}^H\left(\phi_1^2-\Delta\gamma\right)-\rho_2^{-1}\Delta\gamma}{\phi_1^2+\Delta\gamma\left(1+\rho_2 G_{22}^H\right)} \tag{4.36}$$

For low solution concentrations (ρ_2 is close to 0 and ϕ_1 is close to 1) with a good approximation, this reduces to,

$$G_{22}^D = -a + G_{22}^H \tag{4.37}$$

The equations (4.32), (4.33) and (4.37) relate changes in the solute chemical potential due to deuteration of the solvent to the changes of intermolecular distribution between solute molecules. The differences between G_{22} in heavy water and that in normal water, calculated from the activity coefficients[28] and listed in Table 4.1, indicate the changes of ion associations due to the deuteration of the solvent. It is worth mentioning that activity coefficients in different literatures vary greatly.[25,28-30]

Table 4.1 Activity coefficient fitted parameters and the differences between G_{22} in heavy water and that in normal water

	a	b	$G_{22}^D - G_{22}^H$
	kg/mol	kg/mol	cm^3/mol
Dextrose	0.0796	-0.0097	-79.6
LiCl	0.0042	0.0002	-4.2
NaBr	0.0041	9×10^{-5}	-4.1
NaI	0.0041	-0.0002	-4.1
Urea	0.0010	8×10^{-5}	-1.0
KCl	0.0003	-0.0001	-0.3
NaCl	0.0014	-2×10^{-5}	-1.4
CsCl	-0.0006	-3×10^{-5}	0.6
KI	-0.0083	9×10^{-5}	8.3
CsI	-0.0221	0.0003	22.1

Because deuteration should only significantly affect the vibrational partition function, and not nonbonded interaction (W), the above changes in the intermolecular distributions are mainly a direct consequence from changes in the vibrational partition function. Furthermore, in the above example the solute does not have any internal degrees of freedom and therefore, only the contribution from changes in the vibrational frequencies originating from water molecules of solvation is probed.

4.4.4 Protein folding

One of the aims of many biomolecular force fields is to be able to investigate the details of how and why proteins fold into well defined three dimensional structures. Although the contribution of vibrational entropy changes to protein folding has been estimated[31,32], a search of literature gave no direct indication as to the possible contribution of the vibrational partition function to the free energy change for the protein folding process. It is well known that vibrational frequencies of proteins are shifted on folding, so some effects would be expected.

Here we try to estimate this contribution from some characterized frequencies. However, due to the large number of protein vibrational modes, an exact determination is currently infeasible.

An expression for the equilibrium constant for folding $(D \rightarrow N)$ can be obtained by equating the chemical potentials of the native (N) and denatured (D) states at low protein concentrations, which means no significant protein-protein interactions, to give,

$$-\beta \Delta G^{\circ} = \ln K = \ln \left[\frac{\rho^N}{\rho^D} \right] = -\beta \Delta W + \ln \left[\frac{Q_{vib}^N}{Q_{vib}^D} \right] \qquad (4.38)$$

or

$$\ln K = -\beta \Delta W - \sum_i \left[\tfrac{1}{2}(u_i^N - u_i^D) + \ln \left[\frac{(1-e^{-u_i^N})}{(1-e^{-u_i^D})} \right] \right] = -\beta \Delta W - \beta \Delta Q \qquad (4.39)$$

for which a red shift on folding leads to an increase in the population of the folded state, i.e. protein stabilization. The most common group found in peptides and proteins is the peptide group itself. The peptide group NH and CO vibrations give rise to the amide A and I bands in the vibrational spectra of proteins, which are typically red shifted and blue shifted, respectively, on folding. The major corrections required for classical simulations would be the inclusion of these high frequency shifts on folding, hence, we will focus on determining their magnitude. The shifts typically observed for proteins are very similar to those reported for NMA and water mixtures as a function of composition and so these values are used for calculations. The NH vibrational frequency shifts from 3420 cm^{-1} to 3310 cm^{-1} on going from pure water to pure NMA, while the CO frequency shifts from 1635 cm^{-1} to 1653 cm^{-1} for the same process.[33] If the same shifts are considered for a single NH and CO vibration, which upon folding form intramolecular protein hydrogen bonds similar to those in liquid NMA, then one finds contributions of -0.27 (or -0.66 kJ/mol) for NH and 0.04 (or 0.11 kJ/mol) for CO, respectively, to the value of $\beta \Delta Q$. It would indicate that burying a mere 50% of the peptide groups in a 100 residue protein would contribute -27.5 kJ/mol to the folding free energy change for just these two vibrations. This is on the order of the total stabilizing energy for many folded proteins.

It is obvious that this is a simplified calculation which ignores many other vibrational modes and the low frequency modes in particular. However, the changes in low frequency modes in proteins should be highly classical in character, reasonably well reproduced by classical simulations, and therefore requiring little or no correction. The hydrogen bonding frequencies are

somewhat intermediate and may compensate the high frequency shifts. Evidence for this comes from a study of the deuterated (NH to ND) protein CD2 which indicates little change in the denaturation equilibrium.[34] However, there is no evidence that these contributions will always be cancelled.

Other studies of protein folding in heavy water suggest a remarkable contribution from changes in the vibrational modes of water to the folding process.[35] The native conformation of the CD2 protein with 98 residues is stabilized by -5.0 kJ/mol using heavy water.[34] To include water in the chemical equilibrium, one has to evaluate the vibrational partition function for an equation of the form D: $nH_2O \rightarrow N+nH_2O$, which is difficult as n is unknown. However, by assuming all water frequencies are shifted by a factor of -0.27 on deuteration, and the vibrational partition function is dominated by the zero point energy term, then the contribution of the water modes can be estimated as -0.50/-0.27 = 18.5 kJ/mol, which can be compared to the total stabilization energy of -36.5 kJ/mol. Therefore, the folding process in heavy water is favored more than in normal water due to the reduction in the unfavorable change in ΔQ. The vibrational change is 33% of the value of ΔW = -55.0 kJ/mol. It is interesting that this is similar to the contribution from the zero point energies obtained for the formation of water dimers as determined by quantum mechanical calculations.[36] Furthermore, the above calculations suggest that the contribution from electrostatic and VDW interactions is far more favorable than that obtained when not explicitly considering the contribution of ΔQ.

4.4.5 Enthalpy of mixing

As a final example, we provide an analysis for the enthalpy of mixing of acetone and water. Max and Chapados evaluated acetone-water mixtures by considering the water and acetone molecules in a random organization, constrained only by hydrogen bonding between the molecules. They determined the molecule species distribution and corresponding IR frequencies as a function of concentration.[37] Based on their data, a detailed analysis of the vibrational frequency changes for both acetone and water has been provided. The contribution of these changes to the enthalpy of mixing is given by

$$\Delta H^{Ex} = \tfrac{1}{2} h \sum_{i=1}^{5} \sum_{j=1}^{3} \Delta v_{i,j}^{\omega} \omega_i + \tfrac{1}{2} h \sum_{i=1}^{5} \sum_{j=1}^{3} \Delta v_{i,j}^{\alpha} \alpha_i \tag{4.40}$$

where the summations over i represent the various water (or acetone) specie present in solution, and the summations over j represent the major bond vibrational modes of water (or acetone), The ω and α values are the weights due to the composition and fractional population of each form, while the ν values are the frequencies of corresponding species, and $\Delta\nu_{i,j}^{\omega}\omega_i$ (or $\Delta\nu_{i,j}^{\alpha}\alpha_i$) shows the change from the pure liquid to corresponding composition.

Figure 4.4 Top panel: water vibrational frequency changes in acetone-water mixture. Bottom panel: Excess enthalpy of mixing including contributions from the intramolecular vibrations.
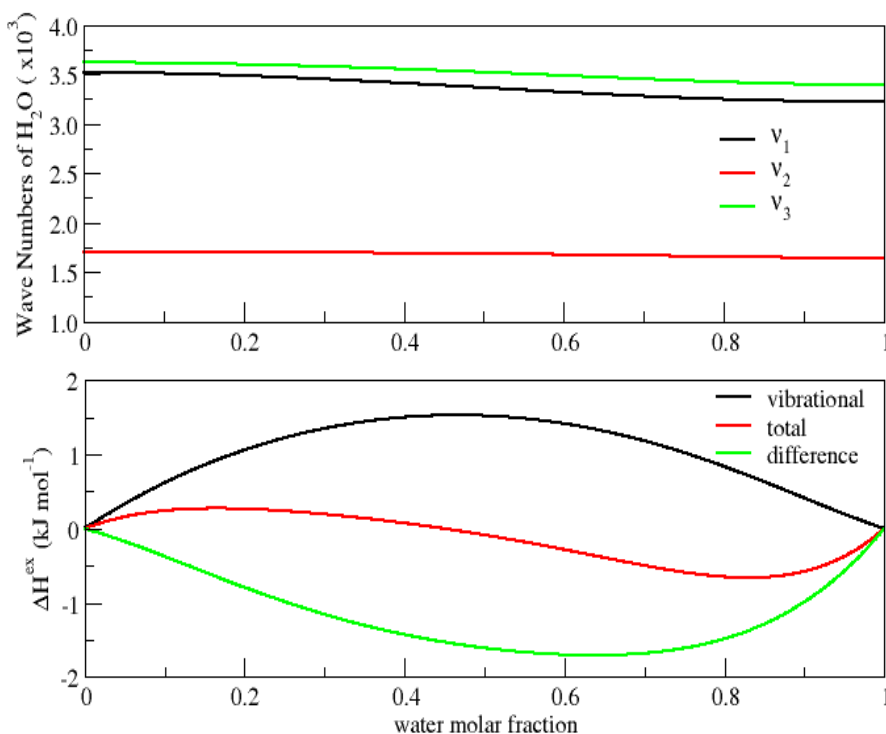
The total water vibrational frequency, indicating the first term in equation (4.40), as a function of water molar fraction is displayed in the top panel of Figure 4.4. The acetone vibrational frequency, indicating the second term in equation (4.40), was obtained as well, but not shown because acetone vibrational frequencies did not have remarkable change with species composition. The results calculated by equation (4.40) is represented as the black line in the

bottom panel of Figure 4.4, the red line is from the experimental enthalpy of mixing[38], and the green line shows the difference between the vibrational and experimental data, which could be explained as the intermolecular interaction. These results clearly indicate a significant contribution from shifts in the bond frequencies, especially for water. In fact, the two contributions, vibration and intermolecular, act in different directions and appear to lead to an overall sigmoidial shape for the enthalpy of mixing. Interestingly, this shape has been attributed to polarization effects and not to vibrational changes.[39]

As discussed in (4.3.4), if all water molecules in the system are substituted by heavy water molecules, the total water vibrational frequencies would be red shifted by 0.27x1.5=0.41 kJ/mol at the highest point in the black curve in Figure (4.4). With the approximations that the acetone frequencies are not affected by the water composition and the intermolecular interaction should keep the same since H and D are chemically equivalent, the total enthalpy of mixing should decrease about 0.41 kJ/mol. However, experimental results show that the enthalpy of mixing for acetone-heavy water system is really close to that for acetone-normal water system, increasing by a negligible amount,[40] which disagrees with the analysis results from the acetone-water IR spectrum. Hence, more frequencies, especially low frequencies, are needed for further analysis.

4.5 Simulation Frequency Analysis

Because the experimental spectrum analysis provides some large deviation from real results and some consequences which are hard to explain, vibrational frequencies from quantum calculations are adopted to analyze the contribution from vibrational partition function. The quantum calculations in density functional theory are used to provide enough frequencies in all the vibrational degrees of freedom. Although they are composed of only two molecules, not a real solution, they can tell the trends of frequency shifts. The molecular dynamics simulation is supposed to mimic the effect of solute composition on the frequency shifts.

4.5.1 Methods

4.5.1.1 Density functional theory calculations

All quantum calculations in this work were performed using density functional theory with the Amsterdam Density Functional (ADF) program[41]. In the ADF program, a triple-ζ

polarized basis set with the frozen core approximation was used for the geometry optimization and frequency calculation. Both the Generalized Gradient Approximation (GGA) BP86 and the hybrid functional B3LYP[42] were compared. All the systems were performed in the gas phase at 298K. The frequencies were visualized by Jmol (http://jmol.sourceforge.net/). Detailed frequency analysis is provided in supporting information.

4.5.1.2 Molecular dynamics simulations

All the classical simulations in this work were performed using Molecular Dynamics with GROMACS program[43]. The enhanced simple point charge water model (SPC/E)[44] was used for water and the acetone force field was from Kirkwood-Buff Force Field developed by Smith et al[45]. All the systems were performed at 300 K and 1 atm. The relaxation times for temperature and pressure were 0.1 and 0.5 ps, respectively using weak coupling technique[46]. A step time of 0.25 fs is adopted for integration of the equations of motion. The PME approach[47] was used for all electrostatic interactions and a twin range cutoffs of 1.0 and 1.5 nm were employed for the LJ and VDW interactions, respectively. The nonbonded update frequency was 10 steps. The initial configurations were generated by random placement of molecules into a cubic box of length 6 nm. All simulation times were 7.5 ns in length, and the final 3.75 ps were used to calculate ensemble averages.

4.5.2 Results

4.5.2.1 NMA and water systems

In order to evaluate the frequency shifts for protein backbones in protein folding process, four systems: an NMA dimer, a water dimer, and two NMA-water complexes with hydrogen bonds formed in different bonding sites were studied using density functional theory. The protein folded states are simplified into the first two systems as shown in Figure 4.5, while the protein unfolded states are simplified into the other two systems as shown in Figure 4.6, in which NMA worked as hydrogen bond donor and acceptor, respectively.

130

Figure 4.5 Snapshots of NMA dimer and water dimer systems used to represent the folded protein.
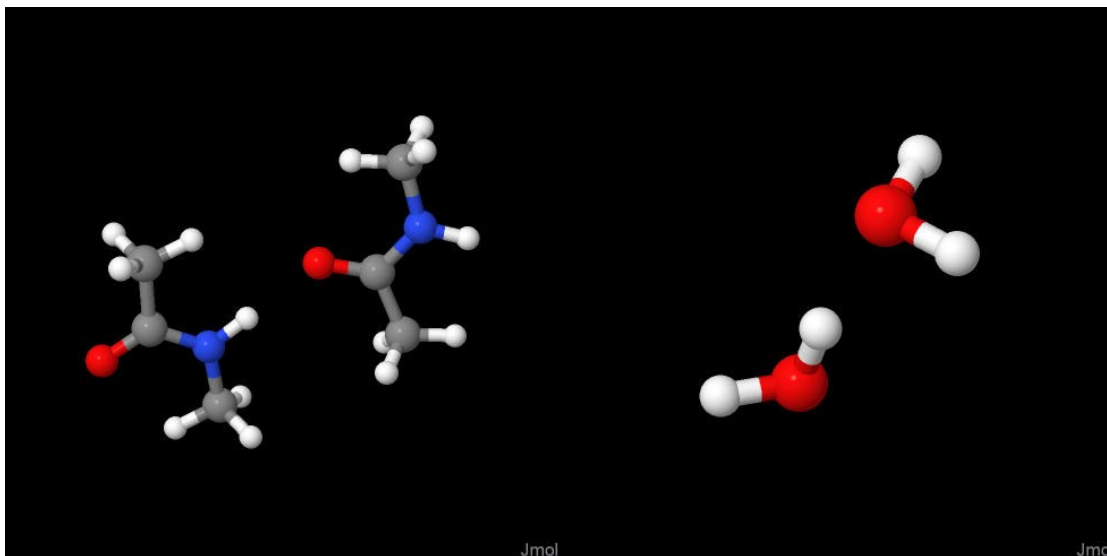
Figure 4.6 Snapshots of NMA and water complex systems with hydrogen bonds formed in different bonding sites used to represent the unfolded protein.
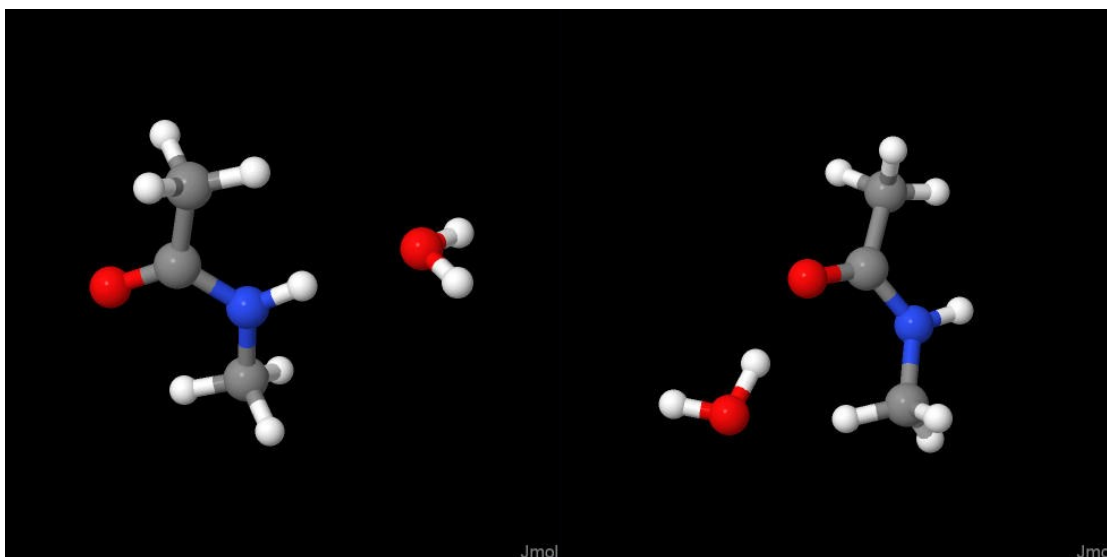
The vibrational partition function contributions and corrections for energy and free energy were calculated by equations (4.11) (4.15) and (4.25), respectively. All of these changes

in the folding process with normal and heavy water are listed in Table 4.2. With normal water, the vibrational partition function gives a large contribution (-9.61 kJ/mol) to the free energy change, and it is much larger than the value obtained in section (4.4.4), which was calculated based on the several high frequencies. This means both the high and low frequencies are important in the calculation of vibrational partition function, and only taking parts of frequencies in the calculation can induce a large error. The corrections for free energy and energy from classical system to real system are calculated by equation (4.15) and (4.25). Although the vibrational partition function contribution is large, the differences between classical dynamics and quantum calculations are small, 0.31 and 0.19 kJ/mol for energy and free energy, respectively.

Table 4.2 Vibrational partition function contributions and corrections for energy and free energy corresponding to the frequency shifts from NMA-water complex to pure dimers with normal and heavy water, respectively .

		$RT\ln Q$ (kJ/mol)		$RT\sum W_E^{\Delta}$ (kJ/mol)		$RT\sum W_A^{\Delta}$ (kJ/mol)	
		H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
Bond Stretching	NMA	-0.49	-0.26	-0.50	-0.27	-0.45	-0.25
	Water	0.61	0.23	0.61	0.23	0.53	0.19
	Total	0.11	-0.03	0.11	-0.04	0.08	-0.06
Angle Bending	NMA	-0.95	-0.92	0.24	0.10	0.16	0.06
	Water	-0.08	0.01	-0.08	0.01	-0.06	0.00
	Total	-1.03	-0.92	0.17	0.10	0.10	0.06
Torsion & Libration		-8.69	-1.35	0.04	0.14	0.01	0.07
Total		-9.61	-2.30	0.31	0.21	0.19	0.07

In order to study the solvent isotopic effect, the four systems were tested with heavy water as well. NMA-D_2O complex has one more negative torsion & libration frequency. If the negative frequency could be adjusted to the corresponding frequency in NMA-H_2O with the reduced mass ratio, the total contribution change from the frequency shifts for folding process in

heavy water would be -2.30-9.14=-11.44 (kJ/mol). This is in agreement with the fact that heavy water stabilized the protein in folded state[48-50].

The folding process of NMA deuterated in amide was also studied and compared with that of normal NMA in Table 4.3. The vibrational partition function contributions with normal and deuterated NMA are very close to each other, which are supported by several protein experimental results[34,51-54]. The differences for energy and free energy between classical dynamics and quantum calculation are very small and do not show obvious protein isotopic effect.

Table 4.3 Vibrational partition function contribution, and corrections for energy and free energy corresponding to the frequency shifts from NMA-water complex to pure dimers with normal and deuterated NMA, respectively.

		$RT\ln Q$ (kJ/mol)		$RT\sum W_E^\Delta$ (kJ/mol)		$RT\sum W_A^\Delta$ (kJ/mol)	
		H(NMA)	D(NMA)	H(NMA)	D(NMA)	H(NMA)	D(NMA)
Bond Stretching	NMA	-0.49	-0.01	-0.50	-0.04	-0.45	-0.09
	Water	0.61	0.62	0.61	0.62	0.53	0.54
	Total	0.11	0.61	0.11	0.58	0.08	0.45
Angle Bending	NMA	-0.95	-1.70	0.24	-0.26	0.16	-0.16
	Water	-0.08	-0.07	-0.08	-0.07	-0.06	-0.06
	Total	-1.03	-1.77	0.17	-0.33	0.10	-0.22
Torsion & Libration		-8.69	-8.18	0.04	0.06	0.01	0.02
Total		-9.61	-9.35	0.31	0.31	0.19	0.26

4.5.2.2 Acetone and water systems

The next question is to evaluate the contribution from the C=O stretching frequency change in the systems without the N-H bonds. Therefore, an acetone dimer and an acetone-water complex, displayed in Figure 4.7, were studied with BP86 functional and triple-ζ polarized basis

set. The acetone dimer was designed as two acetone molecules located in parallel planes with two carbonyl groups overlapping each other and pointing to the opposite ways.

Figure 4.7 Snapshots of acetone dimer and acetone-water complex systems
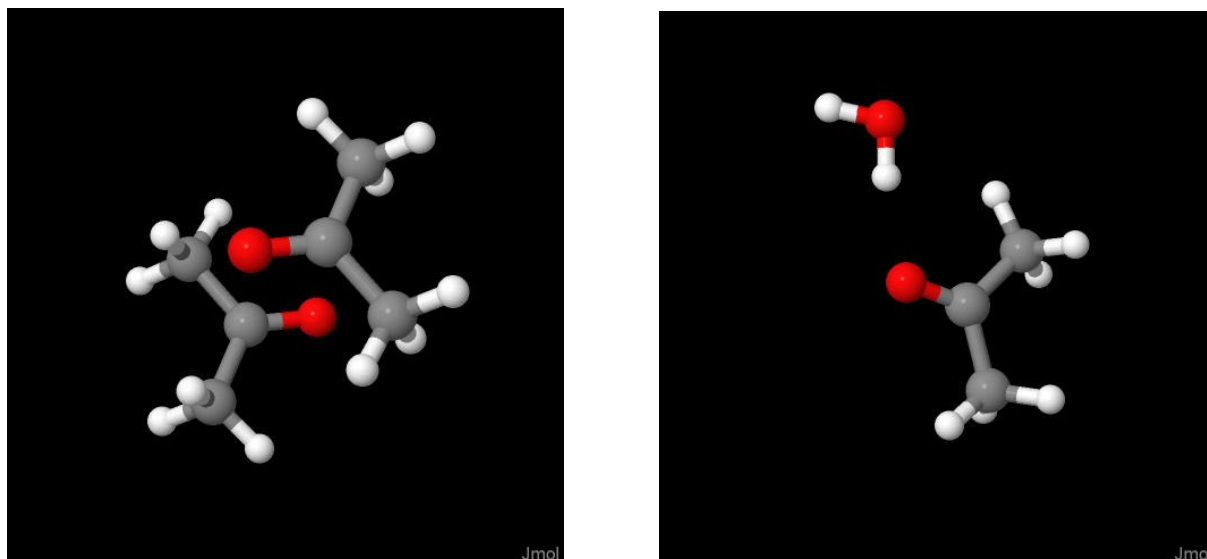
Similar to NMA systems, the vibrational partition functions give large contributions to free energy change. However, the difference in free energy between classical dynamics and real quantum calculations are small, 0.03 kJ/mol. The solvent isotopic effect is negligible in these systems, which are supported by calorimetry experimental results[40]. It is worthy to point out that $RT\sum W_E^\Delta$ terms with both normal and heavy water are very small, which is different from the deduction in section 4.4.5. This is probably due to the fact that only parts, instead of all, of the vibrational frequencies are used to calculate the enthalpy of mixing in section 4.4.5.

Table 4.4 Vibrational partition function contribution, and corrections for energy and free energy corresponding to the frequency shifts from acetone-water complex to pure dimers with normal and heavy water, respectively.

		$RT\ln Q$ (kJ/mol)		$RT\sum W_E^\Delta$ (kJ/mol)		$RT\sum W_A^\Delta$ (kJ/mol)	
		H_2O	D_2O	H_2O	D_2O	H_2O	D_2O
Bond Stretching	Acetone	-0.10	0.10	-0.08	0.11	-0.04	0.12
	Water	2.22	1.29	2.22	1.29	1.96	1.08
	Total	2.12	1.39	2.14	1.40	1.93	1.20
Angle Bending	Acetone	-0.60	-1.47	-0.55	-0.70	-0.34	-0.42
	Water	-0.27	-0.17	-0.27	-0.17	-0.20	-0.12
	Total	-0.87	-1.64	-0.82	-0.87	-0.54	-0.54
Torsion & Libration		-21.61	-20.29	-2.54	-1.06	-1.35	-0.54
Total		-20.35	-20.55	-1.21	-0.52	0.03	0.12

The pure water and acetone liquid and water-acetone mixtures were studied with different compositions by molecular dynamics simulations as well. And the IR spectrum could be obtained from equations in section 4.3.5.2B. The acetone IR spectrums with increasing acetone compositions are displayed in figure 4.8. The acetone spectrums for all compositions almost overlap each other. The only shifts are at the low frequency part near zero, and small red shifts with the increase of acetone are negligible. It is worthy to mention that $-CH_3$ group is represented as one particle in KBFF, a united atom force field. Hence, all the vibrations in $-CH_3$ group are missed in the molecular dynamics simulations. However, since no strong interactions between $-CH_3$ and water molecules, the composition (or environment) changes are not expected to affect the vibration frequencies in $-CH_3$ group very much, which can be supported by data in Table 4.9. In general, acetone vibration frequencies do not change with acetone composition and this is in agreement with Chapados' experimental data[24] and our quantum simulation results.

Figure 4.8 Acetone IR spectroscopy in acetone aqueous solutions with three compositions and pure acetone liquid.
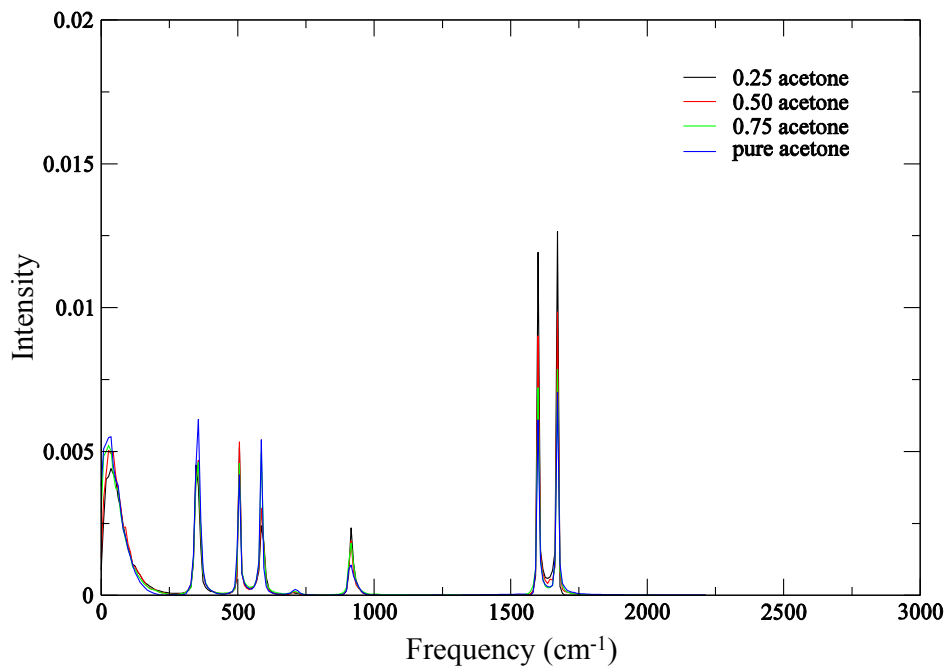
Figure 4.9 Water IR spectroscopy in pure water liquid and acetone aqueous solutions with three compositions.
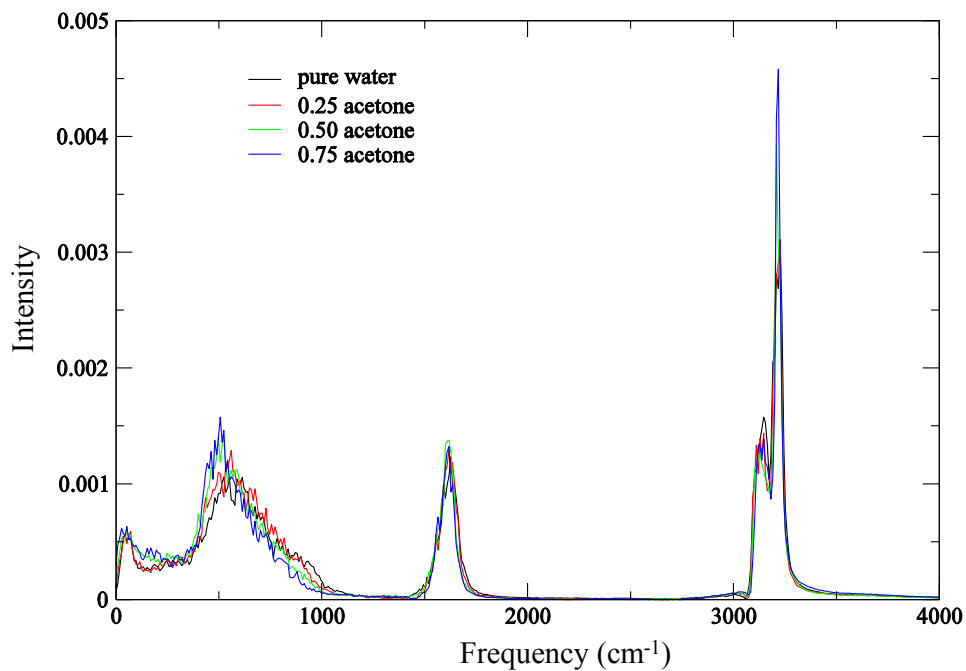
The water IR spectrum with decreasing water compositions is displayed in Figure 4.9. The peak located around 3200 cm^{-1}, which corresponds to the asymmetric stretching vibration, shows a noticeable red shift with decreasing water composition. This agrees with the large decrease when changing from water dimer to acetone-water complex in the density functional theory calculations. The low frequencies below 1000 cm^{-1}, which corresponds to torsions and librations, showed obvious red shifts when water composition decreased. This agrees with the quantum calculation results as well and supports that hydrogen bond switching or environment change has much larger effects on torsional/librational frequencies than bond stretching frequencies.

4.6 Discussion

The calculations above represent estimates of contributions of the vibrational partition function to the energy or free energy change for several processes in aqueous solutions. Although they involve significant degrees of approximation in various aspects, it is clear that the vibrational changes are often large and therefore need to be considered. It is possible that vibrational frequency shifts in high or low frequency regimes could cancel, such as the in the NMA and water quantum calculation, but this seems unlikely for most systems. Experimental frequency analysis in section 4.4 and simulated frequency analysis in section 4.5 exhibit quite different results, which means all vibrational frequencies give contributions to the free energy change. Therefore, only choosing a certain amount of characteristic frequencies in the analysis is not appropriate. Since the quantum corrections need to be performed for all vibrational frequencies, and low frequency vibrations can be treated classically and are supposed to be reasonably well reproduced by an accurate force field, the problem is mainly in the high frequency vibrations. In most molecular dynamics simulations, the high frequency vibrations are constrained in order to use a longer step time. Even though these vibrations are unconstrained, they are still not quantum and different from real vibrations.

Furthermore, it is clearly difficult to correct the results of classical simulations to compare with experiment for a variety of reasons including: 1) insufficient experimental spectra and insufficient spectra from molecular dynamics simulation on the density of vibrational states, 2) infeasibility to mimic the solution environment using density functional theory because of

high computational cost, 3) uncertainty as to whether the harmonic oscillator approximation is valid, and 4) the complicated role of solvent. These problems make the accurate evaluation of the total internal partition function essentially impossible with current approaches. On the basis of these difficulties, current popular force fields will be widely used now and in the near future at least, and some questions involving reliabilities and applications of classical force fields are discussed here.

So, why have classical force fields been able to provide useful data on many of the above processes? A possible answer lies in the approaches used during the parameterization procedure with the fact that the zero point vibrational energy for an interaction can to some extent be incorporated into an effective LJ/Coulomb potential. This point is illustrated from equation (4.30) to equations (4.34), which directly relate changes in the vibrational partition function to changes in the radial distribution function between solute species. For example, it is well known that hydrogen bonds in heavy water are considered to be stronger than in normal water, as exemplified by the larger heat of vaporization[11]. However, this is almost certainly a consequence of changes in the zero point energies, and other populated vibrational energy levels, rather than an increase or decrease in VDW or Coulomb interactions. Hence, in many ways, the interpretation of the vibrational changes in terms of nonbonded interactions has already been used effectively to describe the properties of deuterated systems.[55]

Fortunately, many force fields have already implicitly included some of the above changes in the vibrational partition function. Early force fields based on gas phase quantum calculations were typically not corrected to eliminate zero point energy contributions, nor were the experimental data corrected for quantum contributions to the enthalpy of vaporization for instance. Even for water models in which the quantum corrections were included, they were determined to be essentially negligible and therefore this had little effect on the results. However, recent estimates of the quantum contributions for water are not so negligible.[56,57] Alternatively, force fields which are parameterized to reproduce the experimental KB integrals as a function of composition have implicitly included the vibrational changes, in addition to environmentally dependent polarization effects, as indicated by equation (4.34). For mixtures, this can only be assured by other force field approaches if they reproduce the uncorrected chemical potentials for both the pure solute and the infinitely dilute solute in water. The former can present a problem if the pure solute is a solid. Take OPLS force field as another example, some of OPLS parameters

have been obtained by fitting to ab initio 6-31G(d) results for complexes of water with ions or molecules.[58,59] The changes in zero-point vibrational energy between monomer ion (or molecule) and monomer-water complex were calculated in ab initio studies, and took ~10% of total binding energy.[60] Therefore, the vibrational frequency shifts were included in the OPLS parameterization process, instead of showing in an explicit adjusting term when comparing with experimental thermal dynamic properties.

The consequences of the implicit inclusion of the vibrational partition function changes as an effective intermolecular interaction involve several discussions. Fundamentally, the interactions are different (bonded vs nonbonded) and so it is questionable whether such an approach will be successful. Certainly, the temperature dependence of the classical nonbonded interactions will not be the same as that of the quantum mechanical intermolecular vibrations, because the classical vibrational modes will not contribute to the change in enthalpy for a process as they obey the equipartition theorem, whereas the quantum mechanical modes will. Therefore, the parameterization of an effective classical force field at one temperature may produce errors at other temperatures. While this is clearly not an ideal situation, classical force fields have been shown to perform reasonably well as long as the temperature remains close to that used during the parameterization.

Other consequences of the above observations for force field development are that quantum corrections should not be explicitly taken into account when comparing with experimental data. Rather, the effects of changes in the vibrational partition functions, and therefore the chemical potentials of the species involved, should be included in the form of effective nonbonded interactions. In this way one does not have to perform the difficult process of correcting the results after the simulation is completed. For example, consider an experiment where one observes a 9999:1 (ΔG^o = -23.0 kJ/mol) mixture of native and unfolded protein that buries 50% of 100 residues. The previous example would suggest a contribution to the folding free energy from the NH and CO vibrational modes of -27.5 kJ/mol. In comparison, a classical simulation of the folding process, if this were possible, would require the final results to be corrected for any quantum effects due to the constrained high frequency bond stretching (primarily NH and CO). The above correction indicates that a simulation performed with an accurate force field, but without the implicit inclusion of the vibrational partition function corrections, would have generated a population for the native state corresponding to only 16%,

i.e. an effective free energy difference of $\Delta G^{o'} = 4.5$ kJ/mol. Such large corrections make the raw results of a classical simulation extremely difficult to interpret.

The above issues also have consequences for situations where force fields are modified from their original parameterization. For instance, transforming a force field which originally developed using rigid bonds into a flexible force field introduces the classical high frequency vibrations into the simulations. For protein folding this could significantly shift the observed equilibrium from that predicted by the same rigid bond force field. The classical contribution to the high frequency shifts typically observed in protein folding would contribute -0.06 kJ/mol from each peptide group. In the 100 residue 50% hydrogen bonded example, this equals to -3 kJ/mol, which is a small, but not insignificant ($\approx$RT), shift in the equilibrium depending on the particular process involved. Further shifts would be expected if a flexible water model was also applied. In principle, the whole force field should be reparameterized when introducing something as simple as bond flexibility, although this may not be warranted if the changes are small compared to the errors associated with the terms contributing to ΔW.

Another potential problem arises when trying to develop implicit solvent models for the study of long time biological processes such as protein folding. Here, one usually modifies a force field developed for explicit solvent simulations and attempts to determine the solvation free energy using some simplified equations. There are two ways this can cause problems. First, the removal of explicit solvent molecules means that the bending vibrations (the bonds are usually constrained) of solvent exposed hydrogen bonding groups on the protein will not be blue shifted as they typically would be in an explicit solvent – they will essentially vibrate at the gas phase frequencies. However, any groups that become buried on folding will be shifted if they form intramolecular hydrogen bonds. In effect, the vibrational frequencies of buried groups in the folded state will be similar to the explicit solvent simulation, whereas exposed groups will not. The absence of the (classical) blue shift in the denatured state due to the solvent will tend to stabilize the denatured state. Again, due to the potentially large number of groups that can be buried one can easily unbalance the equilibrium between folded and unfolded states.

Implicit solvation models also have to be reparameterized, although this is usually restricted to the solvation term. If the solvation term is parameterized directly using Poisson-Boltzmann calculations then the vibrational contribution of the solvent is not included in the continuum solvent electrostatic calculations. This may or may not be included in the surface area

term that is used to account for nonpolar solvent effects. A better parameterization would focus on reproducing the experimental free energies of solvation, uncorrected for quantum effects. In the latter case, the use of static solute structures during the parameterization procedure can potentially ignore the sizeable contribution from the classical bending vibrations, which can again lead to differences between the explicit solvent and implicit solvent results for processes such as protein folding.

A final consequence of a significant vibrational contribution to the chemical potential in solutions involves the experimental determination of radial distribution functions for liquids and liquid mixtures. In many cases this is achieved by isotopic substitution and the assumption that the rdfs remain unchanged in the isotopic system. From equations (4.33), (4.35) and (4.36), it is clear that this is, in principle, incorrect. Isotopic substitution, especially D for H, will affect the chemical potential of the solute and solvent, which will produce a change in G_{22}, indicating that g_{22} must also have changed. How significant these changes are will depend on the vibrational frequencies of that particular system. Simulated rdfs will only agree with the experimental data when isotope effects are small.

4.7 Conclusions

An attempt has been made to determine the changes in the quantum vibrational partition function for a series of processes in solution. The exact determination of these contributions from experimental or simulated results is difficult, but we suggest that these contributions are potentially significant, especially for protein folding. It is also argued that correcting the results of classical simulations afterwards is not feasible in the majority of cases. Therefore, it is suggested that the vibrational contributions should, and can, form an implicit part of the intermolecular interactions. Essentially, the fitting procedure should involve μ^* and not just W, or H* in the case of enthalpies of vaporization. Care must be taken as several potential problems can arise when well parameterized classical force fields are subsequently modified.

4.8 Supporting Information

The following tables provide more data and short descriptions about the frequencies obtained from the density functional theory calculation.

All the bond stretching vibrational frequencies for four systems (NMA dimer, water dimer and two different NMA-water complexes) with BP86 functional are listed in Table 4.5.

These frequencies did not shift much when it changes from folded state to unfolded state, except the N-H stretching frequency. It shows a red shift of 83 cm^{-1} during the folding process. On the other hand, O-H stretching frequency in the water molecule which worked as a hydrogen bond donor shows an obvious blue shift of 112 cm^{-1} in the folding process. The O-H frequency blue shift could be canceled for the most part by the red shift of the N-H stretching, and the sum of all bond stretching vibrational frequencies increased by only 18.2 cm^{-1}. Therefore, switching of the hydrogen bond between NMA and water does not have any significant effect on the total bond stretching vibrational frequencies.

Table 4.5 Bond stretching vibrational frequencies with BP86 for NMA and water four systems. D and A in the brackets refer to hydrogen bond donors and acceptor, respectively, and s and a are for symmetric and asymmetric vibrations, respectively.

	Folded state (cm^{-1})				Unfolded State (cm^{-1})			
	1		2		3		4	
	NMA(D)	NMA(A)	H$_2$O(A)	H$_2$O(D)	NMA(D)	H$_2$O(A)	NMA(A)	H$_2$O(D)
N-C amide	844	850			844		849	
C-C	967	964			961		964	
N-C	1081	1071			1077		1067	
C=O	1667	1656			1670		1658	
(N) C-H	2935	2957			2942		2955	
	2983	3014			2991		3026	
	3053	3074			3058		3064	
(C) C-H	2965	2968			2964		2966	
	3038	3047			3041		3047	
	3061	3048			3045		3047	
N-H	3353	3535			3436		3534	
H$_2$O(s)			3665	3499		3670		3387
H$_2$O(a)			3761	3740		3766		3739
total	66786				66768			

The bond stretching vibrational frequencies from B3LYP are displayed in Table 4.6 to compare with those from BP86. It is easy to notice that almost all the frequencies increased to the same extent, and the changes of the frequencies are quite similar to those from BP86. The O-H stretching blue shift and N-H stretching red shift can counteract each other, and during the folding process, the total bond frequencies decrease by only 15.5 cm^{-1}. Although it changed to an opposite direction to that with BP86, both of them can be considered the same as zero based on their very small values. And it indicates that the vibrational frequency shift does not have a significant dependence on the choice of functional.

Table 4.6 Bond stretching vibrational frequencies with B3LYP for NMA and water four systems. D and A in the brackets refer to hydrogen bond donors and acceptor, respectively, and s and a are for symmetric and asymmetric vibrations, respectively.

	Folded state (cm^{-1})				Unfolded State (cm^{-1})			
	1		2		3		4	
	NMA(D)	NMA(A)	H$_2$O(A)	H$_2$O(D)	NMA(D)	H$_2$O(A)	NMA(A)	H$_2$O(D)
N-C amide	875	867			865		860	
C-C	998	996			996		998	
N-C	1107	1096			1104		1079	
C=O	1721	1708			1723		1709	
(N) C-H	3003	3022			3003		3046	
	3050	3088			3050		3095	
	3126	3121			3126		3128	
(C) C-H	3026	3027			3026		3026	
	3094	3102			3097		3103	
	3011	3106			3108		3104	
N-H	3531	3643			3574		3692	
H$_2$O(s)			3775	3659		3782		3560
H$_2$O(a)			3865	3849		3872		3855
total	68566				68582			

Due to the extent of this chapter, we cannot include all the vibrational frequencies here, and only list the quantum contribution from stretching, bending, torsional and librational vibrational frequencies with BP86 in Table 4.7. As mentioned above, the total bond stretching vibration does not change much, and neither of the angle bending vibration. The sum of torsional and librational contribution shows a blue shift, which gives a total small blue shift in folding process. A small value of total quantum contribution shift appears in the results from B3LYP functional as well (not shown). Consequently, the total vibration contribution increased slightly in folding.

Table 4.7 Vibrational partition function contributions corresponding to different species of vibrations with BP86 functional in NMA and water systems

	Unfolded (kJ/mol)	Folded (kJ/mol)	Change in Folding (kJ/mol)
Bond stretching	396.85	396.96	0.11
Angle bending	214.84	213.81	-1.03
Torsion and Libration	-23.51	-32.20	-8.69
Total	588.18	578.57	-9.61

The frequencies of the folding process of NMA deuterated in amide were simulated using BP86, and quantum contributions from each vibrational species are listed in Table 4.8. Compared with Table 4.7 for normal NMA, the deuterated NMA shows a total larger blue shift for bond stretching and a total larger red shift for angle bending. Fortunately, these two shifts are able to almost cancel each other out. Therefore, the total quantum contribution remains as a small number which is close to that in normal NMA systems.

Table 4.8 Vibrational partition function contributions corresponding to different species of vibrations in deuterated NMA and water systems

	Unfolded (kJ/mol)	Folded (kJ/mol)	Change in Folding (kJ/mol)
Bond stretching	385.40	386.01	0.61
Angle bending	208.10	206.33	-1.77
Torsion and Libration	-24.52	-32.70	-8.18
Total	568.99	559.64	-9.35

Table 4.9 Bond stretching vibrational frequencies with BP86 functional for acetone and water three systems. s and a are for symmetric and asymmetric vibrations, respectively.

	Acetone dimer (cm^{-1})		water dimer (cm^{-1})		Acetone & water complex (cm^{-1})	
C-H	3069	3068			3072	
	3068	3067			3067	
	3017	3016			3013	
	3011	3010			3008	
	2953	2952			2953	
	2946	2946			2946	
C=O	1700	1688			1685	
C-C-C(a)	1202	1195			1213	
C-C-C(s)	767	766			773	
H$_2$O(a)			3761	3740		3740
H$_2$O(s)			3665	3499		3405
Total	58108				28875x2=57750	

In the acetone-water complex, which is listed in Table 4.9, the water asymmetric and symmetric stretching frequencies are 3740 and 3405 cm^{-1}, respectively. Compared with the corresponding frequencies of water dimer, both of them exhibit large blue shifts in the acetone-water separation process, especially the symmetric stretching. All of the stretching frequencies in acetone molecules do not show any significant change.

The quantum contributions corresponding to different species of vibrations involved in acetone and water separation process are listed in Table 4.10. Compared with the water stretching blue shift, the acetone stretching shifts could be neglected. Contrarily, the acetone angle bending vibration shows a total remarkable red shift, and finally gives a large decrease in the total contribution during water-acetone separation process. The torsional and librational contributions changed in the same direction but had a larger magnitude, and gave the dominant contribution to the vibrational partition function. In summary, without the N-H, the vibrational partition function contribution changes cannot be canceled by themselves, and would leave a large contribution to the chemical potential and free energy changes.

Table 4.10 Vibrational partition function contributions corresponding to different vibration and molecule species in acetone and water systems

	Acetone dimer & Water dimer (kJ/mol)		Acetone & water complex (kJ/mol)		Separation Change (kJ/mol)
Bond stretching	Acetone	258.18	Acetone	258.28	-0.10
	Water	87.20	Water	84.98	2.22
	Total	345.37	Total	343.25	2.12
Angle bending	Acetone	152.50	Acetone	153.10	-0.60
	Water	19.19	Water	19.46	-0.27
	Total	171.69	Total	172.56	-0.87
Torsion & Libration	-27.14		-5.54		-21.61
Total	489.92		510.28		-20.35

All these results above mean that a total large red shift exists when the acetone-water mixture separated. However, if it was true, according to the isotopic effect mentioned above, it would have a large change on the vibrational partition function when normal water is substituted by heavy water. Assuming the intermolecular interaction W should not be influenced by the deuteration, the large change on the vibrational partition function would finally result in a huge alteration in the free energy. We have to point out that the accuracy of the acetone dimer frequency cannot be guaranteed. The acetone dimer structure is totally artificial, which is just one reasonable structure and could have some deviation from acetone-acetone interaction model in reality. The strong dipole-dipole interaction exists between two vertically parallel carbonyl groups in acetone dimer, not acetone-water complex, while the stronger hydrogen bonding interaction exists between acetone and water in a horizontal plane. Hence from acetone dimer to acetone-water complex, the intermolecular interaction changes in not only the strength, but also the orientation. In real world, where the acetone is surrounded by plenty of molecules and the intermolecular interaction is represented by the sum of individual ones, the changes in orientation is not noticeable, and this could be the reason why these huge shifts in our quantum simulation cannot happen in experimental IR spectroscopy. However, considering the computational cost in quantum calculations and some practical problems, such as distinguishing different species of vibration for different molecules, it is hard and impossible to calculate and analyze the acetone cluster in which one acetone molecule is surrounded by others.

Table 4.11 Frequency comparison between normal water dimer and heavy water dimer

	H_2O (cm^{-1})	D_2O (cm^{-1})	Ratio
Bond stretching	3761	2931	1.38
	3740	2702	1.38
	3665	2611	1.40
	3499	2513	1.39
Angle bending	1627	1239	1.31
	1601	1218	1.31
Libration	687	500	1.37
	408	324	1.26
	215	183	1.17
	191	141	1.35
	174	137	1.27
	129	92	1.40

For the water dimer, heavy water frequencies decrease from normal water by the ratio of 1.3~1.4, which is consistent with the expected ratio according to the reduced mass change. However, to the contrary, most acetone vibrational frequencies listed in Table 4.12, in heavy water-acetone complex do not show any remarkable changes. It is worthy to point out that two torsional and librational frequencies, 684 and 425 cm^{-1} in normal water, red shifted to 471 and 281 cm^{-1} in heavy water and these two frequencies happen to red shifted to around 100 cm^{-1} in acetone complex. It is possible that the changes in these two frequencies from with normal water to with heavy water could be modified by adding more surrounding molecules. It suggests that the deuteration in water only have a large effect on some angle bending vibrations involving carbonyl group, which is in the vicinity of water molecule, but not the bond stretching vibrations and angle bending vibrations which are further away from water.

Table 4.12 Frequency comparison between normal water-acetone complex and heavy water-acetone complex

		H_2O (cm^{-1})	D_2O (cm^{-1})	ratio
Bond Stretching	Water	3740	2697	1.39
		3405	2473	1.38
	Acetone	3072	3074	1.00
		3067	3059	1.00
		3013	3012	1.00
		3008	3007	1.00
		2953	2952	1.00
		2946	2944	1.00
		1685	1683	1.00
		1213	1211	1.00
		773	771	1.00
Angle bending	Water	1637	1243	1.32
	Acetone	1439	1441	1.00
		1423	1423	1.00
		1412	1413	1.00
		1411	1411	1.00
		1341	1341	1.00
		1335	1334	1.00
		1076	1076	1.00
		1052	1052	1.00
		873	873	1.00
		853	853	1.00
		536	539	0.99
		473	479	0.99
		376	376	1.00
		118	129	0.91
		97	101	0.96

	684	471	1.45
	425	281	1.51
Torsion & Libration	180	164	1.09
	84	81	1.04
	79	72	1.10
	41	46	0.89

Table 4.13 Vibrational partition function contributions comparison between acetone-water complexes with normal and heavy water, respectively

		H$_2$O (kJ/mol)		D$_2$O (kJ/mol)		Ratio
Water dimer & Acetone dimer	Bond stretching	Acetone	258.18	Acetone	258.18	1.00
		Water	87.20	Water	62.77	1.39
		Total	345.37	Total	320.95	1.08
	Angle bending	Acetone	152.50	Acetone	152.50	1.00
		Water	19.19	Water	14.60	1.31
		Total	171.69	Total	167.10	1.03
	Torsion & Libration	-27.14		-31.81		0.85
	Total	489.92		456.24		1.07
Acetone-water complex	Bond stretching	Acetone	258.28	Acetone	258.08	1.00
		Water	84.98	Water	61.48	1.38
		Total	343.25	Total	319.56	1.07
	Angle bending	Acetone	153.10	Acetone	153.97	0.99
		Water	19.46	Water	14.77	1.32
		Total	172.56	Total	168.74	1.03
	Torsion & Libration	-5.54		-11.51		0.48
	Total	510.28		476.79		1.07

The quantum contributions in various vibration species in phase separated state and mixture state with heavy water compared with normal water are listed in Table 4.13. Most contributions shifted in water dimer or water in acetone-water complex, and these shifts lead to the decreases in the total contributions by the same ratio for both separated and combined states.

The solvent isotopic effect was also evaluated in NMA-water systems. Table 4.14 exhibits the quantum contribution in NMA-H_2O and NMA-D_2O complexes. The total contributions of NMA-D_2O complex are much less than those of NMA-H_2O complex and most of the differences are from the water. The complex with a hydrogen bond between amide in NMA and oxygen in water has larger isotopic effect than that with a hydrogen bond between carbonyl and hydrogen in water.

Table 4.14 Vibrational partition function contributions comparisons between NMA-water complexes with normal and heavy water, respectively.

		(NMA)CO---H_2O (kJ/mol)		(NMA)NH---OH_2 (cm^{-1})	
		in H_2O	in D_2O	in H_2O	in D_2O
Bond Stretching	NMA	155.58	155.46	154.68	154.57
	Water	42.38	30.71	44.21	31.84
	Total	197.95	186.16	198.89	186.40
Angle Bending	NMA	98.16	98.01	97.41	97.53
	Water	9.76	7.37	9.51	7.22
	Total	107.92	105.39	106.92	104.75
Libration & Torsion		-8.71	-10.89	-14.79	-11.39
Total		297.16	280.66	291.02	266.53

4.9 References

(1)   Berens, P. H.; Mackay, D. H. J.; White, G. M.; Wilson, K. R. *J. Chem. Phys.* **1983**, *79*, 2375.

(2)   Jorgensen, W. L.; Madura, J. D. *Mol. Phys.* **1985**, *56*, 1381.

(3)   Wilson, J., E. B.; Decius, J. C.; Cross, P. C. *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra*; Dover Publications, 1955.

(4)   Levine, I. N. *Molecular Spectroscopy*; John Wiley & Sons, 1975.

(5)   Levine, I. N. *Quantum Chemistry*; Prentice Hall, 2000.

(6)   McQuarrie, D. A. *Statistical Mechanics*; Harper&Row, 1976.

(7)   Reimers, J. R.; Watts, R. O.; Klein, M. L. *Chem. Phys.* **1982**, *64*, 95.

(8)   Brubach, J. B.; Mermet, A.; Filabozzi, A.; Gerschel, A.; Roy, P. *J. Chem. Phys.* **2005**, *122*.

(9)   Owicki, J. C.; Scheraga, H. A. *J. Am. Chem. Soc.* **1977**, *99*, 7403.

(10)   Ben-Naim, A. *Molecular Theory of Solutions*; Oxford University Press, New York, 2006.

(11)   Marcus, Y.; Ben-naim, A. *J. Chem. Phys.* **1985**, *83*, 4744.

(12)   Berens, P. H.; Wilson, K. R. *J. Chem. Phys.* **1981**, *74*, 4872.

(13)   Herzberg, G. *Spectra of Diatomic Molecules*; Van Nostrand, **1950**.

(14)   Wilson, J., m E. B. *Ann. Rev. Phys. Chem.* **1951**, *2*, 151.

(15)   Mueller, M. *Fundamentals of Quantum Chemistry: Molecular Spectroscopy and Modern Electronic Structure Compurations*; Kluwer Academic Publishers, 2002.

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

(17)   Marcus, Y. *Ion Solvation*; Wiley, Chichester, 1985.

(18)   Koddermann, T.; Schulte, F.; Huelsekopf, M.; Ludwig, R. *Angew. Chem. Int. Edit.* **2003**, *42*, 4904.

(19)   Lock, A. J.; Bakker, H. J. *J. Chem. Phys.* **2002**, *117*, 1708.

(20)   Marin, T. W.; Takahashi, K.; Bartels, D. M. *J. Chem. Phys.* **2006**, *125*.

(21)   Postma, J. P. M., University of Groningen, The Netherlands, 1985.

(22)   Woldan, M. *Acta. Univ. Lodziensis, Folia. Chim.* **1988**, *8*, 35.

(23)   Holden, C. A.; Hunnicutt, S. S.; Sanchez-Ponce, R.; Craig, J. M.; Rutan, S. C. *Appl. Spectrosc.* **2003**, *57*, 483.

(24)   Max, J. J.; Chapados, C. *J. Chem. Phys.* **2003**, *119*, 5632.

(25)   Marshall, P. R.; Katz, J. J. *J. I. N. C.* **1974**, *36*, 1589.

(26)   Franks, F. *Water: 2nd Edition A matrix of life*; Royal Society of Chemistry, Cambridge, 2000.

(27)   Kirkwood, J. G.; Buff, F. P. *J. Chem. Phys.* **1951**, *19*, 774.

(28)   Bonner, O. D. *J. Chem. Thermodyn.* **1971**, *3*, 837.

(29)   Schrier, M. Y.; Schrier, E. E. *J. Chem. Thermodyn.* **1973**, *5*, 811.

(30)   R. A. Robinson; Stokes, R. H. *Electrolyte Solutions*; LONDON BUTTERWORTHS, 1959.

(31)   Rossi, M.; Scheffler, M.; Blum, V. *J Phys Chem B* **2013**, *117*, 5574.

(32)   Maffi, C.; Baiesi, M.; Casetti, L.; Piazza, F.; De Los Rios, P. *Nat. Commun.* **2012**, *3*.

(33)   Zhang, R.; Li, H. R.; Lei, Y.; Han, S. J. *J. Mol. Struct.* **2004**, *693*, 17.

(34)   Parker, M. J.; Clarke, A. R. *Biochemistry-Us* **1997**, *36*, 5786.

(35)   Huyghues-Despointes, B. M. P.; Scholtz, J. M.; Pace, C. N. *Nat. Struct. Biol.* **1999**, *6*, 910.

(36)   Scheiner, S.; Cuma, M. *J. Am. Chem. Soc.* **1996**, *118*, 1511.

(37)   Max, J. J.; Chapados, C. *J. Chem. Phys.* **2004**, *120*, 6625.

(38)   Benedetti, A. V.; Cilense, M.; Vollet, D. R.; Montone, R. C. *Thermochim. Acta.* **1983**, *66*, 219.

(39)   Pereyra, R. G.; Asar, M. L.; Carignano, M. A. *Chem. Phys. Lett.* **2011**, *507*, 240.

(40)   Khurma, J. R.; Fenby, D. V. *Aust. J. Chem.* **1981**, *34*, 635.

(41)   te Velde, G.; Bickelhaupt, F. M.; Baerends, E. J.; Guerra, C. F.; Van Gisbergen, S. J. A.; Snijders, J. G.; Ziegler, T. *J. Comput. Chem.* **2001**, *22*, 931.

(42)   Yanai, T.; Tew, D. P.; Handy, N. C. *Chem. Phys. Lett.* **2004**, *393*, 51.

(43)   Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory. Comput.* **2008**, *4*, 435.

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

(45)   Weerasinghe, S.; Smith, P. E. *J. Chem. Phys.* **2003**, *118*, 10663.

(46)   Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.

(47)   York, D. M.; Darden, T. A.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *99*, 8345.

(48)   Itzhaki, L. S.; Evans, P. A. *Protein Sci.* **1996**, *5*, 140.

(49)   Guzzi, R.; Sportelli, L.; La Rosa, C.; Milardi, D.; Grasso, D. *J. Phys. Chem. B* **1998**, *102*, 1021.

(50)   Masson, P.; Laurentie, M. *Biochim. Biophys. Acta.* **1988**, *957*, 111.

(51)   Krantz, B. A.; Srivastava, A. K.; Nauli, S.; Baker, D.; Sauer, R. T.; Sosnick, T. R. *Nat. Struct. Biol.* **2002**, *9*, 458.

(52)   Krantz, B. A.; Moran, L. B.; Kentsis, A.; Sosnick, T. R. *Nat. Struct. Biol.* **2000**, *7*, 62.

(53)   Shi, Z. S.; Krantz, B. A.; Kallenbach, N.; Sosnick, T. R. *Biochemistry-Us* **2002**, *41*, 2120.

(54)   Shi, Z. S.; Olson, C. A.; Kallenbach, N. R.; Sosnick, T. R. *J. Am. Chem. Soc.* **2002**, *124*, 13994.

(55)   Sarangi, S. S.; Reddy, S. K.; Balasubramanian, S. *J. Phys. Chem. B* **2011**, *115*, 1874.

(56)   Ojamae, L.; Hermansson, K.; Probst, M. *Chem. Phys. Lett.* **1992**, *191*, 500.

(57)   Schmidt, J. R.; Corcelli, S. A. *J. Chem. Phys.* **2008**, *128*, 184504.

(58)   Jorgensen, W. L.; Gao, J. *J. Am. Soc.* **1988**, *110*, 4212.

(59)   Jorgensen, W.L.; Tirado-Rives, J. *J. Am. Soc.* **1988**, *5*, 411.

(60)   Gao, J., Garner, D. S. and Gorgensen, W. L. *J. Am. Chem. Soc.* **1986**, *108*, 4784.

Chapter 5 - Summary and Future Work

With the growing power of available computational resources, molecular dynamics simulation will provide more insights into intermolecular interactions and play a more significant role in biological systems. Kirkwood-Buff theory provides a direct connection between the relative distribution of species in solution and the thermodynamics properties of the solution. This theory can be applied on both experimental results and computational data. Therefore, a combination of KB theory and MD simulation becomes a powerful tool to study biomolecular behavior in the condensed phase at atomic level. The preceding chapters have shown our attempts to derive a extend protein force field with KB theory, the application of the new force fields and a KB theory to the study of urea cosolvent effects, and suggested the advantage of a KB derived force field in which the vibrational partition function contribution was implicitly included.

The Smith Group has devoted many years to develop a full biomolecular force field based on KB theory, and a full force field for proteins is almost complete. We are now working on the parameters for phospholipids. With the continuous improvement in the quality of the current parameters set, we hope KB theory and KB force field can help us unveil the mechanisms of many biological behaviors, such as protein folding, aggregation and ligand binding.