# Atomistic simulations of cosolvent/water mixtures:

## Force field development and application to cosolvent effects on nonpolar solubility and hydrophobic interactions

Dissertation
Zur Erlangung des Grades

"Doktor der Naturwissenschaften"

am Fachbereich Chemie, Pharmazie und Geowissenschaften
der Johannes Gutenberg-Universität
in Mainz

Maeng Eun Lee

geboren in Jeonju, Südkorea

Mainz 2006

# Zusammenfassung

In dieser Arbeit werden hydrophobes Lösungsverhalten und hydrophobe Wechselwirkungen in binären Lösungsmittel-Wasser-Mischungen mit Hilfe atomistischer Simulationen untersucht. Viele dieser Mischungen zeigen nicht-ideales Verhalten; die Lösungsmittelmoleküle aggregieren, obwohl sich auf makroskopischer Ebene stabile und homogene Mischungen ausbilden. Auf Kraftfeldern basierende atomistische Simulationen eröffnen eine Möglichkeit, atomare Strukturen und Wechselwirkungen mit den thermodynamischen Eigenschaften des Lösungsmittels in Verbindung zu bringen. Die erwarteten Eigenschaften hängen sehr stark von den Werten der Parameter ab, mit denen die molekularen Wechselwirkungen beschrieben werden.

Im ersten Teil der Arbeit wird als binäre Mischung t-Butanol (TBA) und Wasser verwendet. Ausgehend von der Kirkwood-Buff-Theorie für Lösungen wird ein neues Kraftfeld entwickelt, das es ermöglicht, die Ausbildung von Alkohol-Alkohol-, Wasser-Wasser- und Alkohol-Wasser-Aggregaten in der Mischung zu beschreiben. Weiterhin können die Ableitungen der chemischen Potentiale der an der Mischung beteiligten Stoffe in Übereinstimmung mit experimentellen Daten bestimmt werden.

Mit dem erstellten Kraftfeld ist das bevorzugte Lösungsverhalten und die Thermodynamik einer hydrophoben Verbindung in TBA-Wasser-Mischungen am Beispiel von Methan studiert worden. Dazu wurden Freie Energie-, Enthalpie- und Entropieänderungen des Lösungsvorgangs für verschieden konzentrierte TBA-Wasser-Mischungen berechnet und mit experimentellen Werten verglichen. Im Gegensatz zur Freien Energie verändern sich Enthalpie und Entropie nicht-monoton bei Änderung des Mischungsverhältnisses. Dies lässt sich auf die strukturelle Umverteilung des Lösungsmittels zurückführen, welche keinen Einfluss auf die Freie Energie ausübt, da sich di Beiträge von Enthalpie und Entropie nahezu kompensieren. Daher kann aus dem Verhalten von Lösungsenthalpie und -entropie ein genaueres Bild über die Anordnung der Lösungsmittelmoleküle um die gelöste Substanz gewonnen werden.

Im zweiten Teil werden hydrophobe Wechselwirkungen in Harnstoff-Wasser-Mischungen untersucht. Diese sind von großer Bedeutung in der Biologie (z.B. für das Falten bzw. Entfalten von Proteinen). Dennoch sind die Änderungen der Wechselwirkungen, die durch Erhöhen des Harnstoffanteils verursacht werden, noch nicht vollständig aufgeklärt. Dieses Problem wird anhand der Änderungen der Freien Energie (potential of mean force), der Enthalpie und der Entropie als Funktion des Abstandes zweier unpolarer Moleküle in Wasser und in einer wässrigen Harnstoff-Lösung (6.9 M) diskutiert. Dazu werden in Kapitel 5 die Enthalpie- und der Entropiebeiträge zum "potential of mean force" untersucht, die von Umstrukturierungen im Lösungsmittel abhängen. Dabei stellt man fest, dass der Zusammenschluss unpolarer Moleküle in einer wässrigen Harnstoff-Lösung genau wie in reinem Wasser thermodynamisch bevorzugt wird. Dies steht im Widerspruch zur bisherigen Meinung, dass sich Cluster unpolarer Moleküle vollständig in wässrigen Harnstoff Lösungen auflösen. Die Konsequenzen dieser Beobachtung für die Stabilität von Proteinen in konzentrierten Harnstofflösungen wird eingehend in Kapitel 6 diskutiert.

# Summary

In this thesis, atomistic simulations are performed to investigate hydrophobic solvation and hydrophobic interactions in cosolvent/water binary mixtures. Many cosolvent/water binary mixtures exhibit non-ideal behavior caused by aggregation at the molecular scale level although they are stable and homogenous at the macroscopic scale. Force-field based atomistic simulations provide routes to relate atomistic-scale structure and interactions to thermodynamic solution properties. The predicted solution properties are however sensitive to the parameters used to describe the molecular interactions.

In this thesis, a force field for tertiary butanol (TBA) and water mixtures is parameterized by making use of the Kirkwood-Buff theory of solution. The new force field is capable of describing the alcohol-alcohol, water-water and alcohol-water clustering in the solution as well as the solution components' chemical potential derivatives in agreement with experimental data. With the new force field, the preferential solvation and the solvation thermodynamics of a hydrophobic solute in TBA/water mixtures have been studied.

First, methane solvation at various TBA/water concentrations is discussed in terms of solvation free energy-, enthalpy- and entropy- changes, which have been compared to experimental data. We observed that the methane solvation free energy varies smoothly with the alcohol/water composition while the solvation enthalpies and entropies vary non-monotonically. The latter occurs due to structural solvent reorganization contributions which are not present in the free energy change due to exact enthalpy-entropy compensation. It is therefore concluded that the enthalpy and entropy of solvation provide more detailed information on the reorganization of solvent molecules around the inserted solute.

Hydrophobic interactions in binary urea/water mixtures are next discussed. This system is particularly relevant in biology (protein folding/unfolding), however, changes in the hydrophobic interaction induced by urea molecules are not well understood. In this thesis, this interaction has been studied by calculating the free energy (potential of mean force), enthalpy and entropy changes as a function of the solute-solute distance in water and in aqueous urea (6.9 M) solution. In chapter 5, the potential of mean force in both solution systems is analyzed in terms of its enthalpic and entropic contributions. In particular, contributions of solvent reorganization in the enthalpy and entropy changes are studied separately to better understand what are the changes in interactions in the system that contribute to the free energy of association of the nonpolar solutes. We observe that in aqueous urea the association between nonpolar solutes remains thermodynamically favorable (i.e., as it is the case in pure water). This observation contrasts a long-standing belief that clusters of nonpolar molecules dissolve completely in the presence of urea molecules. The consequences of our observations for the stability of proteins in concentrated urea solutions are discussed in the chapter 6 of the thesis.

# Table of Contents

## Chapter 3. A new force field of aqueous tertiary butanol solutions using atomistic Simulations

## Chapter 1

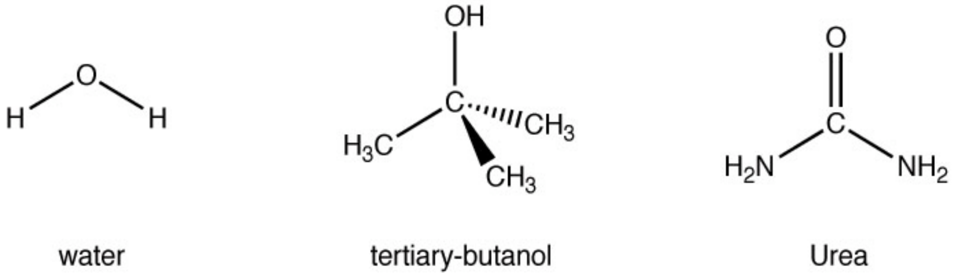
# Introduction

Many biological processes are a matter of chemistry in aqueous solutions. The model studies can help to gain insight into the thermodynamics of these processes. Molecular interactions such as hydrophilic, hydrophobic and hydrogen bonding play an important role in aqueous mixtures. How these interactions affect the structure and thermodynamics of these mixtures, is central question. For the case of alcohol/water binary mixture, water and alcohol molecules are a homogeneous and stable mixture at the macroscopic scale, but inhomogeneities appear at the microscopic level because alcohol and water tend to self-aggregate and a maximum of aggregation is present at intermediate concentrations.[1, 2] Atomistic simulation can provide an insight into molecular scale behavior of those alcohol/water binary mixtures that cannot be easily observed by means of conventional experimental studies. In particular, molecular dynamics (MD) can simulate the time dependent behavior of molecular systems and provide the structure, dynamics and thermodynamics of molecules in solution. The features of simulations of alcohol/water mixtures are usually very sensitive to the atomistic models (force fields) invoked. Generally force fields are parameterized using quantum mechanical calculations combined with an empirical approach for the non-bonded interaction parameters, which are tuned in order to reproduce liquid state properties such as densities, vaporization enthalpy, viscosity, diffusion constant, dielectric permittivity, *etc*. of the *pure* fluid. In my thesis, a force field for binary liquid mixtures of *tert*-butanol and water is developed based on the Kirkwood-Buff (KB) theory of solution.[3] Recently, Weerasinghe and Smith have developed force fields based on KB theory for solution mixtures (*e.g.,* acetone/water,[4] methanol/water,[5] urea/water,[6] sodium chloride/water,[7] guanidium chloride/water[8]). Their KB derived force fields reproduce thermodynamics and molecular aggregation behavior in the binary mixtures. KB theory was first introduced in 1955, when Kirkwood and Buff formulated the relations for the derivatives of the chemical potentials with respect to concentrations, the partial molar volumes and compressibility in terms of volume integrals of the pair correlation functions.[3] Their formula was a one-way theory leading from the pair correlation functions to thermodynamic quantities through the so-called KB integrals (KBIs). After a long absence of studies on KB theory, in 1977, Ben-Naim has reported an inversion procedure using the measure of the thermodynamic quantities applied to ethanol and water mixtures to provide an insight into the molecular structure.[9] In the following years, few studies have been reported by Donkersloot for methanol/water and ethanol/water[10] and Patil for *tert*-butanol/water.[11] Then, in 1984, Matteoli and Lepori extended the KB theory for various binary mixtures such as methanol, ethanol, *1*-propanol, *1*-butanol, *2*-methyl-*2*-propanol, acetonitrile, acetone, dimethyl-sulfoxide, tetrahydrofuranm piperidine, pyridine, *1,4*-dioxane, *2*-aminoethanol, and *2*-(dimethylamino)ethanol.[1] They suggested that one of the specific features of the aqueous/organic solutions was the existence of microheterogeneity[1, 12]; *i.e.,* inhomogeneities at the molecular scale, particularly at intermediate concentrations. The link between microscopic structure and macroscopic properties involving excess chemical potentials, partial molar volumes, and compressibilities has been successfully provided by the KBIs. Nishikawa was the first to provide X-ray scattering data[13, 14] of the KB integrals. Neutron scattering measurements have been provided by Misawa *et al.* and Almásy *et al.*[15, 16] The microheterogeneities in aqueous methanol mixtures have been recently studied by Dixit *et al.*[17] using neutron-scattering techniques coupled with a reverse Monte Carlo procedure. Computer simulations attempted by several authors[4, 18-23] have provided further visual evidence that these mixtures are highly inhomogeneous at the microscopic scale.

Non-ideal behaviors in binary mixtures have been investigated with nonpolar solute solvation thermodynamics. Ben-Naim reported thermodynamic behavior of argon in aqueous solutions such as ethanol/water,[24] methanol/water,[25] *p*-dioxane/water[26] and ethylene-glycol/water[27]. He suggested that the effect of co-solvent on the structure of water plays an important role on the solvation thermodynamics of argon. Some interesting features arise as the qualities of solvent are changed from *pure* water to *pure* organic solvent. General features in binary mixtures are the following. Enthalpy and entropy curves increase steeply from both the effects of argon and co-solvent on the structure of water in the range of $0 \le x(\text{co-solvent}) \le 0.25$ and constantly converge in the high concentration range. He suggested that these features come from large clusters of water molecules linked by hydrogen bonds, and thus depend on the average cluster size. Increasing fraction of co-solvent to water causes a steep decrease of the average water cluster size. As a result, large water clusters are destabilized in the range of $0 \le x(\text{co-solvent}) \le 0.25$. Beyond that region, large water clusters do not exist, so that a gas solute (argon) “sees” a normal liquid mixture. The co-solvent is supposed to destabilize the structure of water at dilute co-solvent concentration. Ben-Naim decomposed the solvation enthalpy and entropy changes of nonpolar solute in binary mixtures into two terms; *i.e.,* a solute binding energy term and a solvent relaxation contribution term. In fact, the solute-solvent binding energy is a weak dispersion energy. The enthalpy and entropy changes will be strongly affected by the solvent reorganization energy. However, increases in both the enthalpy and entropy changes are compensated in the free energy denoted by $\Delta G = \Delta H - T\Delta S$.

Enthalpy and entropy changes – unlike free energy changes – have not been extensively studied separately in the past using atomistic computer simulations due to the large computational expense. Yet, aspects of solvation and reorganization of solvent molecules can usually be probed better through the entropy and enthalpy changes. In order to explain experimentally observed anomalies in the enthalpy and entropy changes, it is studied in terms of atomic-scale solvation mechanisms. With respect to atomistic models, calculations of enthalpy and entropy changes provide also a mean to better assess the quality of the force fields.

In this thesis, two different binary liquid mixtures were examined, in which both hydrophobic solvation and hydrophobic interactions play an important role: 1. aqueous tertiary-butanol (TBA) mixtures and 2. aqueous urea mixtures. Figure 1.1 shows the chemical structures of the molecules. Aspects of hydrophobic solvation and hydrophobic interactions in these solutions are discussed with a focus on the atomic-scale distribution of the solvent molecules and solvation mechanisms.



water

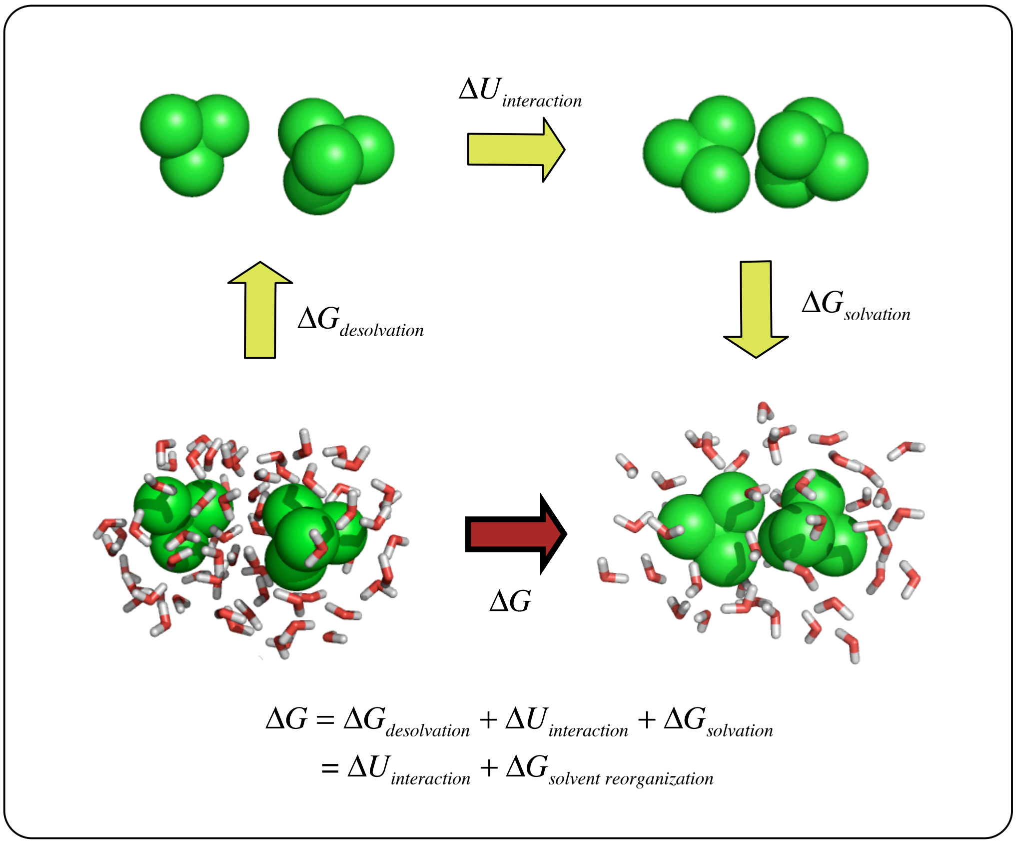
tertiary-butanol

Urea

*Figure 1.1 Chemical structures for water and tertiary-butanol and urea as co-solvents which are simulated in this thesis.*

It has been shown that highly non-ideal behavior occurs in dilute aqueous TBA solutions[28-32] because TBA has the largest nonpolar moiety among the series of alcohols fully miscible with water. This fact makes reliable descriptions of this system based on classical atomistic models very challenging. The solvation thermodynamics of inserting a methane solute in aqueous TBA mixtures is investigated in my thesis by comparing simulation results from our KB derived force field with experimental data on methane solvation free energies, enthalpies, and entropies.

Furthermore, association of two nonpolar solutes (hydrophobic interaction) is also investigated in water and in aqueous urea solution. Figure 1.2 shows the association scheme of two nonpolar solutes in water. Association from solvent-separated distance (left) to contact distance (right) between two nonpolar solutes can be decomposed into three steps (de-solvation, solute association, then solvation). In here, de-solvation and solvation processes are related to solvent-reorganization. Thus, the association process can be considered as solute interaction and the changes of its surrounding solvent. In case of hydrophobic interaction in pure water, water forms a fluctuating hydrogen-bonded network, and insertion of solute molecules cleaves this network, which reorganizes itself around the solutes and eventually clusters them together. The nature of the solute molecule (*e.g.,* the various polar and nonpolar parts that contain the charge distributions) plays an equally important role, but the crucial part is the bonded nature of the solvent. The topology and strength of the hydrogen-bonded network of water give this particular liquid more anomalous properties than any other hydrogen-bonded solvent.[33]



$$\begin{aligned}\Delta G &= \Delta G_{desolvation} + \Delta U_{interaction} + \Delta G_{solvation} \\ &= \Delta U_{interaction} + \Delta G_{solvent\ reorganization}\end{aligned}$$

*Figure 1.2 Association scheme of two nonpolar solutes (green) in water. Solvation free energy change ( $\Delta G$ ) from solvent separated distance to contact distance can be considered by de-solvation at the solvent separated distance ( $\Delta G_{desolvation}$ ), association of two solutes ( $\Delta U_{interaction}$ ), and solvation at the contact distance ( $\Delta G_{solvation}$ ). In particular, de-solvation and solvation processes are related to solvent-reorganization.*

Urea is well known as a protein denaturant and is of high importance in protein folding research. In general, hydrophobic interactions are not stable in aqueous urea solution relative to water and this fact is generally used to argue that urea favours solvent exposure of the nonpolar interior of proteins.[33-38] The mechanism of protein denaturing however still is not well understood today.[38-46] In my thesis, I studied hydrophobic interactions in water and in urea/water mixtures in terms of potentials of mean force obtained from constraint-bias simulations with force averaging. The so-obtained potentials of mean force (*i.e.,* free energies of association along a radial distance coordinate) are decomposed in enthalpic and entropic parts, which are further separated in solute-solvent and solvent-solvent contributions. Altogether this provides a complete picture on the thermodynamics of solute association in the binary mixtures.

The rest of this thesis is organized into six chapters. Chapter 2 describes shortly the techniques used in molecular dynamics simulation and analytical methods to calculate thermodynamic properties of liquid mixtures. Force field development and thermodynamics for the TBA and water mixtures are discussed in chapter 3. The following chapter 4 focuses on the methane solvation in TBA and water mixtures, comparing several force fields over a wide range of TBA concentrations. Our primary attention lies on the dilute TBA concentrations (TBA mole fractions smaller than 0.2) that show highly anomalous behavior of the enthalpy and entropy changes of methane solvation. The quality of the prediction and interpretation of methane solvation thermodynamics using our new force field (LV), discussed in chapter 3, are emphasized in this section. Pair hydrophobic interactions (aliphatic-aliphatic, aromatic-aromatic), induced by the present of urea comparing pure water, are studied in chapter 5 and 6. Chapter 5 will discuss the association free energy of *neo*-pentane - *neo*-pentane in pure water and in an aqueous urea solution in terms of the contributions of the entropy and enthalpy. In particular, enthalpy-entropy compensation in the free energy of hydrophobic interaction is emphasized. Chapter 6 describes new observations of the role of urea on the stability of hydrophobic interactions using the potentials of mean force of *neo*-pentane, toluene, *3*-methylindole as prototypes of nonpolar amino side groups. These new observation is discussed with implication for protein denaturation. The conclusions are given in chapter 7.

# 2. Computational Methods

## 2.1 Introduction

Computational chemistry techniques have become an important tool for studies of energetic, structural, and dynamic aspects of molecular systems. These methods have been used for more than four decades to simulate the behavior of systems from simple molecules to complex biomolecular systems. Simulation techniques are constantly bring improved also through comparison with experimental data, and there now exists an array of methods at the computational chemists disposal. These approaches can be divided into two main categories, quantum chemistry methods, and statistical mechanics methods.

Among the various computational chemistry methods, the most detailed are *ab initio* quantum mechanical (QM) methods which predict the atomic electronic structure. Using a QM level calculation, we can obtain various molecular properties such as the molecular geometry, relative stabilities, vibrational spectra, dipole moments, reactivity, and atomic charges. This method requires expensive computer time and resources. Thus, quantum chemistry methods are limited to just small molecular system.

Large molecular system beyond the reach of QM can be simulated by classical statistical mechanics methods, such as Monte Carlo (MC) or molecular dynamics (MD). These methods are introduced to obtain properties and processes at the molecular scale. Parameters (so called, force fields) are built up to describe bonded/nonbonded atomic interactions. In principle, QM can be used to determine the interactions in small model species. These quantitative results of QM can be utilized as starting points for adjustable parameters for potential functions that constitute an empirical force field. In fact, there are many other methods for choosing the nonbonded parameters for statistical mechanical simulations. In this thesis, a novel approach is introduced by using the Kirkwood-Buff (KB) theory[3] for developing force fields in liquid mixtures, which are our systems of interest. This is because information provided by the KB theory can predict good solution activities and solution structures of liquid mixtures at the various mixture compositions.

Complex fluids or macromolecules such as proteins are modeled by a set of potential energy functions of bonded and nonbonded interactions. The macromolecular model can be simulated in vacuum or in the presence of solvent (typically water or small organic molecule mixtures). The MD method is a deterministic procedure. MD starts from an arbitrary configuration and an initial set of particle velocities, and Newton's equations of motion of the system are integrated numerically as a function of time. (This thesis treats the details of MD techniques in the following section). On the other hand, MC is a stochastic method. The random MC trial moves are accepted with a probability. Depending on the interaction potentials and simulated systems, MD or MC are selected as a simulation method.

In this thesis we are interested in the atomic-scale description of liquids at a degree of resolution where an atom is the smallest elementary entity. The computational chemistry method needed in such a description relies on concepts of classical statistical mechanics. For studying processes at an all-atom scale in condensed phases, computers are used to solve the complex integrals that are used in statistical mechanics to define relations between the fundamental molecular interactions and experimental observables. This approach supposes that the Hamiltonian, or total energy of the system, is known.

This chapter is outlined as follows. First molecular dynamics and Monte Carlo methods are introduced. Then the functional forms of force fields and typical force field parameterizations are addressed. The next section is focused on molecular dynamics techniques, because molecular dynamics were used in all my simulations. Following this, free energy calculations and chemical potentials are presented for analyzing solvation effects, energetic contributions, and aggregations in liquid mixtures. In particular, the Kirkwood-Buff theory of solution, used in chapter 3 and chapter 4 of this thesis, is addressed in detail.

## 2.2 Statistical mechanics

Molecular dynamics and Monte Carlo simulations generate information at the microscopic level, including atom positions and velocities. The conversion of this microscopic information to macroscopic observables such as pressure, energy, heat capacities, etc., requires statistical mechanics. Statistical mechanics calculates average properties starting from a proper atomistic model and then provides the link between these quantities and the atomistic description on a microscopic level. Thus, statistical mechanics methods relate thermodynamic and bulk properties to the distribution and motion of the atoms and molecules of an *N*-body system. In the first place, molecular dynamics and Monte Carlo methods are briefly introduced. Afterward force fields and techniques are following.

### 2.2.1 Molecular dynamics

The Hamiltonian of a classical system of *N* point masses $m_i$ depends on the positions $\{\mathbf{r}_i\}$ of all point masses *i=1,2,…,N* and their momenta $\{\mathbf{p}_i\}$. The classical Hamiltonian *H* is defined as the sum of the kinetic and potential energy.

$$H(\{\mathbf{p}_i,\mathbf{r}_i\}) = \sum_{i=1}^{N} \frac{\mathbf{p}_i^2}{2m_i} + U(\{\mathbf{r}_i\}) \tag{2.1}$$

The kinetic energy contribution to the Hamiltonian in the first term on the right hand side of eq. 2.1 uses particle masses, $m_i$ and velocities, while the potential energy part depends on the set of point mass positions $\{\mathbf{r}_i\}$.

Molecular dynamics is the time dependent simulation of molecules described by Newton's equations of motions. Newton's equation of motion ($\mathbf{f}_i = m_i\mathbf{a}_i$) can be used to generate a dynamical trajectory of a system of interacting particles. The negative gradient of the potential energy function $U(\{\mathbf{r}_i\})$ with respect to particle coordinate $\mathbf{r}_i$ equals the force $\mathbf{f}_i$ acting on a particle *i*,

$$\mathbf{f}_i = -\frac{\partial U(\{\mathbf{r}_i\})}{\partial \mathbf{r}_i} \tag{2.2}$$

Newton's equation of motion can be numerically integrated on a computer to yield particle velocities and positions as a function of time. The method is deterministic; once the positions and velocities of each atom are known at time $t$=0, the state of the system can be predicted at any time in the future. Numerical integration methods are addressed in subchapter 2.2.4.

### 2.2.2 Monte Carlo

Monte Carlo methods are commonly used in statistical mechanics. In particular, the Metropolis Monte Carlo technique[47] has been used extensively in liquid simulations. The Monte Carlo method is a stochastic process, *i.e.,* atoms are moved randomly during simulation. The simulation starts from an arbitrary configuration, which is changed at each step with the probability of Boltzmann distribution, $\exp\left(-\Delta U / k_B T\right)$.

This method obeys the detailed balance condition in the equilibrium state such as,

$$W_{i\to j}\exp(-U_i / k_B T) = W_{j\to i}\exp(-U_j / k_B T) \tag{2.3}$$

where $W_{i\to j}$ is a probability to go from the state $i$ to the state $j$. The state $i$ and $j$ differ by the position of one particle, that is moved at a MC step. The probability to accept the move from state $i$ to state $j$, equal to $\max\left(1, \exp\left(-\left(U_j - U_i\right) / k_B T\right)\right)$ by the Metropolis criteria. Likewise the $W_{i\to j}$, the reverse process probability $W_{j\to i}$ follows the $\max\left(1, \exp\left(-\left(U_i - U_j\right) / k_B T\right)\right)$. Thus the contributions are the same in both direct and reverse probabilities such as,

$$\frac{\max\left(1, \exp\left(-\left(U_j - U_i\right) / k_B T\right)\right)}{\max\left(1, \exp\left(-\left(U_i - U_j\right) / k_B T\right)\right)} = \exp\left(-\left(U_j - U_i\right) / k_B T\right) \tag{2.4}$$

The condition of detailed balance given by eq. 2.3 implies that the average number of moves state *i*→*j* is the same as the average number of inverse moves state *j*→*i*. If the system in equilibrium is submitted to moves that obey the detailed balance condition, there will be no change in the probability of any conformation and the system will remain in equilibrium.

Over the process of several million attempted steps, energetically accessible configurations of the system are explored. Usually the step size is adjusted so that approximately 50% of the attempted steps are accepted. For example, if the size of each step is very large, the change in energy is too large, thus this attempt will be rejected. If, on the other hand, the step size is too small, the step size restricts the sampling of configuration space, leading to slow convergence of calculated properties. Thus, for large biomolecules with many internal degrees of freedom, MC is generally less efficient than MD for the calculation of thermodynamic properties. Moreover MC does not provide direct dynamic information about the system, *i.e.,* in MC the microscopic states are created randomly, there is no intrinsic time scale.

### 2.2.3 Force fields

In molecular systems the potential energy function $U\left(\left\{\mathbf{r}_i\right\}\right)$ should describe all interactions between atoms that are covalently bonded as well as non-bonded interactions between atoms and molecules in the condensed phase. Because in our classical description we do not include electrons (that describe the chemistry of the system) we make the approximation that atoms, molecules and their mutual interactions can be described by analytical potential energy functions in which the parameters model the chemistry. The set of potential energy functions and their parameters is usually referred to as a force field. Some of the force field parameters can be obtained from quantum-chemical calculations (*e.g.,* those that describe torsional transitions in molecules, vibrations of covalent bonds and bond angles) while others are often *parameterized* (*i.e.,* fitted) to reproduce solid or liquid state properties of the system.

In the simplest form a force field is written as a sum of bonded and non-bonded interaction terms

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdw} + U_{Coulomb} \tag{2.4}$$

The total potential is comprised of bonded terms ($U_{bond}$, $U_{angle}$, $U_{dihedral}$) and non-bonded terms ($U_{vdw}, U_{Coulomb}$). The bonded potentials are shown in detail in figure 2.1. The bond potential describes stretching of covalent bonds between atoms *i* and *j* and can be described with an harmonic potential. The bond angle vibration between a triplet of atoms *i-j-k* is described with an angle potential which is often chosen harmonic as well. For the dihedral potential, we can consider two types. Proper dihedral angles are defined according to $\phi$, the angle between the *ijk* and the *jkl* planes. Improper dihedrals are meant to keep planar groups planar (*e.g.,* aromatic rings) or to prevent molecules from flipping over to their mirror image.

The nonbonded potential is mainly divided in dispersion/repulsion interaction and electrostatic interaction. The typical features of intermolecular interaction are an attractive tail at large separations due to the Van der Waals interaction or Coulomb attractive interaction of charged species, while there is a steep repulsion at short separations caused by Pauli repulsion associated with the overlap of electron clouds belonging to non-bonded atoms/molecules. For the nonbonded Van der Waals interaction potential, a Lennard-Jones 12-6 pair potential is often used:

$$U^{LJ}(r) = 4\varepsilon\left((\sigma / r)^{12} - (\sigma / r)^{6}\right) \tag{2.5}$$

where *r* is the interatomic pair separation, $\varepsilon$ is the energy constant (or well depth) and $\sigma$ is Lennard-Jones size parameter.

It can provide a reasonable description of the properties of liquids, via computer simulation, if the parameters $\varepsilon$ and $\sigma$ are chosen appropriately.

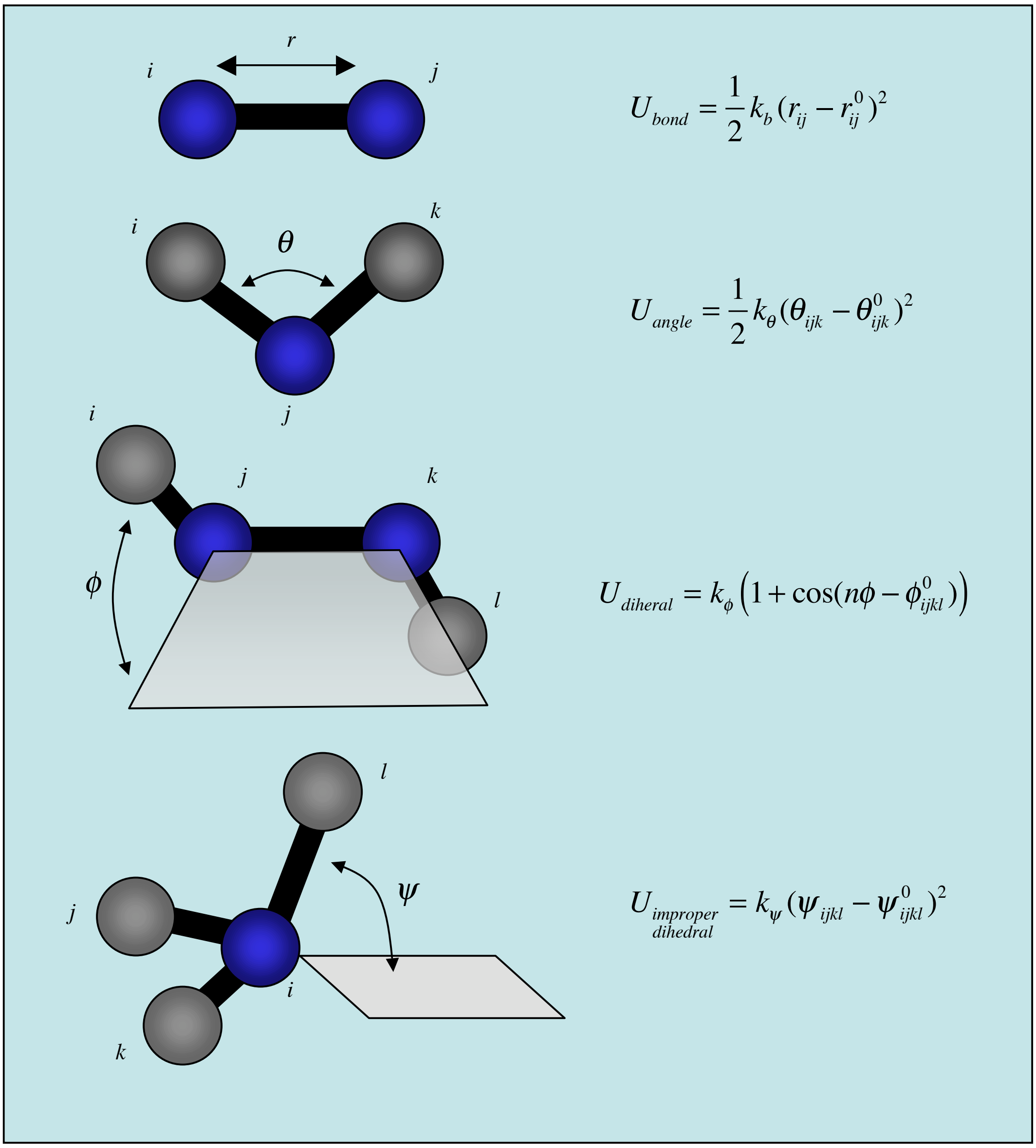


*Figure 2.1 The bond-, angle-, dihedral- and improper dihedral bonded potentials among atoms i, j, k, and l.*

Since at this point we have assumed pairwise additivity of the nonbonded interaction potentials, it should be kept in mind that any nonpairwise (multibody) effect will effectively enter into the values of $\varepsilon$ and $\sigma$ in the parameterization against liquid state properties.

The long-range interaction between charged or polar molecules is usually described by the Coulombic potential

$$U\ (r_{ij}) = \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.6}$$

where $q_i$, $q_j$ are the partial electric charges on interaction sites $i$ and $j$ and $\varepsilon_0$ is the permittivity of free space. Partial electronic charges can be obtained from quantum-chemical calculations (*e.g.,* by using them as parameters that reproduce the electrostatic potential on the molecular surface, which is usually chosen just beyond the Van der Waals radius of the atoms where it is most important to model the intermolecular interactions correctly). In many cases, however, they are used as empirical parameters optimized to describe liquid state properties.

Electrostatic interactions are long ranged and thus exceed half the box length for a typical simulation of ~ 500 molecules.[48] The obvious solution to this problem is to increase the size of the simulation box to reduce the artifact effect by screening neighbors. However, the simulation in large simulation box is expensive. Thus, methods such an Ewald summation and a reaction field provide appropriate tricks to handle long-range interaction. We introduce Ewald summation, particle mesh Ewald and reaction field for long-range forces in detail in section 2.2.6.

So far, the simplest forms of force fields for bonded and nonbonded interactions were addressed. Typically, quantum mechanics and experimental data are used to parameterize these force fields. Accurate description of interaction potential determines the quality of modeling. Validation of results obtained from simulation is performed by comparing with observable values in experiment. To ensure accurate predictions from a given force field, the fitting process is repeated until the deviations between computed and observed values are sufficiently small.

A main physical quantity in determining the structure of molecular systems is the energy. QM calculations provide energies, first and second derivatives of the energy with respect to atomic coordinates, and dipole moments for the models. Many other properties can be obtained as derivatives of the energy with respect to some other parameters such as atomic coordinates or electronic. The dynamics of molecular motion are also determined by these energies.

The use of quantum mechanical calculations, in particular ab initio calculations, to derive force fields has proven successful in a number of applications.[49-51] QM derived force fields have several advantages. It is possible to derive the properties of molecular systems without any input from experiments requiring difficult measurements. This interaction potential has a well-defined physical meaning for two-body or maybe three-body molecular interactions.

The force fields for nonbonded interactions as mentioned above, consist of a Coulomb term and a Van der Waals term. Although it is theoretically possible to obtain charges for the Coulomb term in a force field by fitting to dipole and higher order moments, in practice, it is very difficult to get physically reasonable charges. Therefore, charges and Van der Waals parameters are usually obtained separately. The Van der Waals parameters have usually been obtained by fitting to crystal structure data and to heats of sublimation of molecular crystals.[52-57] It is also possible to obtain parameters from quantum mechanical calculations but it is computationally expensive owing to the large basis sets and high degree of electron correlation required for their description. Because Van der Waals interactions include dispersion, it would be necessary to use a high-level ab initio methods such as a high-order perturbation theory or coupled cluster approaches.

The interaction parameters are capable to describe multi-body effects. Unfortunately, QM derived force fields consist of two-body energies and their physical descriptions. For example, tertiary butanol (it is dealt in chapter 3 and 4) has different dipole moments in the gas (1.66 Debye) and liquid state (2.52 Debye). As an another example, $H_2O$ shows big different dielectric permittivities values in liquid (~78 at 25°C) and gas (~1 beyond 100°C) states. To obtain meaningful transferable parameters, it is necessary to describe interactions including multi-body effects in order to be closed to experimental data.

In practice, empirical and QM derived parameters are often combined. Generally empirical force fields can be obtained from experimental data such as structure parameters (x-ray crystallography, NMR), dielectric properties (static and dynamic dielectric constants), elastic constant, vibrational properties (modes of vibration, phonons), heat capacities, and densities.

Force field parameters derived for the models can in many cases be transferred to the bulk state without any loss of generality. Force field transferability is important in practical simulations. When parameters have been carefully refined to reproduce observation for only a model system, one can make a question how they will behave in new chemical environments. Transferability is also a considerable factor to develop the parameters. Parameters for interaction potentials should be selected in order to be modeling according to transferability and correct descriptions of properties of the systems.

### 2.2.4 Integration of equation of motions

A good MD program requires a good algorithm to integrate Newton's equations of motion. In this sense, the choice of algorithm is important. A finite difference method is used to generate MD trajectories with continuous potential models, which we will assume to be pairwise additive. The essential idea is that the integration is broken down into many small stages, each separated by a fixed time $\Delta t$. The total force on each particle in the configuration at a time $t$ is calculated as the vector sum of its interactions with other particles. From the force we can determine the accelerations of the particles which are then combined with the positions and velocities at a time $t$ to calculate the positions and velocities at a time $t+\Delta t$. First, we introduce Verlet's algorithm[58] which is one of the simplest. The Verlet algorithm uses the positions and accelerations at time t, and the positions from the previous step, $r(t-\Delta t)$ to calculate the new positions $r(t+\Delta t)$ at $t+\Delta t$. The velocity does not appear explicitly in the Verlet integration. The Verlet algorithm has some disadvantages for the implicit velocity term. Indeed the velocity is not available until the positions have been computed at the next step. In addition, the new position can be obtained from the current positions and the position from the previous time step. At *t=0* there is no previous step for starting, starting velocities can be generated by taking random velocities from a Boltzmann distribution for a given temperature.

As an improvement to the Verlet algorithm, the leap-frog algorithm[59] has been developed. It's described by the following relationship,

$$r(t+\Delta t) = r(t) + \Delta t v(t+\frac{1}{2}\Delta t) \tag{2.7}$$

$$v(t+\frac{1}{2}\Delta t) = v(t-\frac{1}{2}\Delta t) + \Delta t a(t) \tag{2.8}$$

The leap-frog algorithm was used in our all MD simulations. The velocity $v(t+\Delta t/2)$ is first calculated, from the velocities at time $t-\Delta t/2$ and the accelerations at time *t*. The positions $r(t+\Delta t)$ are then deduced from the velocities just calculated together with the positions at time $r(t)$. The velocities thus 'leap-frog' over the positions to give their values at $t+\Delta t/2$. The leap-frog has an advantage over the Verlet algorithm as it explicitly includes the velocity. However, it has a disadvantage that the positions and velocities are not synchronized. This means that it is not possible to calculate the kinetic energy contribution to the total energy at the same time as the positions are defined.

The velocity Verlet method[60] has been developed giving positions, velocities and accelerations at the same time.

$$r(t+\Delta t) = r(t) + \Delta t v(t) + \frac{1}{2}\Delta t^2 a(t) \tag{2.9}$$

$$v(t+\Delta t) = v(t) + \frac{1}{2}\Delta t\left[a(t) + a(t+\Delta t)\right] \tag{2.10}$$

The velocity Verlet method has solved the disadvantages in the Verlet algorithm and the leap-frog method. The real advantage of the velocity Verlet algorithm is that it requires less computer memory, because only one set of positions, forces and velocities need to be carried at any one time. Also Beenan's algorithm[61] and Predictor-corrector[62] integration methods have been used for integration methods.

## 2.2.5 Periodic boundary conditions and cutoff distances

Simulations are performed on a finite number of particles in a small volume, but should still represent the properties of infinitely large systems. We therefore need a trick to reduce boundary effects, because many particles will be at the surface of the finite volume. The classical way to minimize boundary effects in a finite system is to use periodic boundary conditions.

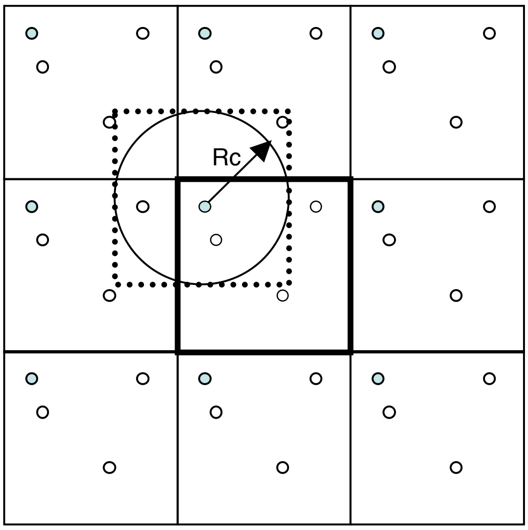


*Figure 2.2 Periodic boundary conditions with a cubic box in two dimensions. Any particle that leaves the central box re-enters at the opposite side. Hence, attention need only be fixed on the central box. In calculating the force on an atom, the minimum image of all neighboring atoms is taken into account within an interaction cutoff radius Rc. The box size should be at least two times larger than correlation distance (g(r)=1) to avoid that the central atom interacts with another atom and its image at the same time.*

The atoms of the system that is to be simulated are put into a cubic, or more general into any periodically space filling box, which is surrounded by identical translated images of itself (figure 2.2).

If a crystal is simulated, boundary condition is not a big artifact for its restricted periodic motion. For a molecule in solution, the periodicity is an artifact of the computation, so the effects of periodicity on the forces on the atoms should not be significant. When calculating the forces of the grey atom in the central box in figure 2.2, all interactions with atoms in the central box or their images in the surrounding boxes that lie within the spherical cut-off radius *Rc* are taken into account. The cutoff is chosen such that a particle in the primary box does not see its image in the surrounding boxes. When an atom leaves the central box at one side, it enters it with identical velocity at the opposite side at the translated image position. There are several possible shapes for unit cells such as cubic, rhombic dodecahedron, truncated octahedron. Cubic periodic conditions were used in our all simulation. (figure 2.2)

Under the periodic boundary conditions, the minimum central box image requires a total of $N(N-1)/2$ pairwise interaction calculations at every MD step. In order to reduce the computation time to calculate interactions, spherical cutoffs are applied by truncating the nonbonded interactions for each atom at large enough distances. The number of pairwise interactions for any atom is proportional to $r^3$ where $r$ is the spherical cutoff radius. Generally the cutoff distances in 7.5-15 Å ranges have been used. In any case, upper limit cutoff distance should be smaller than half box length. Alternative truncation schemes use different cutoffs for different parts of the energy evaluation. Twin-range cutoffs, which were used in our all simulations, employ two types of truncation, an interior and exterior cutoff. The interactions of all atoms in the interior cutoff range are calculated at every step of the simulation, while all others in the region between the interior and exterior ranges are recomputed only when the nonbonded list is updated. Because all interaction calculation at each step is time-consuming, thus it is updated every few steps of the simulation. At the cutoff distance, calculations for force and energy during simulation are required to truncate smoothly not to make any artifact.

## 2.2.6 Electrostatic interactions

In this section, Ewald method, particle mesh Ewald, and reaction field are introduced for describing electrostatic interactions. For accurate calculation of dielectric properties, the long-range forces should be dealt with properly depending on the simulated system environment.

### 2.2.6.1 Ewald summation

Ewald summation[63] is one of the most accurate methods for describing electrostatic interactions when periodic boundary conditions are used.[64] In fact, the electrostatic interactions are calculated between all atoms located in the central box and between all atoms of the central box with their images in the neighboring boxes. Ewald summation was first introduced as a method to calculate long-range interactions of the periodic images in crystals. The electrostatic interaction energy in a periodic system is given by[65]

$$U = \frac{1}{2}\sum_{|\boldsymbol{n}|=0}^{\infty}{}' \left[ \sum_{i=1}^{N}\sum_{j=1}^{N} \frac{q_i q_j}{|\boldsymbol{r}_{ij} + \boldsymbol{n}|} \right] \tag{2.11}$$

where $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, The sum over $\boldsymbol{n}$ is the sum over all simple cubic lattice points, $\boldsymbol{n} = (n_x L,\ n_y L,\ n_z L)$, where $L$ is the length of a cubic box. The lattice vector $\boldsymbol{n}$ reflects the shape of the box and $q_i$ and $q_j$ are net atomic charges centered on atoms $i$ and $j$. The prime indicates that we omit $i=j$ for $|\boldsymbol{n}| = 0$. For simplicity of notation, all factors of $4\pi\varepsilon_0$ are omitted. This sum is conditionally convergent, but very slow and time consuming. The unit cells are added; the first term has $|\boldsymbol{n}| = 0$, *i.e.,* $\boldsymbol{n} = (0,0,0)$; second term, $|\boldsymbol{n}| = L$, *i.e.,* $\boldsymbol{n} = (\pm L,0,0),(0,\pm L,0),(0,0,\pm L)$; *etc.* Further terms were added to build up infinite system.

The idea for slow and time consuming summation is to convert the single slowly-converging sum eq. 2.11 into two quickly-converging terms and a constant term[48, 65] : *i.e.,* $U = U_{direct} + U_{reciprocal} + U_0$. Unfortunately, the computational cost of the reciprocal part of the sum increases as $N^2$ and it is not realistic for use in large systems. Furthermore, Ewald summation is appropriate method for crystal simulation than for solution simulation caused by the implied periodicity. Thus, we didn't use Ewald summation in our simulations.

### 2.2.6.2 Particle mesh Ewald

In order to overcome intensive calculation of Ewald summation, Darden *et al.* proposed Particle mesh Ewald (PME) method for Ewald sums in large systems.[66, 67] PME divides the potential energy into Ewald's standard direct sum and reciprocal sums.

In fact, the PME method improves the performance of the reciprocal sum, *i.e.,* the reciprocal sum is evaluated using fast Fourier transforms with convolutions on a grid where charges are interpolated to the grid point, instead of directly summing wave vectors. As a result, computational cost of PME, which is efficient and accurate for large molecular systems, reduces to an *NlogN* scale. However, Ewald method is better in the very small system because it avoids the overhead in setting up grids and transforms.

### 2.2.6.3 Coulombic interaction with reaction field

The reaction field method assumes that dielectric permittivity is constant beyond the cutoff distance $R_C$. Therefore, the reaction field method considers two contributions. The first is a short-range contribution from molecules located within a cutoff distance and the second comes from molecules outside a sphere, $R_{out}$ which are considered to form a dielectric continuum ($\varepsilon_{RF}$) producing a reaction field.[68] The magnitude of the reaction field $(A_i)$ acting on molecule *i* is proportional to the dipole moment$(\boldsymbol{\mu})$ of the cutoff sphere surrounding *i*,

$$A_i = \frac{2(\varepsilon_{RF}-1)}{2\varepsilon_{RF}+1}\frac{1}{R_C^3}\sum_{j\in R_{out}}\boldsymbol{\mu}_j \tag{2.12}$$

where the summation extends over the molecules in the cutoff sphere, including *i*, and $R_C$ is the radius of the cutoff sphere. The energy from the reaction field is denoted by $-\frac{1}{2}\boldsymbol{\mu}_i \cdot A_i$.

Coulombic interaction (eq. 2.6) can be modified adopting the reaction field. *i.e.,* treating the dielectric continuum environment beyond the cutoff distance $R_C$ with reaction field dielectric constant of $\varepsilon_{RF}$.

In fact, difficulty of a potential calculation with reaction field is required to know the external dielectric constant $\varepsilon_{RF}$. Fortunately, the choice of $\varepsilon_{RF}$ is insensitive to determine the thermodynamic properties of a dipolar fluid.[48] In our simulations, the particle mesh Ewald and reaction field are used for the Coulomb interaction.

### 2.2.7 Temperature coupling

During an MD simulation, Berendsen[69] or Nosé-Hoover[70, 71] methods are commonly used to keep the temperature constant.

First of all, the Berendsen method goes toward the desired temperature $T_0$ by scaling the velocities for thermal drift at each step. The Berendsen algorithm has weak coupling to an external bath to minimize the disturbance of the system, *i.e.,* as possible to minimize influence on the trajectories.

$$m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} = \boldsymbol{F}_i + m_i \gamma \left[ \frac{T_0}{T(t)} - 1 \right] \frac{d\boldsymbol{r}_i(t)}{dt} \tag{2.13}$$

where $\boldsymbol{F}_i$ is the force and the friction constant $\gamma$ determines the strength of coupling to bath with the time dependent temperature, *T(t)* and the desired temperature $T_0$ . Thus, a modified equation represents a proportional scaling of the velocities per time step in the algorithm from $v$ to $\lambda v$ with,

$$\lambda = \left[ 1 + \frac{\Delta t}{\tau_T} \left( \frac{T_0}{T} - 1 \right) \right]^{1/2} \tag{2.14}$$

where $\tau_T$ is a relaxation time. *i.e.,* a long temperature relaxation time couples the system weakly to the heat bath. This approach effectively scales the atomic velocities at each time step to maintain a constant temperature during the simulation.

The Nosé-Hoover thermostat introduces a thermal reservoir and a friction term in the equations of motion. The Nosé-Hoover method is denoted by,

$$\begin{aligned} m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} &= \boldsymbol{F}_i - m_i \xi \frac{d\boldsymbol{r}_i(t)}{dt} \\ \frac{d\xi}{dt} &= \frac{1}{Q}(T - T_0) \end{aligned} \tag{2.15}$$

where $\xi$ is the friction parameter (or 'heat bath' variable), $Q$ is called as the mass parameter of reservoir which determines strength of the coupling in combination with the desired temperature ($T_0$), and others are same descriptions in Berendsen method.

The friction force is proportional to the product of each particle's velocity and a friction parameter. This friction parameter is a fully dynamic quantity with its own equation of motion. The time derivative is calculated from the difference between the current kinetic energy and the desired temperature. The Nosé-Hoover thermostat method involves both position and momentum phase space variables. It describes an additional dynamic degree of freedom which corresponds to a heat bath. However the Nosé-Hoover thermostat produces an oscillating relaxation. The real time it takes to relax with the Nosé-Hoover thermostat takes 4-5 times longer than the relaxation time used with Berendsen weak coupling. The Nosé-Hoover method was used in all our simulations.

## 2.2.8 Pressure coupling

The system can be coupled to a 'pressure bath', similar to coupling to a 'heat bath'. To simulate at constant pressure in an MD simulation, the volume is considered as a dynamical variable that changes during the simulation. Here, the Berendsen algorithm[69] and Parrinello-Rahman[72],[73] method are addressed as pressure coupling methods.

The Berendsen algorithm for coupling to a constant pressure bath can be performed according to the same principle for coupling to a constant temperature. This rescales the coordinates and box vectors with a matrix $\mu$, which has the change of a first-order kinetic relaxation of the pressure towards a desired pressure. An extra term is added to the equations of motion. Pressure scaling factor becomes such as,

$$\mu = \left\{ 1 + \frac{\Delta t}{\tau_P} \beta \big( P(t) - P_0 \big) \right\}^{1/3} \tag{2.16}$$

where $\beta$ is isothermal compressibility of the system, $\tau_P$ is a relaxation time, $P(t)$ is time dependent pressure and $P_0$ is the desired pressure. This is similar to the scaling pressure in constant pressure algorithm of Andersen.[74]

The fluctuations in volume are important to calculate thermodynamic properties. Therefore the exact ensemble should be defined, which is not possible with the Berendsen algorithm. To reduce this problem, the Parrinello-Rahman barostat has been developed, similar to the Nosé-Hoover temperature coupling. This allows both volume and shape of the periodic cell to respond with the pressure tensor. The volume, $V$ of the periodic unit cell, or all three vectors, $\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\}$ of the unit-cell can be added as an extra generalized coordinates. The position $\boldsymbol{r}_i$ of a particle $i$ can be written in terms of $3\times 3$ matrix, $\boldsymbol{h}$ to form $\{\boldsymbol{a},\boldsymbol{b},\boldsymbol{c}\}$ and of a column vector $\boldsymbol{s}_i$ with components $\xi_i,\eta_i,\varsigma_i$,

$$\boldsymbol{r}_i = \boldsymbol{h}\boldsymbol{s}_i = \xi_i\boldsymbol{a} + \eta_i\boldsymbol{b} + \varsigma_i\boldsymbol{c} \tag{2.17}$$

In the Parrinello-Rahman method the dynamics of the unit cell matrix are governed by the equation,

$$\boldsymbol{W}\ddot{\boldsymbol{h}} = (\pi - P_0)\boldsymbol{\sigma} \tag{2.18}$$

where $\boldsymbol{W}$ is the box 'mass' which is set to the total mass of the system in a box, $\boldsymbol{\sigma} = V(\boldsymbol{h}^{-1})^T$ which gives information of size and orientation of the MD cell, the superscript, $T$ stand for transpose, $P_0$ is the desired reference pressure, and $\pi$ is the instantaneous pressure tensor. In scaled variables, pressure tensor $\pi$ is given by

$$\pi_{\alpha\beta} = \frac{1}{V}\left(\sum_i m_i(\boldsymbol{h}\dot{\boldsymbol{s}}_i)_\alpha^T(\boldsymbol{h}\dot{\boldsymbol{s}}_i)_\beta + \sum_i\sum_{j>i}(\boldsymbol{h}\dot{\boldsymbol{s}}_{ij})_\alpha(\boldsymbol{f}_{ij})_\beta\right) \tag{2.19}$$

where $\boldsymbol{f}_{ij}$ is the force on $i$ due to $j$ in unscaled form. In here, $\pi$ is the deterministic part of pressure tensor.

The Parrinello-Rahman barostat is commonly combined with the Nosé-Hoover thermostat because its equations of motion for the particles are also changed like the Nosé-Hoover thermostat. Just as for the Nosé-Hoover thermostat, the Parrinello-Rahman time constant is not equivalent to the relaxation time used in the Berendsen pressure coupling, but it takes 4-5 times longer time. Also the Parrinello-Rahman coupling gives oscillations of the unit-cell vectors. Parrinello-Rahman pressure coupling was used in all our simulations.

## 2.3 Free energy calculations

Free energies play an important role in this thesis. Free energy changes relate to the stability of a chemical system, *i.e.,* the tendency of the system to react or change. In this thesis the free energies of solvation of solute molecules in aqueous solvents are computed. The free energy of solvation is defined as the thermodynamic work of reversibly introducing the interactions between solute and solvent. The lower the solvation free energy the more favourable the solvation process will be. Not only solvation or solubility of solutes, but also the tendency of solutes to associate in solution is determined by a free energy. In the case of nonpolar solutes in aqueous media we speak of hydrophobic interactions.

For small molecules such as noble gases (He, Ne, Ar, Kr, Xe) or methane, the test particle insertion (TPI) method is useful to calculate the free energy of solvation. However TPI converges very poorly in the case of larger molecules, because it is difficult to observe fluctuations in the pure solvent that create a cavity of the appropriate (large) size. In this case, the thermodynamic integration (TI) methods can be used to calculate a free energy. However, it is a more computationally expensive calculation. Additionally, the potential of mean force (PMF) method is useful to obtain a free energy of association along a coordinate. First, the partition function is addressed because it is an important key to calculate all macroscopic properties of the system.

### 2.3.1 Partition function

At the constant temperature, the probability $p_i$ that the system occupies the *i*th microstate with energy $E_i$, is given by the Boltzmann distribution.

$$p_i = \frac{\exp(-E_i / k_B T)}{Z} \tag{2.20}$$

where *Z* is the normalization constant so that the sum of the probabilities is one $\left( \sum_i p_i = 1 \right)$.

Thus, the normalization constant which is called the partition function, can be described such as,

$$Z = \sum_i \exp(-E_i / k_B T) \tag{2.21}$$

where the sum is over all the microstates of the system.

Briefly consider the classical canonical partition function *Q*. The partition function is defined as,

$$Q(T,V,N) = \frac{1}{N!h^{3N}} \int \cdots \int d\boldsymbol{p}^N d\boldsymbol{r}^N \exp\left[-\beta H(\boldsymbol{p}^N, \boldsymbol{r}^N)\right] \tag{2.22}$$

where integration is performed over all particle momenta ($\boldsymbol{p}^N$) and positions ($\boldsymbol{r}^N$) and $H\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right)$ is the *N*-particle Hamiltonian. The *h* is Planck's constant, and the factor $N!$ takes care of indistinguishability of the particles. The variable $\beta$ equals $1/k_B T$. Hamiltonian can be described by kinetic energy and potential energy terms.

$$H\left(\boldsymbol{p}^N, \boldsymbol{r}^N\right) = \sum_{i=1}^{N} \left(\boldsymbol{p}_i^2 / 2m\right) + U_N\left(\boldsymbol{r}^N\right) \tag{2.23}$$

where $\boldsymbol{p}_i$ is the momentum vector of the *i*th particle presumed to posses only translational degree of freedom, $m_i$ is the mass of each particle *i*. The $U_N\left(\boldsymbol{r}^N\right)$ is the total potential energy of the system at the specified configuration $\boldsymbol{r}^N$.
The integration over the momenta in the eq. 2.22 is

$$\begin{aligned} h^{-3N}\int_{-\infty}^{\infty} d\boldsymbol{p}^N \exp\left[-\beta\sum_{i=1}^{N}\left(\boldsymbol{p}_i^2/2m\right)\right] &= \left[h^{-1}\int_{-\infty}^{\infty} d\boldsymbol{p}^N \exp\left(-\beta\boldsymbol{p}^2/2m\right)\right]^{3N} \\ &= \left[h^{-1}\left(2m/\beta\right)^{1/2}\int_{-\infty}^{\infty}\exp(-x^2)dx\right]^{3N} \\ &= \left[\left(2\pi mkT\right)^{1/2}/h\right]^{3N} = \Lambda^{-3N} \end{aligned} \tag{2.24}$$

where $\Lambda = h/\left(2\pi mkT\right)^{1/2}$ is de Broglie wavelength and $\Lambda^{-3N}$ is the momentum partition function. Hydrogen is a light particle with a large uncertainty in its position ( $\Lambda_{Hydeogen}$ is $\approx 1$Å at the room temperature) and can not be properly treated in classical mechanics. Hence, hydrogen atoms in atomistic simulations should be either constrained (for example via the SHAKE algorithm[75]), or coarse grained, via a united atom approach.

The configurational partition function, $Z_N$, is described by $Z_N = \int d\boldsymbol{r}^N \exp\left(-\beta U_N(\boldsymbol{r}^N)\right)$, thus, the canonical partition function can be rewritten,

$$Q\left(T,V,N\right) = Z_N/\left(N!\Lambda^{3N}\right) \tag{2.25}$$

The free energy $A$ is given in terms of the partition function $Q$ by

$$A = -k_B T \ln Q \tag{2.26}$$

## 2.3.2 Widom test particle insertion

A simple method for measuring the solvation free energy of a solute in a pure fluid or in a mixture is the test particle insertion method proposed by Widom.[76, 77] The TPI method has been used extensively to investigate the solvation of small nonpolar molecules by measuring the energy of solute molecules randomly inserted into the solvent. To calculate the free energy change by an additional particle insertion, it is derived starting from a partition function. When an additional particle is inserted into the *N*-solvent system, the free energy change is described as,

$$\mu = \left(\frac{\partial A}{\partial N}\right)_{V,T} = A\left(N+1,V,T\right) - A(N,V,T) \tag{2.27}$$

Substituting eq. 2.25 and 2.26 into eq. 2.27,

$$\begin{aligned}\mu = \left(\frac{\partial A}{\partial N}\right)_{V,T} &= -k_B T \ln\left[\frac{Z_{N+1}}{(N+1)!\Lambda^{3(N+1)}}\right] + k_B T \ln\left[\frac{Z_N}{N!\Lambda^{3N}}\right] \\ &= -\,k_B T \ln\left(\frac{1}{(N+1)\Lambda^3}\frac{Z_{N+1}}{Z_N}\right) \\ &= k_B T \ln\left(\frac{(N+1)}{V}\Lambda^3\right) - k_B T \ln\left(\frac{1}{V}\frac{Z_{N+1}}{Z_N}\right) \\ &= k_B T \ln\left(\rho\Lambda^3\right) - k_B T \ln\left(\frac{1}{V}\frac{Z_{N+1}}{Z_N}\right)\end{aligned} \tag{2.28}$$

where the first term on the right side means that a particle with a kinetic energy wanders freely in entire volume and is indistinguishable with other particles. The second term indicates the free energy induced by insertion of an additional particle.

Energy induced by an additional particle insertion is described by

$$U_{N+1} = U_N + B_S \tag{2.29}$$

where the binding energy $B_S$ is defined as $B_S = U_{N+1} - U_N$ ; *i.e.,* the difference of *N+1* and *N* particle potential energies. The $B_S$ equals the interaction energy of the insertion particle (the solute) with all other molecule.

Therefore, the second term in eq. 2.28 is derived as,

$$\begin{aligned} -k_BT\ln\left(\frac{1}{V}\frac{Z_{N+1}}{Z_N}\right) &= -k_BT\ln\left(\frac{1}{V}\frac{\int d\boldsymbol{r}^{N+1}e^{-\beta U_{N+1}\left(\boldsymbol{r}^{N+1}\right)}}{\int d\boldsymbol{r}^{N}e^{-\beta U_{N}\left(\boldsymbol{r}^{N}\right)}}\right) \\ &= -k_BT\ln\left(\frac{1}{V}\frac{\int d\boldsymbol{r}^{N+1}\int d\boldsymbol{r}^{N}e^{-\beta U_{N}\left(\boldsymbol{r}^{N}\right)}e^{-\beta B_S}}{\int d\boldsymbol{r}^{N}e^{-\beta U_{N}\left(\boldsymbol{r}^{N}\right)}}\right) \\ &= -k_BT\ln\left(\frac{1}{V}\int d\boldsymbol{r}^{N+1}\left\langle e^{-\beta B_S}\right\rangle_N\right) \\ &= -k_BT\ln\left\langle e^{-\beta B_S}\right\rangle_N \end{aligned} \tag{2.30}$$

Therefore, the free energy change by the insertion of an additional particle,

$$\Delta A_S \equiv -k_BT\ln\left\langle e^{-\beta B_S}\right\rangle_{NVT} \tag{2.31}$$

where the ensemble average is obtained over pure solvent configurations at constant pressure $V$ and temperature $T$, subscript 's' refers to solvation. The method relies on the statistically accurate sampling of solvent configurations that permit the insertion of molecules with low values of $B_S$. For that reason it is more efficient for small-sized solutes.

## 2.3.3 Thermodynamic integration

In case of larger solutes (*e.g.,* >$CH_4$), alternative methods are needed in which the solute is gradually introduced into the system. For this purpose we have used the thermodynamic integration (TI) method. This approach performs calculation of a free energy difference through integration. The TI formula for the solvation free energy, $\Delta A_S$, is given by,[78, 79]

$$\Delta A_S = \int_0^1 \left\langle \frac{\partial U(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{2.32}$$

where $\lambda$ is the coupling parameter that couples the interaction between solute and solvent, $U(\lambda)$ is the energy describing the total potential interaction and $\langle\cdots\rangle_\lambda$ denotes an average in a canonical ensemble with Hamiltonian $H(\lambda)$. The path between initial and final states is specified using a coupling parameter $\lambda$, which is impacting simulation efficiency with adjusting a stepsize of $\lambda$.[80] The parameter $\lambda$ regulates the strength of $U$ such that $\lambda=0$ denotes the full solute-solvent interaction whereas $\lambda=1$ corresponds to the solute-solute interaction being completely switched off.

In practice eq 2.32 is calculated by running a series of simulations, each with a different $\lambda$-value, and numerically integrating over the free energy derivative $\left\langle \partial U / \partial\lambda \right\rangle_\lambda$. Figure 2.3 shows the free energy change using the TI method for dissolving one *tert*-butanol molecule in water with different choices for $\Delta\lambda$.

Free energy changes are -13.1 *kJ/mol* ($\Delta\lambda = 0.1$), -11.6 *kJ/mol* ($\Delta\lambda \le 0.05$). The difference of these two vales (2.5 *kJ/mol*) comes from extreme points which are not detected in the case of $\Delta\lambda = 0.1$. Sampling difficulties may arise near the endpoints ($\lambda = 0$ and 1).

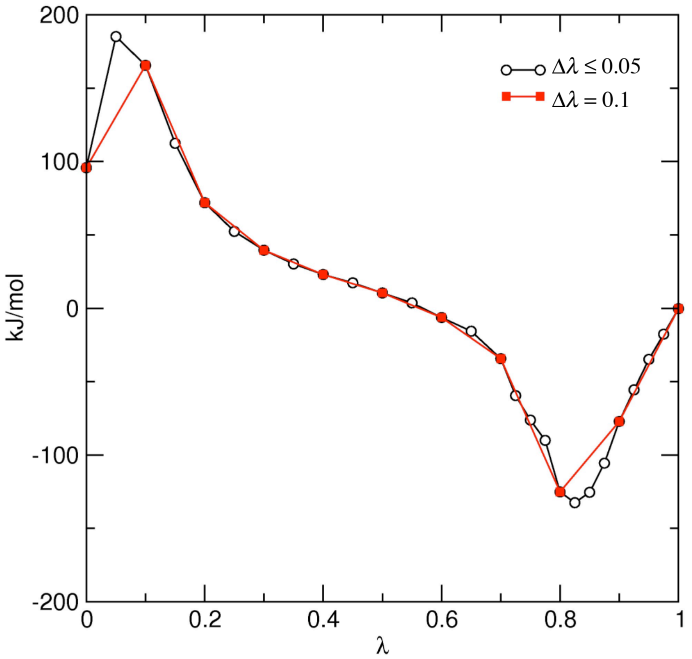


*Figure 2.3 Thermodynamic integration of one tertiary butanol (OPLS)*[81] *and 1000 water (SPC)*[82] *mixtures. Free energy changes are -13.1 kJ/mol (*$\Delta\lambda=0.1$*; red), -11.6 kJ/mol (*$\Delta\lambda \le 0.05$*; black).*

As it will be apparent from the figure, accurate calculation requires the sampling at more $\lambda$-values in regions between $\lambda = 0$ and $\lambda = 0.2$ and regions between $\lambda = 0.7$ and $\lambda = 0.9$. This problem can be overcome to some extent by using smaller $\lambda$ integration steps in the $\lambda$-ranges. Choosing an appropriate $\Delta\lambda$ is required not to miss the significant path. Normally, nonlinear potential functions of $\lambda$ are used.[79, 83] We note that, although we do not specify the $\lambda$-dependence of the Hamiltonian here, in practice this dependence can and must be chosen such that sharp extrema and singularities of $\left\langle \partial U / \partial \lambda \right\rangle_\lambda$ are avoided.

## 2.3.4 Potential of mean force

A potential of mean force (PMF)[84] is a free energy projected onto a coordinate. This may involve a reaction coordinate describing a conformational transition or, as used in this thesis, a simple distance between two solute molecules in a solvent. Then, we also speak of the free energy of association. The PMF involves association and desolvation energies when two molecules approach each other in solvent. There exist many methods to compute PMFs in molecular simulations. Here we discuss the constraint-bias method with force averaging. In this method a distance constraint is applied and the average constraint force – obtained by averaging over degrees of freedom of the solvent – is sampled at various values of the constraint distance $r$. Integrating the solvent-averaged constraint force from large ("infinite") distance to distance $r'$, defines the PMF $w(r)$

$$w(r) = \int_{\infty}^{r} \left\langle F_c(r') \right\rangle_{r'} dr' \tag{2.33}$$

where $F_c$ is the constraint force required to keep the solute pair fixed at distance $r$, and $\langle \cdots \rangle_{r'}$ is an average value over fluctuations of the force. Alternatively the PMF may be obtained from a simulation where the probability distribution $g(r)$ of the degree of freedom $r$ is accumulated. The PMF is then given by,

$$w(r) = -k_B T \ln g(r) \tag{2.34}$$

Because we deal with a solute pair, $g(r)$ is the solute-solute pair correlation function. $k_B$ is the Boltzmann constant and $T$ the temperature of the system.

In case not only the distance between solutes but also their mutual orientation is of interest, one or more angles may need to be included. For example, the interaction between two aromatic rings can be described with a distance $r$ and angle $\theta$ between the two rings. The pair correlation functions for this two-dimensional relation is described by

$$g(r,\theta) = \Delta n(r,\theta) / \Delta V \tag{2.35}$$

where $\Delta V = 2\pi r^2 \sin\theta \Delta r \Delta\theta$. The PMF is hence given by

$$w(r,\theta) = -k_B T \ln\left[ g(r,\theta) \right] \tag{2.36}$$

## 2.4 Chemical potential in molecular fluids

In chapter three of this thesis we will make use of the Kirkwood-Buff (KB) theory of solution in parameterizing an alcohol/water force field. KB theory provides relations between derivatives of the solution component chemical potentials with mixture composition and the distributions of solvent molecules at the molecular scale. To introduce this theory we start here with introducing the concept of an ideal solvent mixture. Then we define non-ideal mixtures and relations between thermodynamic non-idealities and molecular distribution functions through KB theory. For an extensive description we refer to the book of Ben-Naim.[85]

### 2.4.1 Ideal mixtures

A fundamental quantity in the description of liquid mixtures is the chemical potential. The chemical potential measures how much the free energy of a system changes if one mole of particles is added to the system while keeping the number of other particle species (and the temperature $T$ and the pressure $P$) constant. Here we consider binary liquid mixtures composed of the components $A$ and $B$. The chemical potential of component $A$ is defined as

$$\mu_A = \left(\frac{\partial G}{\partial N_A}\right)_{T,P,N_B} = G(T,P,N_A+1,N_B) - G(T,P,N_{A,}N_B)$$

$$= k_B T \ln(\rho_A \Lambda_A^3) + W(A \mid A,B) \tag{2.37}$$

where $G$ is the free energy of the system, $N_A$ and $N_B$ are the number of $A$ and $B$ molecules in the system, $\rho_A$ is the number density of $A$ component, $\Lambda_A^3$ is the momentum partition function of an $A$ molecule, and $W(A \mid A,B)$ is the "coupling work" of $A$ against an environment of $A$ and $B$ molecules. Without loss of generality, we have assumed that $A$ and $B$ have no internal degrees of freedom. This coupling work $W(A \mid A,B)$ is equal to the free energy of solvation in eqs. 2.31 and 2.32.[85] The first term $k_B T \ln(\rho_A \Lambda_A^3)$ account for the fact that $A$ has a kinetic energy, can wander around in a volume $V$ and is indistinguishable from all other $A$-type molecules. The combination of indistinguishability (contribution $k_B T \ln N_A$ in the chemical potential) and volume accessible (contribution $-k_B T \ln V$ in the chemical potential) causes the appearance of the number density $\rho_A$ in $k_B T \ln(\rho_A \Lambda_A^3)$. If we now assume that the $A$ and $B$ molecules are "very similar"[85] in the sense that interactions of $A$ with $A$, $A$ with $B$, and $B$ with $B$ can be approximated equal, the coupling work $W(A \mid A,B)$ will not depend on the ratio of $A$ and $B$ in the mixture and $W(A \mid A,B)$ is a constant (depending, of course, on $P$ and $T$). For the sake of notation used later on, we will rename this constant and denote it $k_B T \ln \gamma_A(P,T)$ where $\gamma_A(P,T)$ is the mole fraction scale activity coefficient of $A$, which in this specific case depends on pressure and temperature but not on composition. Therefore, we can write

$$\begin{aligned} \mu_A(\rho_A,P,T) &= k_B T \ln\left(\rho x_A \Lambda_A^3\right) + k_B T \ln \gamma_A \\ &= k_B T \ln\left(\rho \Lambda_A^3 \gamma_A\right) + k_B T \ln x_A \\ &= \mu_A^*(P,T) + k_B T \ln x_A \end{aligned} \tag{2.39}$$

which is the chemical potential of $A$ in an ideal liquid mixture.[86] To arrive at this expression we used $\rho_A = \rho x_A$, with $\rho$ the total number (or molar) density of the liquid and $x_A$ the mole fraction of $A$ in the system. Since we assumed that $A$ and $B$ are "very similar" in the way they interact we also expect that $A$ and $B$ are distributed randomly in the mixture with no preference of $A$ to be solvated by $A$ (or $B$) or $B$ to be solvated by $B$ (or $A$).

Non-ideal solution behavior occurs if $A$ and $B$ are not "very similar" and $A$'s and $B$'s may prefer to interact with molecules of the same or the other kind. Then, $k_B T \ln \gamma_A$ will depend on the solution composition $x_A$ and the chemical potential of $A$ is

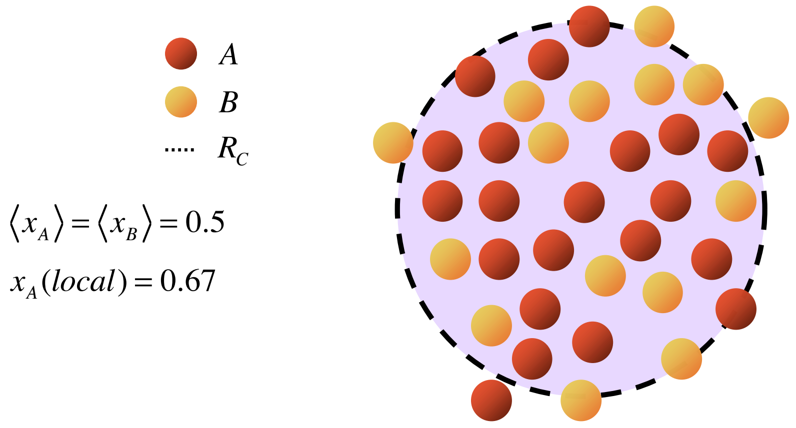
$$\begin{aligned}\mu_A\left(\rho_A,P,T\right) &= k_B T\ln\left(\rho\Lambda_A^3\right) + k_B T\ln x_A + k_B T\ln\gamma_A(P,T,x_A)\\ &= \mu_A^+\left(P,T\right) + k_B T\ln x_A + k_B T\ln\gamma_A(P,T,x_A)\\ &\equiv \mu_A^+\left(P,T\right) + k_B T\ln a_A\end{aligned} \tag{2.40}$$

where $a_A$ is the activity of component $A$. We suppose now that the activity coefficient $\gamma_A$ decreases if $x_A$ increases, *i.e.,* the coupling work of $A$ with the system (solvation free energy of $A$) decreases. It is not difficult to imagine that this now may happen because $A$-type molecules like being surrounded by $A$-type molecules. Therefore the more $A$-type molecules available the less the work needed to add $A$-type molecule. The theory in the next section describes how to obtain quantitative information on the molecular-scale distribution of the molecules based on chemical potential derivatives with the composition.

### 2.4.2 Kirkwood-Buff theory

In this section, we will describe deviation from ideal behavior in liquid mixtures using the Kirkwood-Buff (KB) theory of solution[3] applied to binary systems. In particular, we use KB theory in this thesis to explain the non-ideal behavior of liquid mixtures, which show excess aggregation between molecules of the same species and repulsion among molecules of dislike species.

The non-ideal behavior can be generalized with the number density in the configuration space of molecular pairs and fluctuations. First of all, the grand canonical ensemble ($\mu VT$) which contains $N_1, N_2, \cdots, N_C$ molecules with multi-component mixture is introduced. Molecules can move in and out in the grand canonical system. For example, the average numbers of $A$ and $B$ molecules in mixtures are same in the bulk $\left(\langle x_A \rangle = \langle x_B \rangle = 0.5\right)$. Nevertheless, local aggregation can be observed at the microscopic scale such as the schematic picture below. This picture shows the local aggregation of $A$ molecule (red ball; local mole faction of $A$ molecule, $x_A$ is 0.67) within pair correlation distance ($R_C$) with the open-system (dash line), despite the average mole fractions of $A$ and $B$ molecules are same in the bulk. We assume that the central solute $A$ sees the bulk composition $\left(\langle x_A \rangle = \langle x_B \rangle = 0.5\right)$ beyond the correlation distance and the pair correlation function $g(r > R_C)$ over the correlation distance is assumed to be unity. The correlation volume ($V_C$) equals the spherical volume with a diameter $R_C$.



Excess coordination of an “A” molecule (in the center of the dashed circle)

Therefore excess coordination number $N_{AB}$ is denoted by

$$N_{AB} = \int_{V_C} \rho_B g_{AB}^{\mu VT}(r)\, d\mathbf{r} - \rho_B V_C = \rho_B \int_{V_C} \left[ g_{AB}^{\mu VT}(r) - 1 \right] d\mathbf{r}$$

$$= \rho_B \int_0^{\infty} \left[ g_{AB}^{\mu VT}(r) - 1 \right] 4\pi r^2 dr \tag{2.40}$$

where $g_{AB}^{\mu VT}(r)$ is the pair correlation function between species *A* and *B* in grand canonical ensemble. The quantity $\rho_B g_{AB}(r) 4\pi r^2 dr$ measures the average number of *B* molecules in a spherical shell of width *dr* at a distance *r* from the center of a *A* molecule. On the other hand, the quantity $\rho_B 4\pi r^2 dr$ measures the average number of *B* molecules in the same spherical shell, which has been chosen at random. Therefore, eq. 2.40 defines the excess (or deficiency) to the average number of *B* molecules in the spherical shell at distance *r* from a *A* molecule relative to the number obtained by eliminating the condition that molecule *B* is located at the origin. A positive value of $N_{AB}$ indicates an excess of species *B* surrounding species *A* while a negative value of $N_{AB}$ presents depletion of species *B* in the vicinity of species *A*. This quantity gives information about the affinity of *B* molecules towards *A* molecules. The excess coordination numbers are simply expressed such as $N_{AB} = \rho_B G_{AB}$. Here, the $G_{AB}$ are called Kirkwood-Buff integrals.

Kirkwood-Buff integrals (KBIs) are defined as

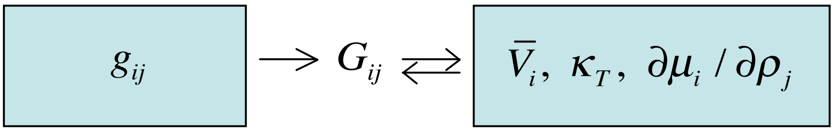
$$G_{AB} \equiv \int_0^{\infty} \left( g_{AB}^{\mu VT}(r) - 1 \right) 4\pi r^2 dr$$

$$= \int_0^{R_C} \left( g_{AB}^{\mu VT}(r) - 1 \right) 4\pi r^2 dr \tag{2.41}$$

where $G_{AB}$ (=$G_{BA}$) is the KBI for solvent pairs *A* and *B*, $g_{AB}^{\mu VT}$ is the corresponding pair correlation distribution function in the grand canonical ($\mu VT$) ensemble, and *r* is the distance between the centers of molecules *A* and *B*. The integration from zero to correlation distance, $R_C$, should be same with the integration from zero to infinity.

Ben-Naim suggested how to calculate the KBIs from measured thermodynamic properties.[9] He presented relationships between experimental data such as partial molar volumes, isothermal compressibilities, partial vapor pressures and the quantities $G_{ij}$.

Using thermodynamic quantities of mixtures such as partial molar volumes $\bar{V}_i$, isothermal compressibilities $\kappa_T$, and partial vapor pressures from experimental work, one can calculate $G_{ij}$ values.[85] By computing the $G_{ij}$ values from the pair correlation function $g_{ij}$ obtained by simulations of liquid mixtures, we can directly connect simulated structural data with experimental thermodynamic data. This idea is schematically shown in the figure below.



As a theory of solution, it is conveniently applied to the entire range of compositions. Thus, the KB theory has become a powerful tool in the study of particularly complex fluids such as water and aqueous solutions.

From now on, these useful relationships are derived in detail.[85] The grand canonical ensemble description is characterized by variables $\boldsymbol{\mu}, V, T$, where $\boldsymbol{\mu} = (\mu_1, \mu_2, \cdots, \mu_c)$ stands for the vector comprising the chemical potentials of all the *c* components of the system. For specified configurations, the indices $\alpha$ and $\beta$ were used to denote the species, *i.e.,* $\alpha$, $\beta$ $= 1, 2, \cdots, c$. In this case, normalization condition for singlet and the pair distribution functions are expressed in following as,

$$\int \rho_{\alpha}^{(1)}(\boldsymbol{r}) d\boldsymbol{r} = \langle N_\alpha \rangle \tag{2.42}$$

$$\begin{aligned} \iint \rho_{\alpha\beta}^{(2)}(\boldsymbol{r}, \boldsymbol{r}') d\boldsymbol{r} d\boldsymbol{r}' &= \langle N_\alpha (N_\alpha - 1) \rangle \quad \text{if } \alpha = \beta \\ &= \langle N_\alpha N_\beta \rangle \quad \text{if } \alpha \neq \beta \\ &= \langle N_\alpha N_\beta \rangle - \langle N_\alpha \rangle \delta_{\alpha\beta} \end{aligned} \tag{2.43}$$

where the symbol $\langle \cdots \rangle$ indicates an average in the grand canonical ensemble, $\rho$ is the number density, superscript (1) and (2) are denoted by a singlet and pair species, and the $\delta_{\alpha\beta}$ is the Kronecker delta function which is unity for $\alpha = \beta$ and zero for $\alpha \neq \beta$. $N_\alpha$ and $N_\beta$ are the number of molecules $\alpha$ and $\beta$ in the system.

The singlet and pair densities are given by,

$$\rho_\alpha^{(1)}(\boldsymbol{r}) = \rho_\alpha \tag{2.44}$$

$$\rho_{\alpha\beta}^{(2)}(\boldsymbol{r},\boldsymbol{r}') = \rho_\alpha \rho_\beta g_{\alpha\beta}^{(2)}(\boldsymbol{r},\boldsymbol{r}') \tag{2.45}$$

where $\rho_\alpha$ is the number density of species $\alpha$ $\left(\rho_\alpha = N_\alpha / V\right)$, $\rho_\beta$ is the number density of species $\beta$ ($\left(\rho_\beta = N_\beta / V\right)$, and $g_{\alpha\beta}^{(2)}(\boldsymbol{r},\boldsymbol{r}')$ is the pair correlation function between species $\alpha$ and $\beta$.

From eq. 2.44 and eq. 2.45, the next relation is derived.

$$\begin{aligned} \iint \left[\rho_{\alpha\beta}^{(2)}(\boldsymbol{r},\boldsymbol{r}') - \rho_\alpha^{(1)}(\boldsymbol{r}) \rho_\beta^{(1)}(\boldsymbol{r}')\right] d\boldsymbol{r} d\boldsymbol{r}' &= \rho_\alpha \rho_\beta \iint \left[g_{\alpha\beta}^{(2)}(\boldsymbol{r},\boldsymbol{r}') - 1\right] d\boldsymbol{r} d\boldsymbol{r}' \\ &= \left\langle N_\alpha N_\beta \right\rangle - \left\langle N_\alpha \right\rangle \left\langle N_\beta \right\rangle - \left\langle N_\alpha \right\rangle \delta_{\alpha\beta} \end{aligned} \tag{2.46}$$

The $G_{\alpha\beta}$ according to eq. 2.44 is expressed as the following formula,

$$G_{\alpha\beta} \equiv \int_0^\infty \left[g_{\alpha\beta}(r) - 1\right] 4\pi r^2 dr = V\left(\frac{\left\langle N_\alpha N_\beta \right\rangle - \left\langle N_\alpha \right\rangle \left\langle N_\beta \right\rangle}{\left\langle N_\alpha \right\rangle \left\langle N_\beta \right\rangle} - \frac{\delta_{\alpha\beta}}{\left\langle N_\beta \right\rangle}\right) \tag{2.47}$$

where $G_{\alpha\beta}$ is KBI to measure the preferential solvation and the right hand side term consists of the particle number fluctuations. The integral extends over all position of a pair of type $\alpha\beta$. This is a connection between the fluctuation in the number of various species and an integral involving only the spatial pair correlation functions for the corresponding pair of species $\alpha$ and $\beta$. This is an important point: the KBI, $G_{\alpha\beta}$, is the mediated relation between the radial distribution functions and the density fluctuations.

Meanwhile, the fluctuation averages are taken over all numbers of molecules $N_1,\cdots,N_C$ and over the accessible phase space of each set of molecules with the probability distribution function. In here, the partition function $\Xi$ is referred as the grand partition function or the open-system partition function.

$$\Xi = \sum_{\boldsymbol{N}} Q_{NVT} \exp\left(\beta \boldsymbol{\mu} \cdot \boldsymbol{N}\right) \tag{2.48}$$

where the summation in eq. 2.48 is over all possible values of ***N***. In eq. 2.48, $\boldsymbol{N} = N_1, N_2, \cdots, N_c$ is the vector representing the composition of the system, where $N_i$ is the number of molecules of species *i*. The corresponding vector $\boldsymbol{\mu} = \mu_{1,}\mu_{2,}\cdots\mu_c$ includes the chemical potential of each of the species. This is generalized with respect to all components for an open system. In eq. 2.48, $\boldsymbol{\mu} \cdot \boldsymbol{N} = \sum_{i=1}^{c} \mu_i N_i$ is the scalar product of the two vectors $\boldsymbol{\mu}$ and $\boldsymbol{N}$. Composition fluctuations in the grand canonical ensemble ($\mu VT$) are related to derivatives of the chemical potentials of the components.

$$\begin{aligned} \left\langle N_\alpha \right\rangle &= \frac{1}{\Xi} \sum_{\boldsymbol{N}} N_\alpha Q_{NVT} \exp\left(\beta \boldsymbol{\mu} \cdot \boldsymbol{N}\right) \\ &= k_B T \left( \frac{\partial \ln \Xi}{\partial \mu_\alpha} \right)_{VT\mu_{\alpha'}} \end{aligned} \tag{2.49}$$

where $\boldsymbol{\mu}_{\alpha'}$ stands for the set $\mu_1, \mu_2, \cdots, \mu_c$ excluding $\mu_\alpha$.

Differentiating eq. 2.49 with respect to $\mu_\beta$, we get

$$\begin{aligned} k_B T \left( \frac{\partial \left\langle N_\alpha \right\rangle}{\partial \mu_\beta} \right)_{VT\mu_{\beta'}} &= \frac{-1}{\Xi^2} \left( \sum_{\boldsymbol{N}} N_\beta Q_{NVT} \exp\left(\beta \boldsymbol{\mu} \cdot \boldsymbol{N}\right) \right) \cdot \left( \sum_{\boldsymbol{N}} N_\alpha Q_{NVT} \exp\left(\beta \boldsymbol{\mu} \cdot \boldsymbol{N}\right) \right) \\ &\quad + \frac{1}{\Xi} \sum_{\boldsymbol{N}} N_\alpha N_\beta Q_{NVT} \exp\left(\beta \boldsymbol{\mu} \cdot \boldsymbol{N}\right) \\ &= \left\langle N_\alpha N_\beta \right\rangle - \left\langle N_\alpha \right\rangle \left\langle N_\beta \right\rangle \end{aligned} \tag{2.50}$$

With respect to interchanging the $\alpha$ and $\beta$,

$$k_BT\left(\frac{\partial\langle N_\alpha\rangle}{\partial\mu_\beta}\right)_{VT\mu_{\beta'}}=k_BT\left(\frac{\partial\langle N_\beta\rangle}{\partial\mu_\alpha}\right)_{VT\mu_{\alpha'}}=\langle N_\alpha N_\beta\rangle-\langle N_\alpha\rangle\langle N_\beta\rangle \tag{2.51}$$

The following equations are derived by combining eq. 2.51 with eq. 2.47 and the number densities of species $\alpha$ and $\beta$.

$$\frac{k_BT}{V}\left(\frac{\partial\langle N_\alpha\rangle}{\partial\mu_\beta}\right)_{VT\mu_{\beta'}}=k_BT\left(\frac{\partial\rho_\alpha}{\partial\mu_\beta}\right)_{T\mu_{\beta'}}=\rho_\alpha\rho_\beta G_{\alpha\beta}+\rho_\alpha\delta_{\alpha\beta} \tag{2.52}$$

$$\left(\frac{\partial\mu_\beta}{\partial\rho_\alpha}\right)_{VT\mu_{\beta'}}=\frac{k_BT}{\rho_\alpha\rho_\beta G_{\alpha\beta}+\rho_\alpha\delta_{\alpha\beta}} \tag{2.53}$$

Finally, eq. 2.53 shows a connection between chemical potential derivatives and molecular distribution function. Ben-Naim presented how to derive the KBIs relationships.[85]

Particle number fluctuations are related to derivatives of chemical potentials with respect to compositions. However, the derivatives of eq. 2.53 are taken at constant chemical potentials. In most practical cases, the experimentalist is more interested in derivatives at constant pressure, *P* and temperature, *T*. Using some simple manipulations in partial derivatives, eq. 2.54 and eq. 2.55 follow from eq. 2.53.[85]

In the previous section 2.4.1, we introduced the chemical potential in the ideal/nonideal mixture using activity coefficient. The following relations between activity coefficient and KB integrals can be also derived with respect to molar concentration and mole fraction such as,

$$\left(\frac{\partial\ln y_A}{\partial\ln\rho_A}\right)_{T,P}=\frac{-\rho_A\left(G_{AA}-G_{AB}\right)}{1+\rho_A\left(G_{AA}-G_{AB}\right)} \tag{2.54}$$

$$\left(\frac{\partial \ln \gamma_A}{\partial \ln x_A}\right)_{T,P} = \frac{-\rho_B x_A\left(G_{AA}+G_{BB}-2G_{AB}\right)}{1+\rho_B x_A\left(G_{AA}+G_{BB}-2G_{AB}\right)} \tag{2.55}$$

In here, different notations for molar activity coefficient, *y* and mole fraction activity coefficient, $\gamma$ to distinguish the scale are used.

If species *A* and *B* are "very similar" like an ideal mixture, interactions of *A*-*A*, *A*-*B*, and *B*-*B* might be same, thus, $G_{AA} = G_{AB} = G_{BB}$. As a result, $\left(G_{AA} - G_{AB}\right)$ term in molar scale and $\left(G_{AA} + G_{BB} - 2G_{AB}\right)$ term in the mole fraction scale will be close to zero. In another case, "*A* likes *A*" or "*B* likes *B*" in the *A* and *B* mixtures, *i.e.,* $G_{AA} > G_{AB}$ or $G_{BB} > G_{AB}$, $\left(G_{AA} - G_{AB}\right)$ term in molar scale and $\left(G_{AA} + G_{BB} - 2G_{AB}\right)$ term in the mole fraction scale will be positive. These relations give direct information of aggregation of binary liquid mixtures.

So far, this section showed the details to induce relations between chemical potential derivatives and KB integrals. The KB theory is applicable for any kind of particles, not only for spherical molecules. Without any assumption, only the pair correlation functions appear in the relations. These are outstanding features in the KB theory. Therefore, the KB theory has become a powerful tool in the study of particularly complex fluids such as water and aqueous solutions.

# 3. A new force field of aqueous tertiary butanol solutions using atomistic simulations

## 3.1 Introduction

The structures of solutions depend on a balance of various intermolecular forces. In classical computer simulations, force fields are often developed by combining quantum chemically derived properties (bond lengths, angles, torsion barriers, partial atomic charges) and Lennard-Jones 12-6 potentials as well as partial atomic charges empirically parametrized to reproduce pure liquid properties such as its density and heat of vaporization.[87-90] The validity of these force fields in simulations of liquid mixtures may then be tested by comparison of, *e.g.,* solution densities, heats of mixing, dielectric constants, diffusion constants, compressibility, etc. with experimental data[91-93] or by comparison of calculated and experimental binary phase behavior.[94, 95] Although reproduction of target properties such as the density and heat of mixing provides a reasonable assessment of the quality of intermolecular potential energy functions, a realistic picture of the solution structure cannot always be guaranteed to emerge. In recent studies, Weerashinge and Smith [4, 6-8] adopted an alternative strategy using the Kirkwood–Buff (KB) theory[3] of solution in parametrizing co-solvent force fields applicable for atomistic simulations of binary water/co-solvent mixtures. The KB theory relates integrals over pair distribution functions (*i.e.,* KB integrals) to thermodynamic quantities such as isothermal compressibilities, partial molar volumes, and chemical potential (solution activity) derivatives. No approximations are made in this theory; therefore KB integrals obtained by computer simulation can be directly compared with thermodynamic data obtained from experiment. In addition, the KB integrals provide a computational inexpensive route to solution activities, which normally are not available during force field development due to a large computational cost of calculating free energies by thermodynamic integrations (TI). In nonideal mixtures, activity coefficients (defined on a molar concentration scale) vary with the solution composition in such a way that the corresponding changes in the excess chemical potentials are of the order of a few *kJ/mol*. For *tert*-butanol/water mixtures the maximum change amounts to 2–3 *kJ/mol* (fig. 3.8). This small free energy change cannot be calculated with sufficient accuracy in TI calculations, which usually provide error bars of 1–2 *kJ/mol*, but can be obtained very accurately from KB analyses, as we will show later on. Based on the KB approach Weerashinge and Smith developed models for urea,[6] acetone,[4] sodium chloride,[7] and guanidinium chloride[8] combined with three-site models (SPC/E,[96] SPC,[82] TIP3P[97]) for water. The advantage of the KB approach over alternative force field parametrization methods is that it guarantees the proper association behavior of solution components resulting from preferential interactions of molecules with molecules of the same or another kind, which must be described realistically when studying aspects of molecular recognition, ligand binding, chemical denaturation of proteins, *etc*.

In this chapter we present a tertiary butanol (TBA) model developed to reproduce the solution thermodynamics and aggregation behavior of TBA/water mixtures over the full composition range at 298 *K* and 1 atm. The physical chemistry of alcohol/water mixtures is particularly interesting as a direct result of the balance between the complex mixtures of hydrophobic and hydrogen bonding interactions that occur between the amphiphilic alcohols and hydrophilic water molecules. TBA is the largest alcohol molecule in the series of monohydric alcohols remaining fully miscible with water. The solution, however, behaves highly nonideal and significant excess coordination of TBA by surrounding alcohol molecules occurs at low alcohol concentrations as a result of hydrophobic interactions between the *tert*-butyl groups. Computer simulation studies of dilute aqueous *tert*-butanol solutions have been reported previously.[28-32] In these studies detailed analyses of the structural aspects of *tert*-butanol hydration were made. However, none have described the solution thermodynamics and aggregation of *tert*-butanol or water in detail. Simulations of aqueous TBA solutions, using parameters taken from the OPLS[81] and GROMOS[89] force fields to describe TBA combined with the SPC model[82] for water, indicated excessive aggregation of the alcohol and water molecules with almost a tenfold overestimate of the alcohol-alcohol and water-water KB integrals (see fig.3.1). The infinite dilution hydration free energies of the OPLS and GROMOS models in SPC water were found 6–7 *kJ/mol* larger than experiment. These observations clearly indicate that both TBA models are too strongly hydrophobic. We therefore decided to reparametrize the distribution and magnitude of the partial atomic charges of the GROMOS model resulting in a new, nonpolarizable, model with a larger dipole moment. The new model presented in this chapter reproduces the experimental KB integrals and their compositional dependencies. As a result, derivatives of solution activity coefficients, and the activity itself, are reproduced over the full range of solution compositions.

In the chapter 2.4.2, we have introduced KB theory which is used in the current work. First of all, the simulation details of the molecular dynamics (MD) and a new force field for TBA and water mixtures are described. Then in the next section, the results in terms of KB analysis and solution structure using our KB-derived force field are presented and discussed.

## 3.2 Simulation details

Simulations of TBA and water mixtures were carried out using the GROMOS96 biomolecular simulation package.[98, 99] In all simulations the temperature was kept constant at 298 $K$ by weakly coupling to a temperature bath with a relaxation time of 0.1 ps.[69] The pressure was maintained at 1 atm by also applying the weak coupling algorithm with a relaxation time of 0.5 ps[69] and an isothermal compressibility of $45.75 \times 10^{-5}$ ($kJ\ mol^{-1}\ nm^{-3}$)$^{-1}$. For nonbonded interactions, a twin-range method with cutoff radii of 0.8 and 1.4 nm was used. Outside the longer cutoff radius a reaction field correction was applied with a relative dielectric permittivity in the range between 12.5 (liquid TBA) and 64.8 (2 *mol %* TBA in water). The integration time step was 2 fs, the pairlist for pairs within the inner cutoff and the energies and forces for pairs between the inner and outer cutoff radii were updated every 10 fs. All bond lengths were kept constant using the SHAKE algorithm[75] with a relative geometrical tolerance of $10^{-4}$. Details on the system sizes, compositions, and simulation times are summarized in table 3.1. For all mixtures, a cubic box with an edge length of $\approx 4$ nm was prepared and simulated over a time span of 10–15 ns. All systems were equilibrated for at least 1 ns.

In addition to calculating the solution activity coefficients, free energy calculations based on thermodynamic integration (TI; see the chapter 2.3.3) were performed. Free energy changes were calculated using at least 50 $\lambda$ values. At each new $\lambda$, the system was first equilibrated for 50 ps after which the derivative in eq. 2.32 was sampled for at least 1 ns. A soft-core $\lambda$ scaling was used[99] to avoid singularities of the derivative at the end points. TI was used to calculate alcohol solvation free energies at mixture compositions $0.001 \le x_A \le 1$. Except for the pure liquid TBA system, the free energies were computed by decoupling the nonbonded interactions between a single alcohol molecule and all other alcohol and water molecules of the system at constant $P$ and $T$. The solvation free energy of TBA in its own pure liquid was calculated by decoupling the nonbonded interactions between all molecules of the system at constant $V$ and $T$.

*Table 3.1 Summary of MD simulations of TBA/water mixtures. The average system volumes were obtained using the partial atomic charges of set V in table 3.3. The numbers in parentheses correspond to larger system sizes.*

| $x_A$ | $N_{TBA}$ | $N_{H2O}$ | $t_{MD}$(ns) | $\varepsilon_{RF}$ | $V$ (nm$^3$) | $T$ (K) |
|---|---|---|---|---|---|---|
| 0.02 | 200 | 9800 | 14 | 64.8 | 329.5 | 301.8 |
| 0.04 | 74(171) | 1776(4096) | 10(10) | 64.8 | 65.29(150.82) | 299.7(301.4) |
| 0.06 | 105(262) | 1645(4096) | 10(10) | 64.0 | 65.26(164.53) | 299.3(301.0) |
| 0.10 | 168(500) | 1512(4096) | 10(10) | 50.7 | 71.73(201.76) | 298.6(300.3) |
| 0.16 | 210 | 1100 | 10 | 36.1 | 66.14 | 299.7 |
| 0.20 | 275(1000) | 1100(4096) | 10(10) | 36.1 | 76.24(281.07) | 297.5(299.4) |
| 0.30 | 336(1728) | 784(4096) | 10(10) | 30.0 | 76.70(397.75) | 296.9(298.9) |
| 0.50 | 432(2733) | 432(2744) | 10(10) | 18.0 | 82.16(523.07) | 296.3(298.2) |
| 0.60 | 450 | 300 | 10 | 15.0 | 81.36 | 296.1 |
| 0.70 | 470 | 200 | 10 | 12.7 | 81.97 | 295.9 |
| 0.80 | 480 | 120 | 10 | 11.9 | 81.50 | 295.8 |
| 0.86 | 490 | 80 | 10 | 11.8 | 82.50 | 297.8 |
| 0.90 | 500 | 55 | 10 | 11.8 | 83.04 | 295.7 |
| 0.95 | 1045 | 55 | 10 | 12.5 | 172.95 | 297.6 |
| 0.98 | 2695 | 55 | 10 | 12.5 | 444.10 | 297.8 |
| 1.00 | 512 | 0 | 5 | 12.5 | 83.92 | 297.5 |

## 3.3 *tert*-butanol and water models

For the *tert*-butanol (TBA) model, a six-site TBA model which is treated with a united atom-methyl group, was used. The bonded as well as nonbonded interaction parameters for TBA were taken from the GROMOS53A6 force field.[89] All internal nonbonded interactions were excluded.

The GROMOS53A6 Van der Waals interaction parameters of the central carbon atom and united atom methyl groups have been parameterized to, amongst other properties, reproduce free energies of hydrophobic hydration in SPC water[100] and are therefore expected to provide a reasonable description of hydrophobic interactions between *tert*-butyl groups.

*Table 3.2 Summary of the tert-butanol and SPC water force field parameters.*

| | Bonded terms | | |
|---|---|---|---|
| | Bond lengths, nm | | |
| O-H | 0.100 | | |
| C-O | 0.143 | | |
| $C\text{-}CH_3$ | 0.153 | | |
| | Angles, $V(\theta)=\frac{1}{2}k_\theta\left(\cos\theta-\cos\theta_0\right)^2$ | | |
| | $k_\theta\left(kJ/mol\right)$ | $\theta_0\left(\deg\right)$ | |
| $CH_3\text{-}C\text{-}CH_3$ | 520 | 109.5 | |
| H-O-C | 450 | 109.5 | |
| $O\text{-}C\text{-}CH_3$ | 520 | 109.5 | |
| | Torsions, $V(\varphi)=K_\varphi\left[1+\cos(\delta)\cos(m\varphi)\right]$ | | |
| | $K_\varphi\left(kJ/mol\right)$ | $\cos(\delta)$ | $m$ |
| $H\text{-}O\text{-}C\text{-}CH_3$ | 1.26 | 1.0 | 3 |
| | Nonbonded terms[a] | | |
| Atom types | $C_{12}(i,i)\left(10^{-6}kJ\ mol^{-1}nm^{-12}\right)$ | $C_6(i,i)\left(10^{-3}kJ\ mol^{-1}nm^{-6}\right)$ | $q_i(e)$[b] |
| | *tert*-butanol[c] | | |
| $CH_3$ | 26.64624 (34.2124) | 9.613802 (9.57465) | 0.0 |
| C | 205.3489 (7.59129) | 2.397081 (2.52124) | 0.337 |
| O | 1.505529[d]/1.265625[c](1.99550) | 2.261954 (2.38253) | -0.76 |
| H | 0.0 (0.0) | 0.0 (0.0) | 0.423 |
| | SPC water | | |
| OW | 2.634129 | 2.617346 | -0.82 |
| HW | 0.0 | 0.0 | 0.41 |

[a] Nonbonded interaction function,
$V\left(r_{ij}\right)=C_{12}\left(i,j\right)/r_{ij}^{12}-C_6\left(i,j\right)/r_{ij}^6+q_iq_j/4\pi\varepsilon_0\left[1/r_{ij}+(\varepsilon_{RF}-1)r_{ij}^2/(2\varepsilon_{RF}+1)R_C^3-3\varepsilon_{RF}/(2\varepsilon_{RF}+1)R_C\right]$,
$R_C$, long-range cutoff radius. $\varepsilon_{RF}$, reaction field relative dielectric permittivity. Combination rules, $C_6(i,j)=C_6(i,i)^{1/2}C_6(j,j)^{1/2}$; $C_{12}(i,j)=C_{12}(i,i)^{1/2}C_{12}(j,j)^{1/2}$.

[b] Derived in this work by parametrizing against solution activity data. The GROMOS and OPLS partial charges are listed in table 3.3.

[c] OPLS nonbonded parameters[81] are listed within parentheses.

[d] Used for O–O and O–OW nonbonded interactions.

[e] Used for O–C and O–$CH_3$ nonbonded interactions.

TBA molecule was also modeled by nonbonded 12-6 LJ parameters and partial atomic charges taken from the OPLS force field[81] in combination with the GROMOS bonded parameters. For water, the SPC model[82] was used. The GROMOS force field parameters have been summarized in table 3.2 together with the parameters for the SPC water.

Short test runs indicated that the KB integrals were very sensitive to the magnitude of the TBA dipole moment, modeled by the set of partial charges located on the hydroxyl oxygen and hydrogen atom, and the central carbon atom, while not much being affected by changes of the nonbonded Lennard-Jones parameters. We therefore decided to only reparametrize the partial atomic charges. Table 3.3 summarizes the charge distributions examined in this work. The initial set of partial charges (denoted by the Latin symbol I) was taken from the GROMOS force field of methanol.[89, 101] The second set (II) was taken from the OPLS force field.[81] Because with both charge distributions the extent of alcohol-alcohol as well as water-water aggregation was too strong (see below), we decided to systematically increase the molecular dipole moment of TBA, which is shown for each set (I–VIII) of partial charges in the last column of table 3.3.

*Table 3.3 Partial atomic charges and corresponding TBA dipole moments evaluated in this work.*

| | $q_C(e)$ | $q_O(e)$ | $q_H(e)$ | $\mu(D)$ |
|---|---|---|---|---|
| I(GROMOS,$CH_3OH$)[101] | 0.266 | -0.674 | 0.408 | 2.19 |
| II (OPLS)[81] | 0.265 | -0.700 | 0.435 | 2.26 |
| III | 0.295 | -0.740 | 0.445 | 2.41 |
| IV | 0.320 | -0.760 | 0.440 | 2.49 |
| V | 0.337 | -0.760 | 0.423 | 2.52 |
| VI | 0.350 | -0.760 | 0.410 | 2.55 |
| VII | 0.370 | -0.780 | 0.410 | 2.65 |
| VIII | 0.410 | -0.820 | 0.410 | 2.85 |

# 3.4 Results

## 3.4.1 KB analysis

In KB theory [3, 85] thermodynamic properties of a solution mixture are expressed in terms of KB integrals between solution components. In the chapter 2.4.2, we have introduced the KB theory. As we have seen eq. 2.41, KB integrals integrated from zero to correlation region, $R_C$, are assumed to be equal to KB integrals integrated from zero to infinity. If $R_C$ is small compared to the simulated system size, the KB integrals can be determined from simulation data in the *NPT* ensemble by assuming that

$$G_{ij} = \int_0^{R_C} \left[ g_{ij}^{NPT}(r) - 1 \right] 4\pi r^2 dr \tag{3.1}$$

where $G_{ij}$ is the KB integral for the pair of species *i* and *j*, $g_{ij}^{NPT}(r)$ is the corresponding radial distribution function in the isothermal-isobaric (*NPT*) ensemble, and *r* is the center of mass–center of mass distance.

Figure 3.1 shows the KB integrals, obtained by using the GROMOS and OPLS force fields (see tables 3.2 and 3.3), presented as a function of the alcohol mole fraction of solution. The experimental KB integrals (red circle symbols) were obtained from small-angle x-ray scattering.[102] Although the KB integrals obtained from experiment have a large error $\left(\sim \pm\, 100\ cm^3 / mol\right)$, the GROMOS and OPLS force fields give rise to a strong overestimation of alcohol-alcohol and water-water aggregation. Analyses of the heats of mixing for $0.001 \le x_A \le 0.9$ (table 3.4) and the free energy of hydration ($x_A$=0.001; table 3.5) are consistent with this observation. While the heats of mixing are consistently positive and large with deviations up to 0.7 *kJ/mol*, the free energies of hydration are overestimated by 6–7 *kJ/mol*.

The experimental free energy of hydration was obtained from the vapor pressure data reported by Koga[103] and the solution densities reported by Nakanishi[104] using the expression[105] $RT \ln y_A = RT \ln\left(P_A / \rho_A RT\right)_{eq}$, where $y_A$ is the molar scale activity coefficient and $P_A$ is the alcohol vapor pressure at equilibrium with the solution at an alcohol molar concentration $\rho_A$. The experimental heats of mixing were also obtained from the work of Koga *et al*.

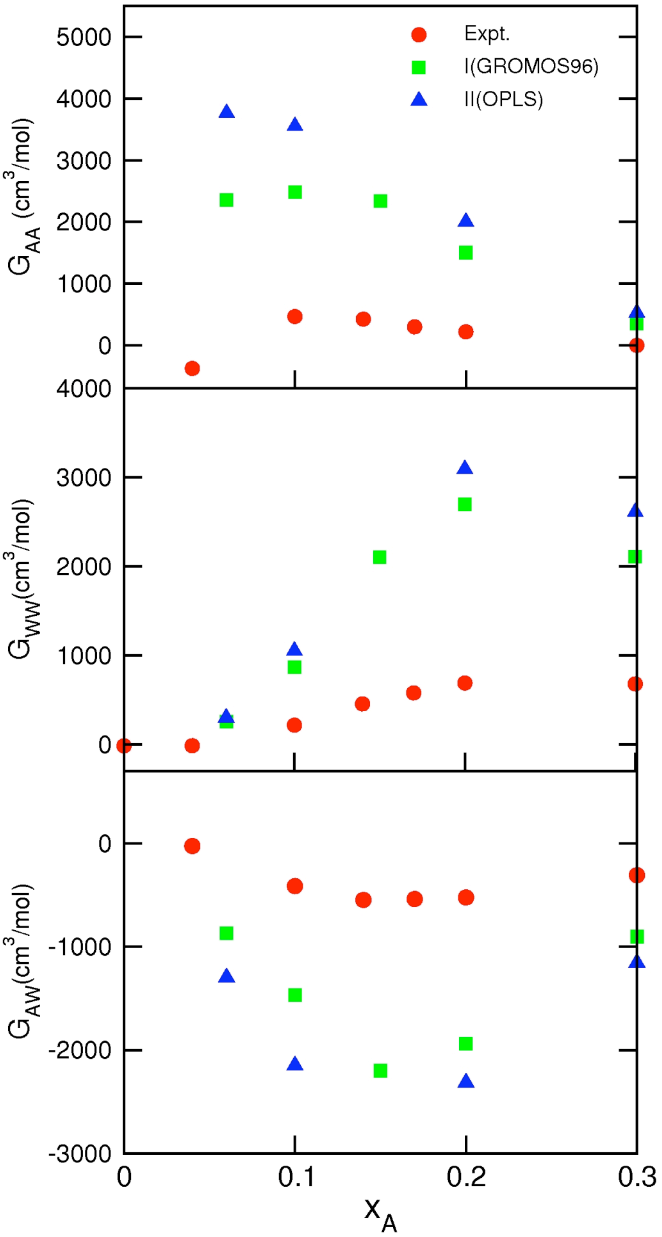


*Figure 3.1 Kirkwood–Buff integrals for TBA-TBA pairs ($G_{AA}$), water-water pairs ($G_{WW}$), and TBA-water pairs ($G_{AW}$) vs the TBA mole fraction ($x_A$) in solution. Experimental data (red circle)[102], GROMOS nonbonded van der Waals parameters and partial atomic charges (green square)[106], OPLS[81] nonbonded van der Waals parameters and partial atomic charges (blue triangle).*

*Table 3.4. Heats of mixing, $\Delta H_m$ (kJ mol$^{-1}$) using TBA partial atomic charge distributions I (GROMOS), II (OPLS), and V (table 3.3).*

| $x_A$ | Expt. | I | II | V |
|---|---|---|---|---|
| 0.001 | ··· | -0.12 | -0.12 | -0.12 |
| 0.04 | ··· | 0.07 | -0.06 | -0.10 |
| 0.06 | ··· | 0.00 | 0.22 | -0.09 |
| 0.1 | -0.52 | 0.08 | 0.09 | -0.03 |
| 0.2 | -0.42 | 0.24 | 0.27 | 0.10 |
| 0.3 | -0.25 | 0.57 | 0.49 | 0.26 |
| 0.5 | 0.17 | 0.84 | 0.73 | 0.53 |
| 0.6 | 0.29 | 0.92 | 0.81 | 0.64 |
| 0.7 | 0.35 | 0.87 | 0.62 | 0.60 |
| 0.8 | 0.36 | 0.71 | 0.62 | 0.45 |
| 0.9 | 0.29 | 0.40 | 0.25 | 0.29 |

Using $\Delta H_m = x_A H_m^E(TBA) + (1 - x_A) H_m^E(H_2O)$, in which $H_m^E(TBA)$ and $H_m^E(H_2O)$ are the excess partial molar enthalpies of the solution components reported in Ref. [103]. Solvent-solvent radial distribution functions (rdf), used to compute KB integrals (eq. 3.1), were defined as follows. The alcohol-alcohol radial distribution function, $g_{AA}(r)$, is based on the distances $r$ between the central carbon atoms of two TBA molecules, the water-water radial distribution function, $g_{WW}(r)$, is based on oxygen-oxygen distances, and the alcohol-water radial distribution function is based on central carbon-oxygen distances.

The rdf and the corresponding KB integrals as a function of the integration distance $r$ in eq. 3.1 are displayed in fig. 3.2 for a 0.5 mole fraction of TBA. While for long enough integration distances the KB integrals should approach a constant value, we observe oscillations at distances up to 1.6 nm indicating that packing effects still play a role at these distances. One may assume, however, that these oscillations slowly die out and the $G_{ij}$s will converge to values obtained by averaging over the last period of oscillation (between 1.1–1.6 nm).

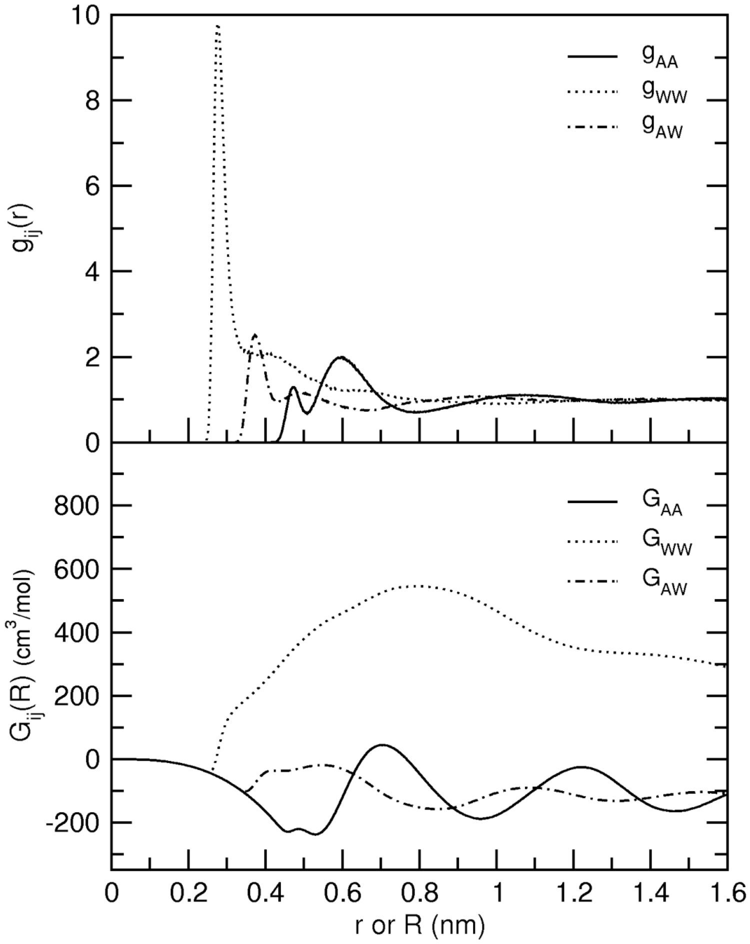


*Figure 3.2 Upper panel, Alcohol-alcohol (solid line), water-water (dotted line), and alcohol-water (dash-dotted line) radial distribution functions. Lower panel, The corresponding Kirkwood–Buff integrals (eq. 2.45) as a function of integration distance. The rdf and KB integrals correspond to a TBA mole fraction 0.5 for the small system size ($N_{TBA}$=432, table 3.1). (TBA partial atomic charges were from set V in table 3.3.)*

Consequently, some error is introduced in the values of $G_{AA}$ and $G_{AW}$ (±10–30 *cm*$^3$*/mol*), while the $G_{WW}$ value seems to have a larger uncertainty. We note that converged values of the KB integrals were obtained only after 7–10 ns sampling time. Therefore, with all systems, simulation times up to 10 ns were performed. A second source of error in our calculations of the KB integrals may arise due to finite size effects. TBA/water mixtures exhibit fairly strong aggregation behavior, which may be artificially suppressed when simulating too small systems. Figure 3.3 shows the KB integrals obtained at, respectively, 0.1 (left panel) and 0.2 (right panel) mole fraction alcohol solution with two system sizes. The box lengths of the smaller systems are $\approx 4$ nm, while those of the larger systems are $\approx 6$ nm. In table 3.1, more details of the small and larger systems are summarized. Full lines represent the KB integrals obtained from simulations of the large systems, while dashed lines represent KB integrals obtained with the smaller system sizes. The major differences between the two systems are manifested in the longer-range correlations.

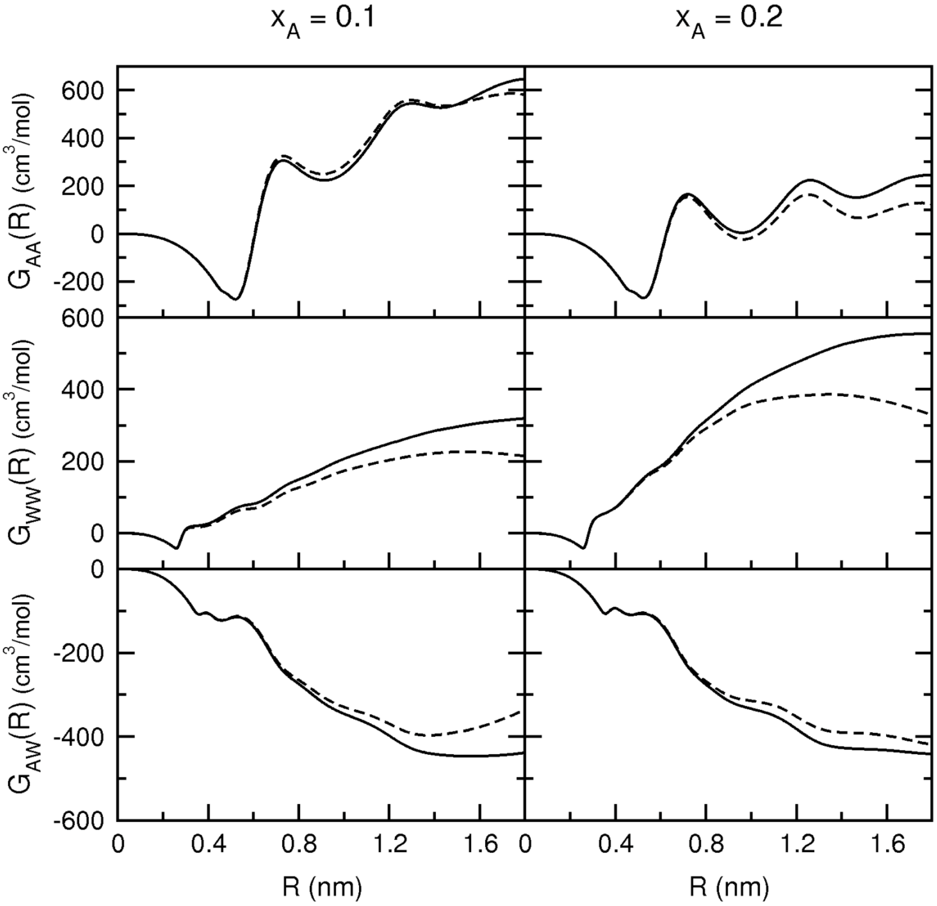


*Figure. 3.3 Alcohol-alcohol (upper panel), water-water (middle panel), and alcohol-water (lower panel) KB integrals for large (full line) and small (dashed line) system sizes (cf. table 3.1) at 0.1 mole fraction of TBA solution (left column) and at 0.2 mole fraction of TBA solution (right column). (TBA partial atomic charges were from set V in table 3.3.)*

The larger systems clearly permit a stronger degree of water-water and TBA-TBA aggregation, while solvation of TBA by water becomes weaker. Unlike $G_{AA}$, which shows an oscillating trend, $G_{WW}$ increases monotonically at distances between 1.0 – 2.0 nm.

Typically, the water-water rdf slowly decays from $\approx 1.03$ to 1.0 in this region. The alcohol-water rdf on the other hand slowly increases from 0.97 to 1.0 at distances between 1.0 and 2.0 nm, causing $G_{AW}$ to decrease and reaching a constant value at about 2.0 nm. We therefore decided to prepare larger systems (cf. table 3.1) for the solutions with alcohol mole fractions between 0.02 and 0.50 where the largest extent of excess aggregation behavior occurs (see below). All data presented below were extracted from the simulations with the large system size.

Figure 3.4 shows the KB integrals for a given solvent composition presented versus the molecular dipole moments of the TBA models summarized in table 3.3. The dashed horizontal lines are the experimental values[33] to be reproduced by our optimized TBA model. Clearly for $\mu < 2.5\ D$ the systems exhibit too strong alcohol-alcohol and water-water aggregation while at the same time the solvation of alcohol by water is much too weak. For $\mu \rightarrow 2.85D$ the limit of ideal mixing is approached and all KB integrals tend to approach zero.

*Table 3.5 Infinite dilution free energies of hydration ($x_A$=0.001) of TBA in SPC water for TBA partial atomic charge distributions I–VIII (table 3.3).*

| | $RT \ln y_A (kJ\ mol^{-1})$ |
|---|---|
| Expt. | -19.0 |
| I(GROMOS) | -12.1 |
| II(OPLS) | -13.1 |
| III | -19.3 |
| IV | -18.8 |
| V | -19.4 |
| VI | -17.4 |
| VII | -18.3 |
| VIII | -32.6 |

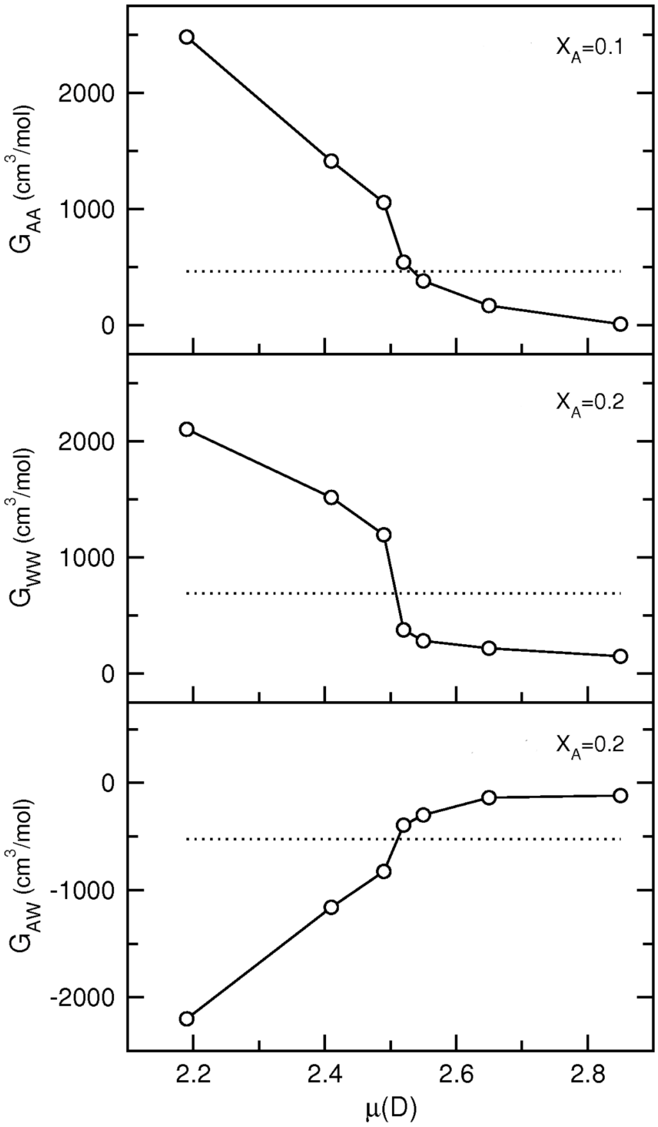


*Figure 3.4 Alcohol-alcohol (upper panel, $x_A$=0.1), water-water (middle panel, $x_A$=0.2), and alcohol-water (lower panel, $x_A$=0.2) KB integrals (○) presented vs the molecular dipole moment of the TBA models in table 3.3. Dashed horizontal lines correspond to the experimental KB integrals [102].*

In the vicinity of the experimental KB integrals the curves make a rather sharp transition between these two extremes of solution behavior. If the TBA model was assigned a molecular dipole moment of ≈ 2.5 *D* the experimental KB integrals were closely reproduced. Our final model (TBA partial charge distribution corresponding to set V in table 3.3) has a dipole moment $\mu = 2.52\ D$ compared to the gas phase experimental value of 1.66 $D$.[107] The infinite dilution free energy of hydration of this TBA model is in agreement with the experimental value (table 3.5).

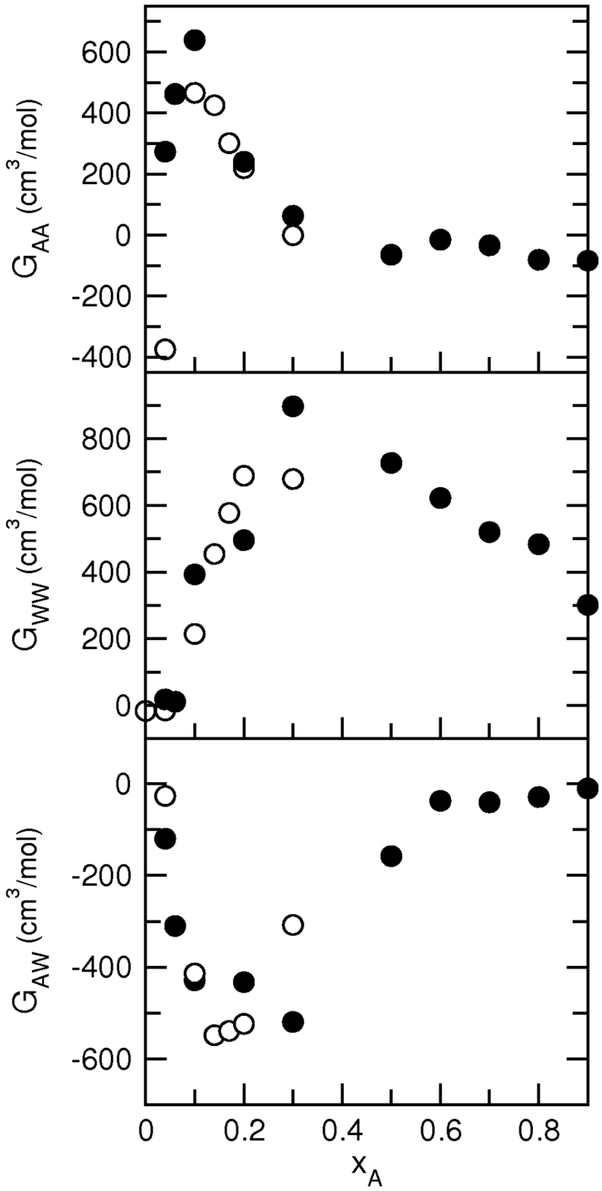


*Figure 3.5 Alcohol-alcohol (upper panel), water-water (middle panel), and alcohol-water (lower panel) KB integrals vs solution composition. Values computed using charge distribution set V (table 3.3) (●), experimental data (○)*[102].

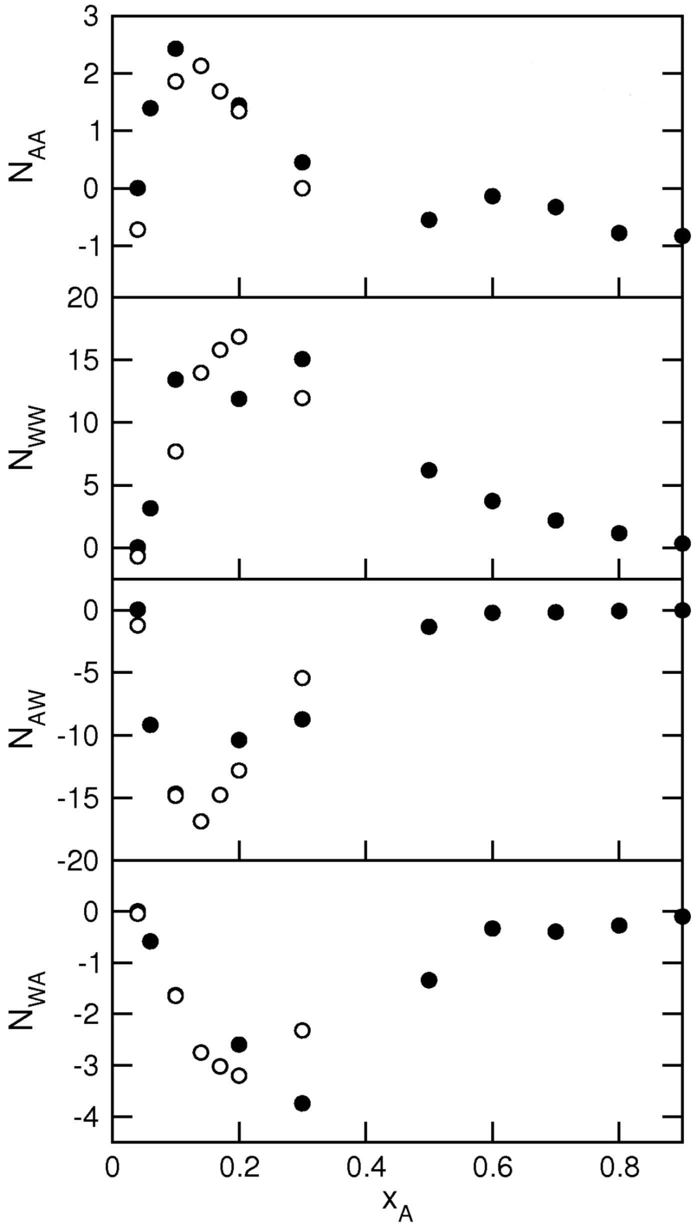


*Figure 3.6 Excess coordination numbers (●), $N_{ij}=\rho_j G_{ij}$. The experimental data points (○) were obtained by combining the KB integrals reported by Nishikawa et al.[102]) and the solution densities reported by Nakanishi et al.[104].*

Figure 3.5 and 3.6 show the compositional dependence of the KB integrals and excess coordination numbers, $N_{ij} = \rho_j G_{ij}$, (see the chapter 2.4.2) respectively, obtained with our final TBA model. The KB integrals are summarized in table 3.6. The agreement with experiment is satisfactory, in particular, when considering the uncertainties between different sets of experiments that may amount to differences of the KB integrals up to 100 *cm³/mol*.[102] Excess alcohol-alcohol aggregation runs through a maximum at an alcohol mole fraction $x_A \cong 0.1$, while a maximum in excess water-water aggregation occurs at $x_A \cong 0.3$.

*Table 3.6 Summary of total number densities,* $\rho = \rho_A + \rho_W$, *and Kirkwood–Buff integrals for TBA partial atomic charge distribution V in table 3.3. For* $x_A \leq 0.5$, *KB integrals were obtained by averaging the* $G_{ij}(R)$*s over intervals* $1.7 \leq R \leq 2.5$ *nm. For* $x_A > 0.5$, *intervals* $1.1 \leq R \leq 1.6$ *nm were typically used.*

| $x_A$ | $\rho\left(mol\ cm^{-3}\right)$ | $G_{AA}\left(cm^3 mol^{-1}\right)$ | $G_{WW}\left(cm^3 mol^{-1}\right)$ | $G_{AW}\left(cm^3 mol^{-1}\right)$ |
|---|---|---|---|---|
| 0.02 | 0.0504 | 110 | -10 | -100 |
| 0.04 | 0.0471 | 273 | 18 | -120 |
| 0.06 | 0.0445 | 462 | 11 | -310 |
| 0.1 | 0.0389 | 639 | 393 | -429 |
| 0.2 | 0.0300 | 240 | 496 | -432 |
| 0.3 | 0.0243 | 63 | 897 | -519 |
| 0.5 | 0.0175 | -65 | 727 | -158 |
| 0.6 | 0.0153 | -15 | 622 | -37 |
| 0.7 | 0.0136 | -34 | 520 | -40 |
| 0.8 | 0.0122 | -81 | 483 | -29 |
| 0.9 | 0.0111 | -84 | 301 | -10 |
| 0.95 | 0.0106 | -172 | 592 | -118 |
| 0.98 | 0.0103 | -72 | 1047 | 12 |

Table 3.7 shows the first shell coordination numbers obtained by integrating the rdf up to their first minimum. The numbers within parentheses denote *excess* first shell coordination numbers obtained by subtracting the number of molecules obtained by integrating over the same distance from a randomly chosen location in the fluid (*i.e.*, the number of solvent components found by releasing the constraint of having solvent species *A* (or *W*) at the center of the spherical volume). Despite the excluded volume repulsion of the central solvent component playing a significant role within short radii up to 7 Å, the excess first shell TBA-TBA and water-water coordination numbers indicate that a significant attraction occurs (*i.e.*, both numbers are positive) between components of the same kind.

*Table 3.7 First shell coordination numbers and excess first shell coordination numbers (within parentheses) for TBA partial atomic charge distribution V in table 3.3. The distance to the first minimum of the rdf was 7.8 Å for TBA-TBA, 3.4 Å for water-water and 4.1 Å for TBA-water.*

| $x_A$ | $N_{AA}$ | $N_{WW}$ | $N_{AW}$ | $N_{WA}$ |
|---|---|---|---|---|
| 0.04 | 2.8 (0.6) | 4.4 (0.1) | 2.7 (-5.2) | 0.1 (-0.2) |
| 0.06 | 3.6 (0.9) | 4.6 (0.1) | 2.6 (-4.8) | 0.2 (-0.3) |
| 0.10 | 5.8 (1.1) | 4.2 (0.7) | 2.3 (-3.8) | 0.3 (-0.4) |
| 0.20 | 8.0 (0.8) | 3.6 (1.2) | 1.9 (-2.3) | 0.5 (-0.6) |
| 0.30 | 9.1 (0.3) | 3.2 (1.5) | 1.7 (-1.3) | 0.7 (-0.6) |
| 0.50 | 10.4 (-0.1) | 2.5 (1.6) | 1.2 (-0.3) | 1.2 (-0.3) |
| 0.60 | 10.9 (-0.2) | 2.1 (1.4) | 1.0 (-0.1) | 1.5 (-0.2) |
| 0.70 | 11.2 (-0.2) | 1.7 (1.3) | 0.7 (0.0) | 1.7 (0.1) |
| 0.80 | 11.4 (-0.4) | 1.2 (0.9) | 0.5 (0.1) | 2.0 (0.3) |
| 0.86 | 11.4 (-0.5) | 1.0 (0.8) | 0.4 (0.1) | 2.2 (0.4) |
| 0.90 | 11.5 (-0.5) | 0.6 (0.5) | 0.3 (0.1) | 2.4 (0.7) |
| 0.95 | 11.5 (-0.6) | 0.4 (0.3) | 0.1 (0.0) | 2.6 (0.8) |
| 0.98 | 11.6 (-0.6) | 0.2 (0.1) | 0.1 (0.0) | 2.8 (1.0) |

On the other hand, the excess first shell TBA-$H_2O$ and $H_2O$-TBA coordination numbers are negative for $x_A < 0.7$ while becoming positive at higher alcohol mole fractions indicative of alcohol-water attraction in the alcohol-rich solutions. When the excess first shell coordination numbers are compared to the excess coordination numbers in fig. 3.6, it shows that a significant contribution to $N_{AA}$ stems from direct, first shell interactions. The major contribution to $N_{WW}$, however, stems from indirect, longer-ranged correlations. It is probably this fact that causes water-water aggregation to equilibrate slower and $G_{WW}$ to have larger uncertainties (fig. 3.2). In fact, the largest uncertainties in experimental data sets[102] exist in the water-water KB integrals as well.

Figure 3.7 (upper panel) shows the derivative of the mole fraction scale activity coefficient (eq. 2.55) versus the alcohol mole fraction of the solution. The experimental activity coefficient derivatives (open circles) were calculated from vapor pressure data reported by Koga[103] by taking finite differences.

In terms of the notation used in Ref. [103], the mole fraction scale activity coefficient, $\gamma_A$ (*cf.* the molar scale activity coefficient, $y_A$) is written as $\ln \gamma_A = G_m^E(TBA)/RT$, where $G_m^E(TBA) = RT\ln(p_A / x_A p_A^0)$ is the excess partial molar free energy of TBA on the mole fraction scale ($p_A$ denotes the TBA partial vapor pressure and $p_A^0$ the vapor pressure of the pure liquid component). Also included are the activity coefficient derivatives obtained from SAXS studies by Nishikawa[102] (open squares). Clearly, the correct trends and magnitudes are observed in the solution activity derivates obtained from the simulations. With the GROMOS and OPLS models the activity derivatives lie below the experimental values (not shown) and reach a value of $-1$ at $x_A = 0.2$ indicating that the solutions became unstable at this composition.[85] The line included in fig. 3.7 (upper panel) is a numerical fit using Wilson's equation[108] for the excess free energy of a binary solution. Wilson's fit assumes ideal mixing entropy. This means that all deviations from ideal behavior come from differences between alcohol-alcohol, alcohol-water, and water-water interactions. In alcohol and water mixtures, the assumption of ideal mixing entropy is invalid. This equation has two adjustable parameters, which we used as fit parameters, but, in the original derivation, are related to the pure-component molar volumes and characteristic energy differences. In the water-rich region, where solvent-mediated hydrophobic interactions dominate, the fit poorly matches the data.

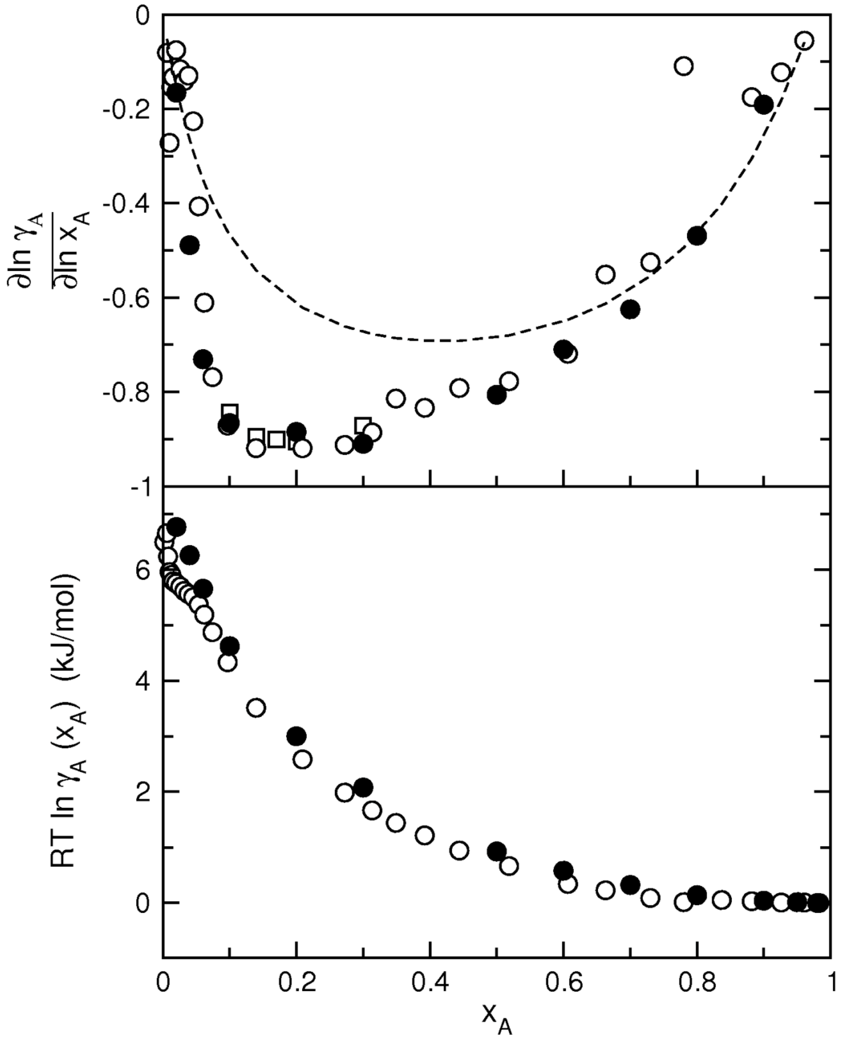


*Figure 3.7 Upper panel, mole fraction scale activity coefficient derivatives as a function of the TBA mole fraction. Computed values (eq. 2.55) (TBA partial atomic charges from set V in table 3.3) (●). Activity coefficient derivatives derived based on vapor pressure data (○)*[103]*. Activity coefficient derivatives obtained by combining (cf. eq. 2.55) SAXS-derived KB integrals of Nishikawa et al.* [102] *and solution density data of Nakanishi et al. (□)*[104] *The dashed line is a fit to the data of Ref. [103] based on Wilson's equation*[108]*). Lower panel, $RT \ln \gamma_A$ vs the alcohol mole fraction obtained by numerical integration of the data in upper panel.*

Most importantly, the data in fig. 3.7 indicate that the solution activities would be correctly reproduced on integration of these derivatives. To perform the integration we write the chemical potential of the alcohol as[85]

$$\mu_A(T,P,x_A) = \mu_A^p(T,P) + RT\ln x_A + RT\int_0^{x_W} \frac{\rho x'_W \Delta_{AW}}{1+\rho x'_A x'_W \Delta_{AW}} dx'_W$$

$$\equiv \mu_A^p(T,P) + RT\ln x_A + RT\ln\gamma_A \qquad (3.2)$$

where $\mu_A^p$ is the chemical potential of pure TBA. The lower panel in fig. 3.7 shows $RT\ln\gamma_A$ presented versus the alcohol mole fraction. The agreement between the simulated and experimental data is quite satisfactory except in the lowest concentration region ($x_A \leq 0.04$).

Figure 3.7 provides a route to interpret compositional derivatives of mole fraction scale-based activity coefficients in terms of solution structures. The $RT\ln\gamma_A$ in terms of the free energy change due to interactions of species *A* with surrounding molecules *A* and *W* is interpreted (chapter 2.4.1). Changes of this quantity with the solution composition are determined by the values of the KB integrals $G_{AA}$ and $G_{AW}$ (eq. 2.55) only; *i.e.,* the free energy ("coupling work") of solvating *A* changes with solution composition if *A* attracts or repels species of the same kind relative to species of the other kind. If *A* preferentially attracts *A* ($G_{AA}$>$G_{AW}$), the coupling work will have a negative contribution (lowering the chemical potential) upon increasing *A*, while the opposite occurs if *A* preferentially attracts *W* ($G_{AW}$>$G_{AA}$). If *A* neither attracts nor repels species of type *A* in comparison to species of type *W* (ideal solution), compositional changes will leave $RT\ln\gamma_A$ unaffected.

Here, we also show $RT\ln y_A$ (with $y_A$ the activity coefficient on the molar concentration scale) in order to account for these interactions. Figure 3.8 shows $RT\ln y_A$ presented versus the alcohol molar concentration. The open circles represent the experimental values (*i.e.,* $RT\ln y_A = RT\ln(\rho_A^{vap}/\rho_A)_{eq}$)[105] obtained from Koga's vapor pressure data[103] together with the solution density data determined by Nakanishi.[104] Three well-defined regions can be distinguished. For $0.8 < \rho_A < 2$ *mol/l*, Henry's law applies ($RT\ln y_A$ is constant) and neither preferential alcohol-alcohol nor preferential alcohol-water interactions occur. For

$2<\rho_A<9$ *mol/l*, preferential alcohol-alcohol interactions dominate (negative slope), while for $\rho_A>9$ *mol/l* preferential alcohol-water interactions take over (positive slope). The closed circles in fig. 3.8 are data obtained by integration of the molar scale activity coefficient derivatives (eq. 2.54) obtained by the KB analysis. Integration of eq. 2.54 yields $RT\ln y_A$ determined up to an arbitrary constant, which we used to match the experimental curve. It is important, however, to also know the absolute values of $RT\ln y_A$ predicted by our final model (model V). These we determined with TI calculations (eq. 2.32) The TI results are included in fig. 3.8 (squares).

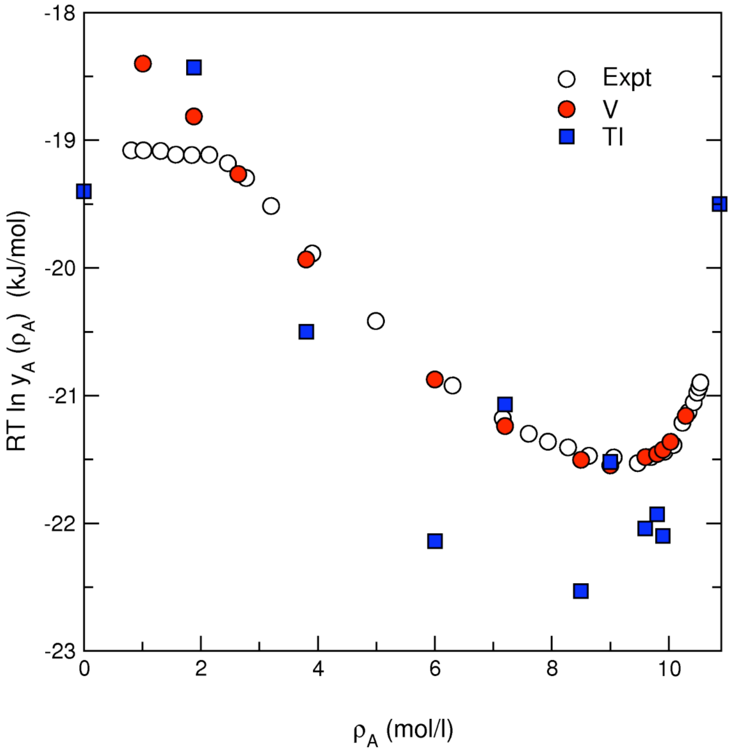


*Figure 3.8 RT ln* $y_A$ *vs the alcohol molar concentration (mol/l). Computed values obtained by numerical integration of eq. 2.54 (TBA partial atomic charges from set V in table 3.3) (red circle). Experimental data* $RT\ln y_A = RT\ln\left(\rho_A^{vap}/\rho_A\right)_{eq}$ *(see text) computed using vapor pressure*[103] *and solution density*[104] *data (white circle). Calculations based on free energy calculations by TI (eq. 2.32) (blue square).*

The statistical error in these calculations amounts to ±1 *kJ/mol*. The nonideality of the solution causes the experimental solvation free energies to decrease by $\approx 2.5$ *kJ/mol* when increasing the alcohol concentration from $\rho_A = 1$ *mol/l* to $\rho_A = 9$ *mol/l.*

On this energy scale, the predictions obtained from the TI calculations scatter considerably with varying solution composition and therefore do not allow extracting compositional dependencies of the alcohol activity. The KB derived data, on the other hand, reproduce the experimental trend well. Despite the error bars intrinsic to the method, the TI results indicate that our final model closely reproduces the absolute values of the alcohol solvation free energies. The comparison of KB and TI predicted solution activities in fig. 3.8 clearly highlight the benefits of predicting the solution thermodynamics by the KB approach. We note that the model does not reproduce Henry's law behavior which occurs in the ideal dilute solution because the TBA/water mixture is a highly nonideal behavior at low TBA concentration.

The solution density of model V is compared with the experimental densities[104] in fig. 3.9 (upper panel). Also included are the solution densities as predicted using the GROMOS and OPLS models. Over the full composition range our model underestimates the solution density by 2.5% (pure SPC water) up to 4% (pure TBA). A slight improvement was obtained over the GROMOS model. The lower two panels in fig. 3.9 show the TBA and water partial molar volumes as a function of solution composition obtained by performing a tangent construction to the density data (circles). We also calculated the partial molar volumes (squares) of alcohol $\bar{V}_A$ and water $\bar{V}_W$ using the following relations,[85]

$$\bar{V}_A = \frac{1+\rho_W(G_{WW}-G_{AW})}{\eta}, \quad \bar{V}_W = \frac{1+\rho_A(G_{AA}-G_{AW})}{\eta} \tag{3.3}$$

where $\rho_W$ *and* $\rho_A$ are the number densities and $\eta = \rho_A + \rho_W + \rho_A\rho_W\left(G_{AA}+G_{WW}-2G_{AW}\right)$. The uncertainties in the KB integrals, however, introduced significant errors. Qualitatively correct trends are observed in the partial molar volumes; the alcohol partial molar volume (middle panel) is a strongly increasing function of composition in low mole fraction alcohol solution, while at higher mole fractions ($x_A$>0.1) a much weaker variation with composition is observed.

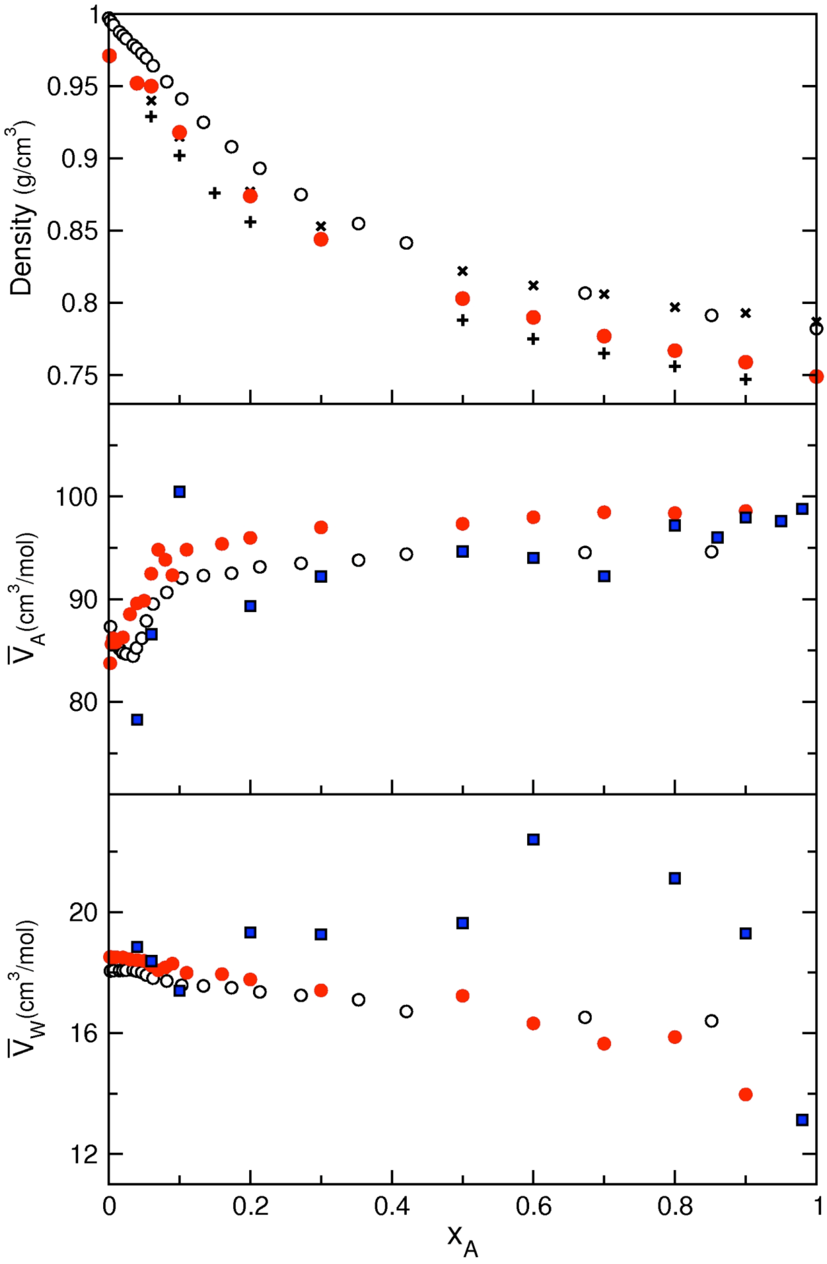


*Figure 3.9 Upper panel, mass densities (g/cm$^3$) of TBA/$H_2O$ solutions. Experimental data*[104] *(○). Calculated densities using GROMOS charges (table 3.3) (plus). Calculated densities using OPLS charges (table 3.3) and van der Waals nonbonded parameters (table 3.2) (cross). Calculated densities using partial atomic charge set V (table 3.3) (red circle). Middle panel, partial molar volumes of TBA (KB derived partial molar volumes are depicted by the blue squares (eq. 3.3). Other symbols; see upper panel). Lower panel, partial molar volumes of water. The partial molar volumes were obtained by taking tangents in upper panel.*

Our simulation results systematically run above the experimental values as a consequence of the too low solution density. Moreover, we do not observe a minimum at low alcohol concentrations, typical of alcohol/water systems. The water partial molar volume (lower panel), as observed experimentally, decreases from 18.1 $cm^3/mol$ (pure water) to $\approx$ 16.5 $cm^3/mol$ in 0.85 mole fraction TBA solution. The prediction from simulation again is in reasonable agreement with the experimental data. In the water-rich regions we overestimate the water partial molar volume because of a too low liquid density of SPC water. In the TBA-rich regions the simulation predictions lie below the experimental data. Figure 3.10 shows the heat of mixing, $\Delta H_{mix}$, the isothermal compressibility and the static relative dielectric permittivity, $\varepsilon$ (0) of model V versus the mole fraction of alcohol solution with experimental data. For the heat of mixing, $\Delta H_{mix}$, our final model (V) follows the experimental[103] trend, but systematically overestimates $\Delta H_{mix}$ by $\approx$0.5 *kJ/mol.* Although there is an improvement over the OPLS and GROMOS models, the agreement with the experimental data could have been even better by using additional parameterizations. We found, however, that further improvement of $\Delta H_{mix}$ occurred at the expense of introducing larger disagreement with the experimental KB integrals. Enthalpies of mixing for nonionic aqueous solutions are difficult to reproduce in atomistic simulations and have recently been discussed by Perera and Sokolic for the case of the acetone-water mixture.[19] There it was shown that with various acetone force fields deviations of about 1 *kJ/mol* are common. Perera and Sokolic discuss that the mixing enthalpy has a S-shape dependence on the mole fraction of the organic solute and is exothermic in the water-rich region while turning endothermic in the solute-rich region. Similar behavior occurs in TBA-water. In Ref. [19] it is argued that in the solute-rich region the hydrogen-bonded network of water is strongly disrupted due to dominating solute percolation (leading to an unfavourable enthalpy of mixing), while in the water-rich region solute insertion stabilizes the water network (leading to a favourable enthalpy of mixing). From this interpretation it becomes clear that changes of water-water interactions upon solute insertion must be described realistically to reproduce mixing enthalpies. The 0.5 *kJ/mol* offset observed in fig. 3.10 could therefore be a feature caused by our choice of the SPC water model. Indeed solute-solvent interactions also play an important role in describing the heat of mixing correctly.

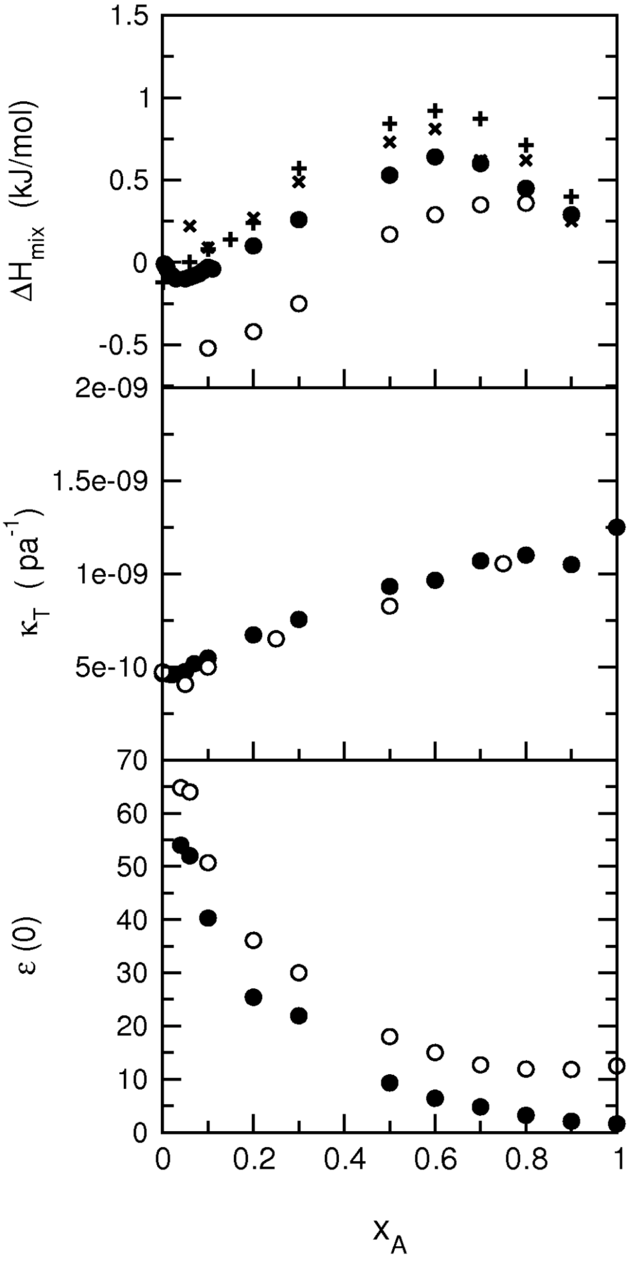


*Figure 3.10 Upper panel, enthalpies of mixing* $\Delta H_{mix}(kJ/mol)$ *; experimental data (○)*[103]*, GROMOS partial charges (plus), OPLS partial charges and nonbonded van der Waals parameters (cross), and TBA partial atomic charge set V (black circle). Middle panel, Isothermal compressibilities* $\kappa_T$ *(*$Pa^{-1}$*); experimental data (white circle)*[109]*, TBA partial atomic charge set V (black circle). Lower panel, Relative dielectric permittivities (0); experimental data (white circle)*[110]*, TBA partial atomic charge set V (black circle).*

We are, however, confident that our model describes these interactions correctly. As we discussed above, the solute activity ($RT\ln y_A$) in fig. 3.8 is a direct measure of solute-solvent interactions. The agreement with the experimental data in fig. 3.8 indicates that these interactions are well described by our model.

The isothermal compressibility $\kappa_T$ (fig. 3.10, middle panel) was calculated by a finite difference approximation. Two additional 1 ns NVT simulations close to the original state point were performed using a ±5% change in density.[111] The compressibilities (model V) are slightly higher than experimental[109] counterparts because the density of model V is lower than experimental data. The compositional dependence of the calculated $\kappa_T$ closely follows the experiment. The lower panel in fig. 3.10 shows the static relative dielectric permittivity, $\varepsilon(0)$, (model V) presented versus the solution composition. $\varepsilon(0)$ was computed from the fluctuation of the total dipole moment of the system. The relation between the relative permittivity and the fluctuation of the total dipole moment depends on the way the long-range forces are evaluated. When applying a reaction field method, $\varepsilon(0)$ can be obtained according to[112]

$$(\varepsilon(0)-1)\left(\frac{2\varepsilon_{RF}+1}{2\varepsilon_{RF}+\varepsilon(0)}\right)=\frac{\langle \boldsymbol{M}^2\rangle-\langle \boldsymbol{M}\rangle^2}{3\varepsilon_0 V k_B T} \tag{3.4}$$

where $\varepsilon_{RF}$ is the relative dielectric permittivity of the continuum that is used in the simulation (table 3.1), $\boldsymbol{M}$ the total dipole moment of the system, $V$ the volume of the simulation box, $k_B$ is the Boltzmann constant, $T$ the absolute temperature, and 0 is the dielectric permittivity of vacuum. In all cases (0) was computed from 10 ns trajectories to obtain well-converged values of $\boldsymbol{<M^2>}$. The experimental dielectric permittivities[110] are shown as open symbols. The computed values of $\varepsilon(0)$ are lower than the experimentally observed values. In the water-rich region the underestimation can be explained from the fact that the SPC water model has a dielectric permittivity of 67 (compared to the experimental value of 78.5).[111] The underestimation in the TBA-rich region must be due to approximations in the TBA model and its parametrization. Also, the neglect of electronic distortional polarizations may figure significantly in the low dielectric *tert*-butanol. Recall that only the reorientational contribution to $\varepsilon(0)$ has been calculated.

### 3.4.2 Solution Structure

Figure 3.11 shows alcohol-alcohol (upper panel), alcohol-water (middle panel), and water-water (lower panel) rdf at four alcohol concentrations. The alcohol-alcohol rdf (upper panel, left) shows a broad peak at ~ 5.8 Å due to hydrophobic interactions of the *tert*-butyl groups. With increasing alcohol concentration $x_A \geq 0.10$, alcohol-alcohol hydrogen bonding (upper panel, right) becomes more dominant at the expense of the hydrophobic interactions between *tert*-butyl groups. At the lowest alcohol concentration (4 *mol %*) no alcohol-alcohol hydrogen bonding is visible. Alcohol-water hydrogen bonding as well as water-water hydrogen bonding gets more pronounced with increasing alcohol concentration. Comparison of the H-OW and O-HW rdf reveals that *tert*-butyl alcohol is a slightly better hydrogen bond donor than acceptor, in agreement with the work of Bowron *et al.*[113] Comparison of alcohol-water and water-water hydrogen bonding shows that water prefers to hydrogen bond to itself. At $x_A \geq 0.10$, the OW-OW and OW-HW rdf exceed unity up to distances of about 8–10 Å, suggesting that spatially extended network structures of hydrogen bonded water molecules exist at these solution compositions.

Bowron *et al.* studied the aqueous solution structure of *tert*-butanol in the dilute and concentrated regions by neutron diffraction (ND) techniques using hydrogen/deuterium isotopic substitution.[113, 114] They extracted detailed information relating to the solution structure using an empirical potential structure refinement (EPSR) procedure. In this Monte Carlo (MC) procedure intermolecular potentials were iteratively adjusted such as to reproduce composite radial distribution functions, consisting of several site-site contributions. Based on the optimized potential energy functions, all site-site radial distribution functions were extracted from an ensemble of configurations generated in a subsequent MC simulation. This modeling procedure introduces some ambiguity in the derived pair correlation functions because the quality of the potential energy function often remains unclear. Figure 3.12 compares the TBA-TBA packing obtained from the MD simulations performed in the current work and the work from Bowron *et al.* in the dilute region.[113] Shown are the radial distribution functions based on distances between the central carbon atoms of the alcohol molecules. The solid lines represent the results obtained in our simulations while the symbols are the data of Bowron *et al.*[113] The overall agreement between the two is quite satisfactory, given the differences in approximations made in this work and the EPSR modeling.

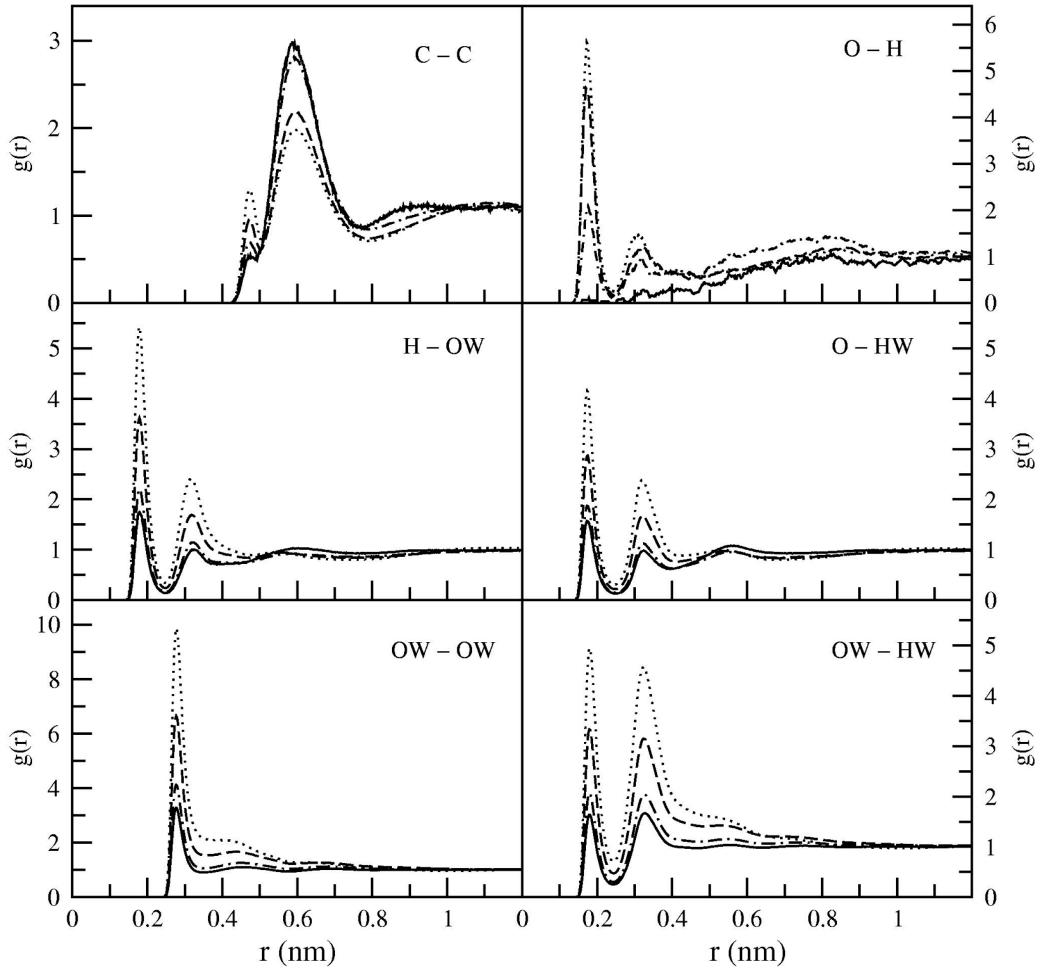


*Figure 3.11 Radial distribution functions at TBA mole fractions 0.04 (solid line), 0.1 (dash dotted), 0.3 (dashed), and 0.5 (dotted). C, central TBA carbon atom; O, TBA hydroxyl oxygen; H, TBA hydroxyl hydrogen; OW, HW, water oxygen and hydrogen.*

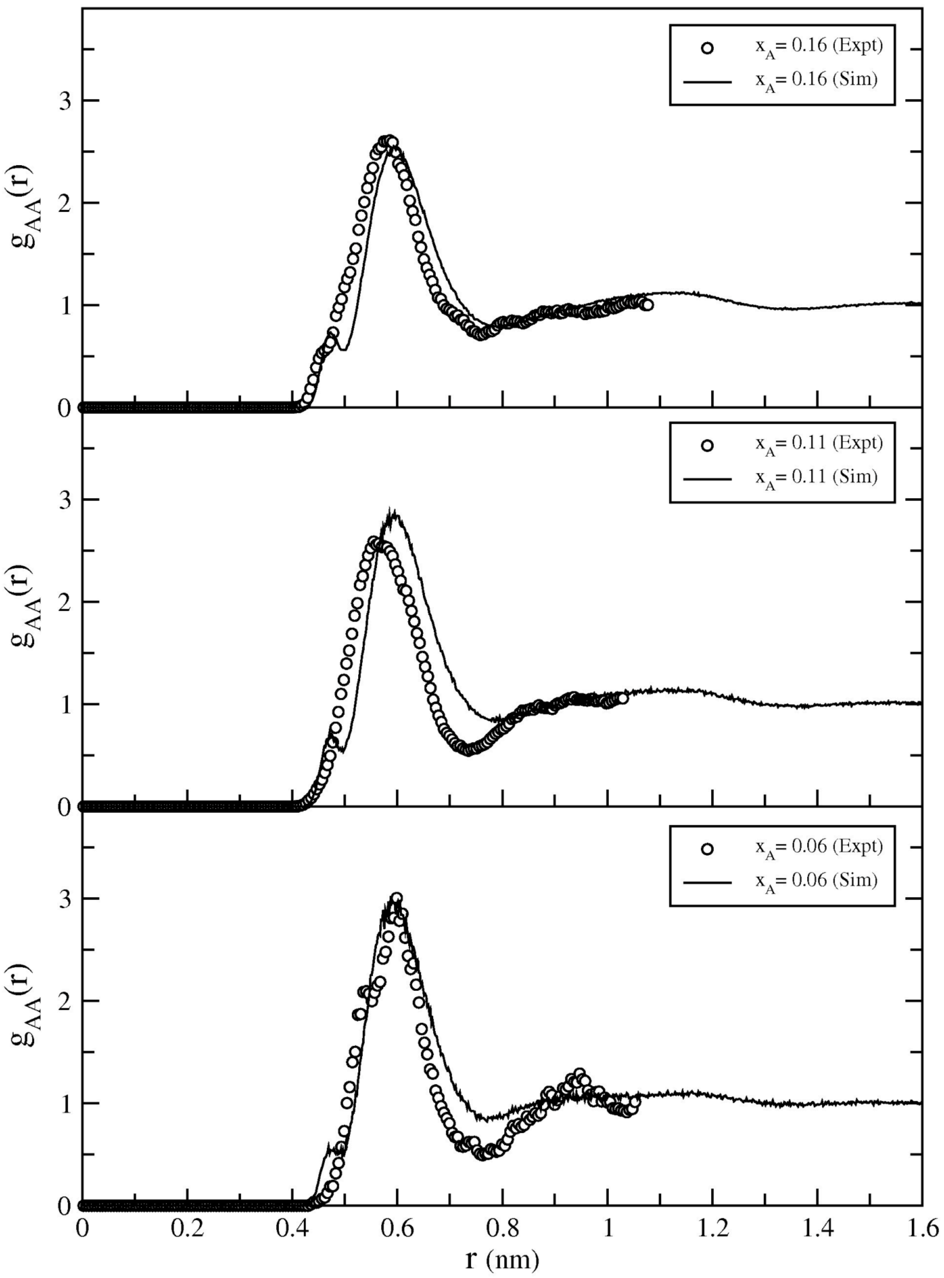


*Figure 3.12 TBA-TBA radial distribution functions at 0.16 (upper panel), 0.11 (middle panel), and 0.06 (lower panel) mole fractions. Full lines, simulation results (TBA partial atomic charge set V). RDFs (○) of Bowron et al. obtained from MC simulation based on neutron diffraction data.*[113]

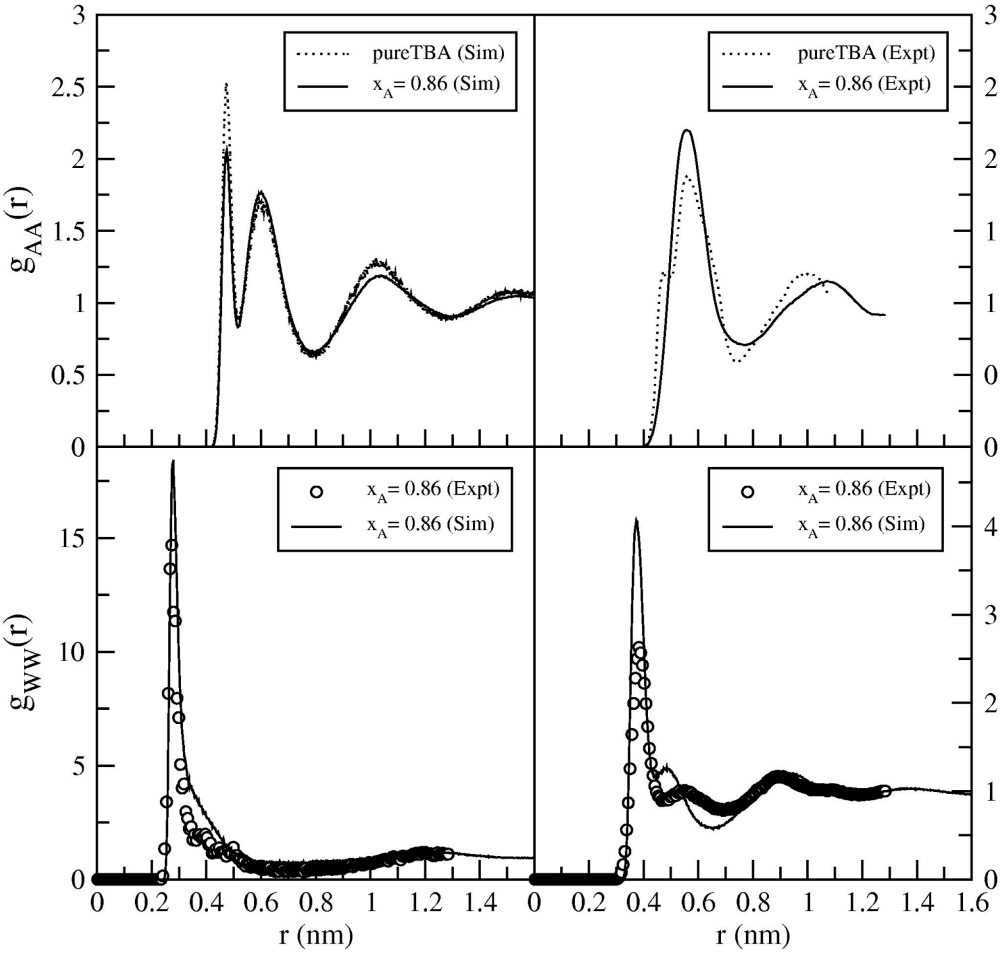


*Figure 3.13 Upper panel, alcohol-alcohol (upper panel, left) radial distribution function in the pure liquid (dashed line) and at 0.86 alcohol mole fraction (solid line) from simulation using TBA partial atomic charges from set V in table 3.3. Alcohol-alcohol (upper panel, right) radial distribution function in the pure liquid (dashed line) and at 0.86 alcohol mole fraction (solid line) from neutron diffraction studies of Bowron et al.*[114] *Lower panel, left, water-water radial distribution functions at 0.86 mole fraction alcohol from simulation (solid line) and neutron diffraction (○). Lower panel, right, alcohol-water radial distribution functions at 0.86 mole fraction alcohol from simulation (solid line) and neutron diffraction (○).*

The first peak at the 0.06 mole fraction alcohol solution (lower panel in fig. 3.12), which corresponds to the hydrophobic interaction between tertiary butyl groups, is observed at approximately the same distance (between 4.0 and 7.0 Å) and has equal peak height. The peak area corresponds to 3.3 alcohol centers (this work) and 2.8 alcohol centers (ND data), indicative that, on average, trimer solute aggregate configurations are favoured in aqueous solution of this concentration. The alcohol-alcohol correlations observed in our work are, however, somewhat weaker as can be seen from a slightly broader first peak and a weaker first minimum. Moreover, we observe evidence of alcohol-alcohol hydrogen bonding (the small shoulder at 4.7 Å) which is absent in the ND data at this concentration. For the 0.11 and 0.16 mole fraction alcohol solutions the qualitative agreement with the ND data remains good, although a slight shift of the first peak to larger distances is observed. The peak areas for the 0.11 mole fraction alcohol solution correspond to 4.6 (this work) and 4.4 (ND data) alcohol centers. For the 0.16 mole fraction alcohol solution these are 6.2 (this work) and 5.2 (ND data). At this mole fraction the longer range (>7 Å) features of the two alcohol-alcohol radial distribution functions are in good agreement, while at shorter distance evidence of alcohol-alcohol hydrogen bonding now becomes visible in the ND data too.

Figure 3.13 show the solution structure at 0.86 mole fraction alcohol solution. The upper panel shows the alcohol-alcohol center radial distribution functions obtained from our simulations (left) and the ND study of Bowron *et al.* (right)[114]. For comparison the alcohol-alcohol center radial distribution functions of pure liquid TBA are included. Clearly our simulations indicate much stronger alcohol-alcohol hydrogen bonding in the pure alcohol as well as in the 0.86 mole fraction alcohol solution. In pure TBA integrating the peak area up to 7.5 Å corresponds to 10.8 alcohol centers in agreement with the number found by ND. Integration up to 5.1 Å corresponds to 2 alcohol centers. In our simulations of pure TBA, therefore, a TBA molecule makes on average two hydrogen bonds with neighboring TBA molecules while the remaining 8.8 direct neighbors are not hydrogen bonded. Upon addition of water to pure liquid TBA, the ND results indicate loss of alcohol-alcohol hydrogen bonding and an enhancement of the packing of alcohol molecules with respect to themselves: the shoulder at 4.7 Å (H-bonding) in $g_{AA}(r)$ of pure TBA disappears (upper right panel), while the first peak at $\approx$5.6 Å increases to a larger value upon addition of water. A similar but much weaker evidence of this phenomenon is seen in our simulations in the upper left figure; the number of hydrogen bonding TBA neighbors decreases to 1.7, while the number of nonhydrogen bonded first neighbors increases to 9. The water molecular center's distribution function, $g_{WW}(r)$ (lower panel left), shows good agreement between the ND data and our simulations. This function has a striking form characterized by a steep peak at $\approx$2.8 Å and a clear region of reduced water-water contacts between 5 – 10 Å. Our simulation thus supports the observation reported by Bowron of water molecules being segregated at this length scale and forming pockets containing clusters of two or three water molecules in concentrated solution of aqueous *tert*-butanol. The alcohol center—water radial distribution function observed in this work is compared to the one obtained by ND in the lower right panel of fig. 3.13. Although both functions are in good agreement at distances larger than 7 Å, two functions at short distance are qualitative different in the concentrated TBA solution.

## 3.5 Conclusions

A nonpolarizable *tert*-butanol model for atomistic simulations of *tert*-butanol and water mixtures based on the KB integrals was developed. The model combined with the SPC water model, performs very well at all *tert*-butanol mole fractions. The extent of the correlation volume in which radial distribution functions differ significantly from unity is such that large system sizes (box lengths larger than 6 nm) are required. Over the full range of compositions, strong aggregation behavior is observed; while alcohol-alcohol aggregation is strongest in the water-rich regions ($x_A$<0.2), water-water aggregation persists over the full range of compositions. Short-ranged hydrophobic interactions between *tert*-butanol molecules are mainly responsible for aggregation of the alcohol molecules, but the relatively large size of the *tert*-butanol molecules causes correlations due to packing features being visible up to distances beyond 1.8 nm. Long-ranged correlations persist also between water-water pairs and contribute almost exclusively to the excess aggregation ($N_{WW}$) of water at alcohol mole fractions between 0.1 and 0.5, causing the water-water KB integrals to converge slowly. The long-ranged water-water correlations are not due to simple packing restrictions, but are caused by a significant strengthening of water-water hydrogen bonding induced by presence of the alcohol in solution. Because water prefers to hydrogen bond to itself, spatially extended water network structures occur in water-rich solutions. In alcohol-rich solutions ($x_A$>0.8), water molecules no longer form extended hydrogen bonded structures, however, long-ranged correlations remain. Water aggregates in small clusters of 3 – 4 molecules, which in the $g_{WW}(r)$ features as a high and broad first peak at distances $2.6<r<5$ Å, while a water-depleted region occurs for larger distances up to 10 Å, followed by tail oscillations characteristic to inter-cluster correlations. These structural features require, in addition to using large system sizes, long simulation times (10 ns), to equilibrate, in particular, the water structure and corresponding water-water KB integrals.

The *tert*-butanol model developed in this work correctly predicts the aggregation behavior of the aqueous mixtures and, as a consequence, reproduces the experimental solution activities shown in figs. 3.7, 3.8. As an alternative method, thermodynamic integration (TI) has been used to compute the solution activities. It, however, clearly showed that the KB-based approach performs superior. The statistical uncertainties in the TI calculations are $\approx \pm 1\text{–}2$ *kJ/mol*, while over the entire range of solution compositions the (molar concentration scale) activity coefficients change by 2.5 *kJ/mol* at maximum. This small change of energy is accurately reproduced by the KB method. Results obtained for other properties not included in the original parameterization reproduce the experimental trends with solution composition. The solution densities and static dielectric permittivities are, however, underestimated. The TBA model parameterized in this work will be useful in future studies of TBA/water mixtures addressing aspects of, *e.g.*, preferential (co)solvent interactions with dissolved solutes,[115] effects of TBA as a cosolvent on chemical reaction kinetics,[116] and salting-out mechanisms of this model amphiphile. In chapter 4, methane solvation thermodynamics in various aqueous TBA mole fractions using the current KB-derived force field are discussed.

# 4. Methane solvation in *tert*-butanol and water mixtures

## 4.1 Introduction

In chapter 3, we developed a new force field for *tert*-butanol and water mixtures based on Kirkwood-buff integrals. Tertiary butanol is an amphiphilic alcohol with the largest hydrophobic moiety remaining fully miscible with water. Our new force field for *tert*-butanol and water mixtures (further on in this thesis called as LV (Lee and Van der Vegt)-model) reproduced the solvation thermodynamics and aggregation behavior over various TBA mole fractions.

Binary solvent mixtures are used in many applications in chemistry and biology. They are used for example as reaction media to control the rates of chemical reactions, as selective solvents where chemically different groups of a solvated solute each preferentially interact with one of the solvent mixture components and as media for studying the stability of biomolecules. Mixtures of organic solvents (but also co-solutes such as inorganic salts, sugars, urea etc.) and water are of particular interest in biochemistry. Proteins are known to loose their stability in concentrated solutions of aqueous urea or guanidinium chloride[117] whereas they may be stabilized at low temperatures (cryopreservation)[118] in aqueous solutions of disaccharides (*e.g.,* trehalose) or diols (*e.g.,* ethylene glycol).[119, 120] The atomic-scale mechanisms and their relation with physical-chemical properties of the solvent system are in many cases still incompletely understood and are topic of ongoing discussion in the literature.[22, 33, 38, 121]

To investigate what happens at a molecular scale it would be desirable to use detailed atomistic computer simulations. The applicability of this method relies on accurate force fields being available to describe the system under study. The solute/cosolvent/water systems mentioned above are however examples where available biomolecular force fields, despite their careful parameterizations and successful applications, may lack the required accuracy. A sensible approach would be to validate the thermodynamic properties of the mixed solvent system (without solute) predicted by the force fields against known experimental quantities such as the heat of mixing, density, viscosity, etc. But not only the latter quantities, also the solvent species' activity (chemical potential) derivatives should be reproduced. These derivatives provide information on preferential interactions that cause 'excess coordination' of molecules of the same or the other kind.[3, 85] In aqueous/organic mixtures observed macro-miscibility often does not hold down to molecular length scales (micro-immiscibility) even in single-phase thermodynamically stable regions of the phase diagram away from any demixing line.[1, 20] It is not unlikely that this type of solvent phase nonideality is correlated with aspects of preferential solvation of solvated solute molecules: solutes can 'choose' to preferentially stay inside those regions of solvent richer in one type of solvent component.

Pioneered by Weerasinghe and Smith,[4-8] a series of Kirkwood-Buff (KB)-derived force fields have recently been published, which guarantee the appropriate description the excess coordination in aqueous solution mixtures. More recently, the solution osmotic pressure, which is also very sensitive to the excess coordination of solution components, has been considered.[122] On the basis of some of these force fields, Van der Vegt *et al.*[115, 123] studied the preferential solvation of small gaseous molecules as well as hydrophobic interactions[124] in terms of thermodynamic driving forces. Özal and Van der Vegt[125] performed a detailed study of solvent reorganization contributions to solvation entropies and enthalpies of methane in mixtures of dimethyl sulfoxide and water. Although the solvation entropies and enthalpies, more than the free energies, depend sensitively on the proper description of solvent-solvent, cosolvent-cosolvent and solvent-cosolvent interactions[125] a direct validation was not possible due to the lack of corresponding experimental data. Interestingly, however, the authors[125] showed that the solvent reorganization enthalpy from the $\left(\Delta E_{VV}\right)_P \simeq \left(T\left(\alpha_P / \kappa_T\right)\Delta V_S^*\right) = \left[\left(\frac{\partial U}{\partial V}\right)_T + P\right]\Delta V_S^*$ relation, where $\alpha_P$ is the isobaric thermal expansion coefficient of solvent, $\kappa_T$ is the solvent isothermal compressibility, $\Delta V_S^*$ is the solvation volume of solute S and $\left(\partial U / \partial V\right)_T$ is the solvent internal pressure, can be obtained indirectly from solvent equation of state data (*i.e.,* the solvent 'internal pressure').

In the chapter 3, a developed nonpolarizable force field for aqueous *tert*-butanol solutions was introduced. At 298 K, water and *tert*-butanol are miscible in all proportions, however, the solution is strongly non-ideal and exhibits significant aggregation of the amphiphilic alcohol molecules as well as the water molecules that try to keep their water-water hydrogen bonded network intact. The chapter 3 has shown that with one set of force field parameters the experimental Kirkwood-Buff integrals and the thermodynamic activity derivatives can both be reproduced as a function of solution composition. In chapter 4, this is extended to the study of methane solvation in *tert*-butanol and water mixtures. By comparison with experimental methane solubility data available for this system, including the methane solvation entropy and enthalpy, and comparison with alternative force fields, this work aims to provide a further validation of the *tert*-butanol-water KB-derived force field. Moreover, it gives some insights in relations between solvation free energies, entropies and enthalpies and aspects of excess coordination of solvent components.

## 4.2. Computational details

### 4.2.1 *tert*-butanol, water and methane models

All parameters for LV force fields in *tert*-butanol and water mixtures are same in chapter 3 (see the table 3.2). We also used the GROMOS[89] and OPLS[81] *tert*-butanol models described in the chapter 3.[126] The GROMOS TBA-model was combined with the SPC water model, while the OPLS TBA-model was combined with the SPC and TIP4P[97] water models. In combination with the SPC water model, the latter two force fields produce too large alcohol-alcohol and water-water aggregation (chapter 3).[126]

Methane was modeled as a united atom LJ-particle with parameters $\sigma$= 0.373 nm and $\varepsilon$= 1.277 *kJ/mol*. Geometric mean mixing rules such as $C_6(i,j) = C_6(i,i)^{1/2} C_6(i,i)^{1/2}$, $C_{12}(i,j) = C_{12}(i,i)^{1/2} C_{12}(i,i)^{1/2}$ were used to describe LJ interactions between chemically different atom types and groups.

### 4.2.2 Simulation details

For this chapter, all simulations were performed using the GROMACS package,[127] which is the fastest MD simulation software, whereas the previous work in chapter 3 was simulated using GROMOS96 package.

*Tert*-butanol and water mixtures were studied at the TBA mole fractions and summarized in table 4.1. Details on system sizes and simulation times are also given in table 4.1. Geometries of all molecules were kept rigid by applying constraints to the interatomic distances within the molecules, using the SHAKE algorithm[75] with a relative geometric tolerance of $10^{-4}$. A twin-range cutoff scheme with 0.8 and 1.4 nm cutoff radii was applied. The nonbonded interactions in the range between these radii were updated every fifth time step. The equations of motion were integrated using the leap-frog algorithm using a time step of 2 fs. A reaction field approximation was used to account for truncation of electrostatic forces beyond the long range cutoff (1.4 nm). The reaction field relative dielectric permittivities are listed in table 4.1. Constant pressure (1 atm) and temperature simulations were performed using the Nose-Hoover thermostat[71, 128] and Rahman-Parrinello barostat[70, 73] with coupling times $\tau_T$ = 1.5 ps and $\tau_P$ = 2.5 ps. Simulations were performed at three temperatures: 278 K, 298 K, and 318 K.

*Table 4.1 Summary of MD simulations of tert-butanol/water mixtures. The systems were also studied including ten methane solutes. V* is the volume of the pure solvent mixtures.*

| $x_{TBA}$ | $N_{TBA}$ | $N_{H2O}$ | $T_{MD}$(ns) | $\varepsilon_{RF}$ | $V$ (nm$^3$) | $V$*(nm$^3$) | $T$ (K) |
|---|---|---|---|---|---|---|---|
| | | | tert-butanol(LV-model)/water(SPC-model) | | | | |
| 0 | 0 | 1000 | 100 | 78.0 | 31.31 | 30.68 | 298 |
| 0.04 | 74 | 1776 | 150 | 64.8 | 65.83 | 65.29 | 298 |
| 0.06 | 105 | 1645 | 150 | 64.0 | 66.68 | 65.26 | 298 |
| 0.10 | 168 | 1512 | 100 | 50.7 | 72.48 | 71.73 | 298 |
| 0.20 | 275 | 1100 | 100 | 36.1 | 77.24 | 76.24 | 298 |
| 0.30 | 336 | 784 | 100 | 30.0 | 77.72 | 76.70 | 298 |
| 0.50 | 432 | 432 | 100 | 18.0 | 83.19 | 82.16 | 298 |
| 0.70 | 470 | 200 | 100 | 12.7 | 83.08 | 81.97 | 298 |
| 0.90 | 500 | 55 | 100 | 11.8 | 84.36 | 83.04 | 298 |
| 1.0 | 500 | 0 | 100 | 11.8 | 83.10 | 82.13 | 298 |
| | | | tert-butanol (GROMOS-model)/water(SPC-model) | | | | |
| 0.06 | 105 | 1645 | 100 | 64.0 | 67.61 | 66.69 | 298 |
| 0.1 | 168 | 1512 | 100 | 50.7 | 73.76 | 72.82 | 298 |
| 0.2 | 275 | 1100 | 100 | 36.1 | 78.97 | 77.96 | 298 |
| | | | tert-butanol (OPLS-model)/water(SPC-model) | | | | |
| 0.06 | 105 | 1645 | 100 | 64.0 | 67.46 | 66.56 | 298 |
| 0.10 | 168 | 1512 | 100 | 50.7 | 73.54 | 72.71 | 298 |
| 0.20 | 275 | 1100 | 100 | 36.1 | 78.24 | 77.51 | 298 |
| | | | tert-butanol (OPLS-model)/water(TIP4P-model) | | | | |
| 0.06 | 105 | 1645 | 100 | 64.0 | 66.74 | 65.76 | 298 |
| 0.1 | 168 | 1512 | 100 | 50.7 | 72.89 | 71.78 | 298 |
| 0.2 | 275 | 1100 | 100 | 36.1 | 78.03 | 76.83 | 298 |

### 4.2.3 Solvation free energies, enthalpies and entropies

The solvation free energies are calculated using the Widom test particle insertion method (see the chapter 2.3.2).[76] The solvation free energy $\Delta G$ is obtained by evaluating the expression,[115]

$$\Delta G = -RT \ln\left[ \left\langle V\, e^{-\psi/RT} \right\rangle_{NPT} / \left\langle V \right\rangle_{NPT} \right] \tag{4.1}$$

where $\psi$ denotes the methane binding energy with the solvent, $R$ denotes the gas constant, $T$ the temperature on the Kelvin scale and $V$ the volume. The methane binding energy is evaluated at 125000 random locations in each solvent configuration stored every 1 ps by summing over all methane-solvent LJ interactions. The angular brackets $\left\langle \cdots \right\rangle_{NPT}$ denote an averaging over constant pressure and temperature configurations of the solvent. In the simulation the averaging is performed over 100 ns simulation trajectories (table 4.1). The solvation entropy $\Delta S = -\left( \partial \Delta G / \partial T \right)_P$ is calculated using a finite-difference assumption,

$$\Delta S(T) = -\frac{\Delta G(T+dT) - \Delta G(T-dT)}{2dT} \tag{4.2}$$

where $T = 298\ K$ and $dT = 20\ K$ in our calculations. The solvation enthalpy (298 *K*) is obtained from the relation

$$\Delta H = \Delta G + T\Delta S \tag{4.3}$$

The statistical uncertainty of $\Delta G$ (eq. 4.1) is smaller than 0.05 *kJ/mol* at all mixture compositions. The statistical uncertainty of $T\Delta S$ (eq. 4.2) and $\Delta H$ (eq. 4.3) is below 0.8 *kJ/mol* at all mixture compositions. Finite difference assumption for an entropy change (eq. 4.2) is only valid for if the solvation entropy is constant over the selected temperature range. Because the solvation heat capacities may however be large,[129] we decided to also determine $\Delta H$ and $T\Delta S$ with a direct method. As an independent method, we calculated $\Delta H$ by subtracting the total potential energy of the solvent without solute from the potential energy of the solvated system ('direct method'). This method requires the simulation of two systems, in which the potential energies were averaged over 100 ns trajectories. A contribution $RT^2\alpha_P$ (with $\alpha_P$ the isobaric thermal expansion coefficient of the solvent mixture) needs to be subtracted from the enthalpy difference obtained by the direct method in order to remove the liberation contribution.[85, 125]

Thus, $\Delta H(solvation) = \Delta H(direct) - RT^2\alpha_P$. The correction term $RT^2\alpha_P$ is non-negligible and amounts to 0.6-0.8 *kJ/mol* for the systems studied in this work.

## 4.2.4 Solubility data

We obtained the experimental methane solvation free energies, enthalpies and entropies from the solubility data reported by Wang *et al.*[120] These data are reported in mole fraction methane ($x_{CH_4}$) in the solution at equilibrium with a methane pressure *P* at temperature *T*.[120] We use the Ben-Naim definition of the solvation process and the corresponding free energy change[130]

$$\Delta G = -RT\ln\left(\rho_{CH_4} / \rho^{ig}_{CH_4}\right)_{eq} \tag{4.4}$$

where $\rho_{CH_4}$ and $\rho^{ig}_{CH_4}$ are the methane molar densities in the solution and ideal gas phase, respectively. Both phases are at equilibrium as denoted by the subscript *eq*. In terms of the variables $x_{CH_4}$, *P* and *T* used in ref. [120] the solvation free energy becomes (assuming $CH_4$ is present at infinite dilution in the binary solvent)

$$\begin{aligned}\Delta G &= -RT\ln x_{CH_4} - RT\ln\left(RT\rho_{solv} / P\right)\\ &= RT\ln K_H - RT\ln\left(RT\rho_{solv}\right)\end{aligned} \tag{4.5}$$

where $\rho_{solv}$ is the molar density of the pure solvent phase (*i.e.,* the alcohol/water mixture) and $K_H$ the Henry coefficient. Experimental information on $\rho_{solv}$ was obtained from ref [104]. We note that $K_H$ is usually expressed in units of atmospheres in which case $RT\rho_{solv}$ (second line in eq. 4.5) should also be expressed in these units. Wang *et al.*[120] reported methane solubility data at 283.15 *K*, 288.15 *K*, 293.15 *K*, and 298.15 *K*. By applying the finite difference method (eq. 4.2) to the free energies obtained from eq 5 at 288.15 *K* and 298.15 *K* we obtain the experimental solvation entropies at 293.15 *K*. The experimental solvation enthalpies at 293.15 *K* are next obtained from eq. 4.3 using the solvation entropies together with the experimental solvation free energies at 293.15 *K*. Note that the so-derived solvation enthalpies and entropies all apply at a temperature 5 *K* below the temperature (298 *K*) used in our simulations. The discrepancy due to this difference in temperature we estimate based on the experimental heat capacity change $\Delta c_p$ (142 *J/mol K*) of solvating methane in water at 298 *K*.[131] Assuming this value remains constant in the 5 *K* temperature range results in a 0.7 *kJ/mol* difference in $\Delta H$ and $T\Delta S$ (*i.e.,* the values being 0.7 *kJ/mol* higher at the higher temperature).

## 4.3 Results and discussion

### 4.3.1 Free energy of methane solvation

Figure 4.1 shows the methane solvation free energies at 298 *K* and 1 atm presented versus the alcohol mole fraction $x_{TBA}$ of the solution. Transfer of the methane solute from pure water to pure *tert*-butanol causes the free energy to decrease with approximately 8 *kJ/mol* corresponding to a 25-fold increase of the methane solubility at 298 *K*. The shape of the curve ($\partial \Delta G / \partial x_{TBA} < 0$) indicates[85, 115] that methane in TBA/water solution preferentially interacts with the alcohol molecules. Four TBA/water force field combinations have been used to calculate the $\Delta G$ data in figure 4.1: LV/SPC, GROMOS/SPC, OPLS/SPC and OPLS/TIP4P. The best agreement with the experimental data (squares) is observed for the LV/SPC model; the calculated data match the experimental values for $x_{TBA}$ up to 0.06; at larger alcohol mole fractions the free energies are slightly underestimates by the model. For $x_{TBA}$ up to 0.2, the GROMOS/SPC and OPLS/SPC models result in almost identical methane solvation free energies, which are however systematically lower than the experimental and LV/SPC values. The $\Delta G$ values obtained based on the OPLS/TIP4P solvent model are slightly improved over those values obtained based on the corresponding OPLS/SPC model but are still systematically below the experimental and LV/SPC data.

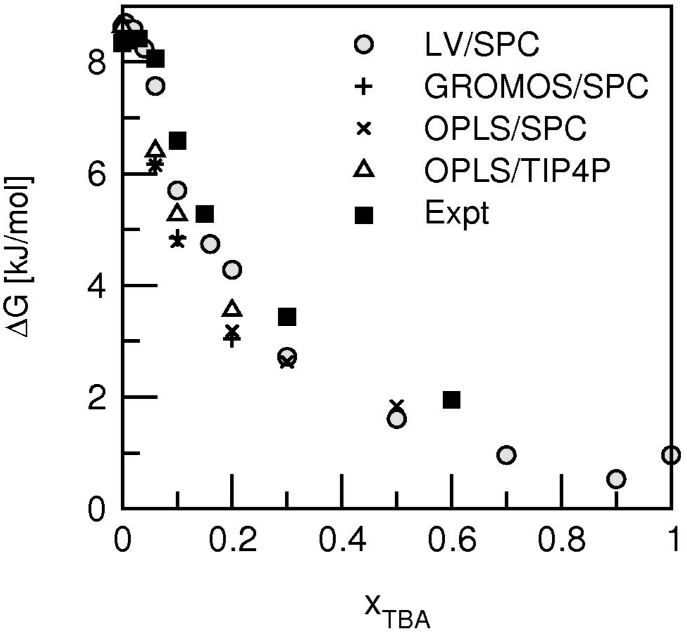


*Figure 4.1 (a) Solvation free energy change over the various TBA/water concentration mixtures; experimental data (black square), LV/SPC (grey circle). GROMOS/SPC (plus), OPLS/SPC (cross), OPLS /TIP4P (triangle). All points have* $\pm 0.01\ kJ/mol$ *error.*

## 4.3.2 Preferential methane solvation

Figure 4.2 shows methane-solvent radial distribution functions g(r). The radial distribution functions (RDFs) were obtained from simulations including ten methane solutes in the systems summarized in table 4.1. The left panel in fig. 4.2 shows the methane-TBA (methyl) RDF at TBA mole fractions 0.06, 0.1 and 0.2. The right panel shows the methane-water (oxygen) RDFs at the same solution compositions. Methane preferentially interacts with TBA-methyl groups. All RDFs (left panel) show a first peak at 0.4 nm followed by a second and third peak superimposed on an exponential type decay of the RDF toward unity. The limiting value (RDF=1) is reached at approximately 1.5 nm with the LV/SPC model, while with all other models box sizes are too small to observe this limit. For the methane-water RDFs (right panel) the physical picture is reversed: a first peak is observed at 0.36 nm, followed by a water depletion region (RDF<1) up to 1.5 nm for the LV/SPC model; with all other models the limiting value RDF=1 is not reached for distances up to 2 nm. First shell coordination numbers obtained by integrating the RDFs up to the first minimum are summarized in table 4.2. To quantify the excess amount of TBA and water vicinal to the methane solute, we calculated methane-solvent excess coordination numbers $N_j^{ex}$ defined as,

$$N_j^{ex} = \rho_j \int_0^R \left[ g_{Sj}(r) - 1 \right] 4\pi r^2 dr \tag{4.6}$$

In eq. 4.6, $\rho_j$ denotes the solvent component number density, $g_{Sj}(r)$ the solute ($CH_4$)-solvent RDF, and $R$ is an integration cutoff distance. Figures 4.3a and 4.3b show the methane-TBA and methane-water excess coordination numbers ($R$ = 1 nm), respectively. The $CH_4$-TBA excess coordination number (LV/SPC model) goes through a maximum at $x_{TBA}$=0.1 while the $CH_4$-water excess coordination number goes through a minimum at this solvent composition. With the GROMOS and OPLS models excess coordination numbers are found that exceed those obtained with the LV/SPC model by 50-60%.

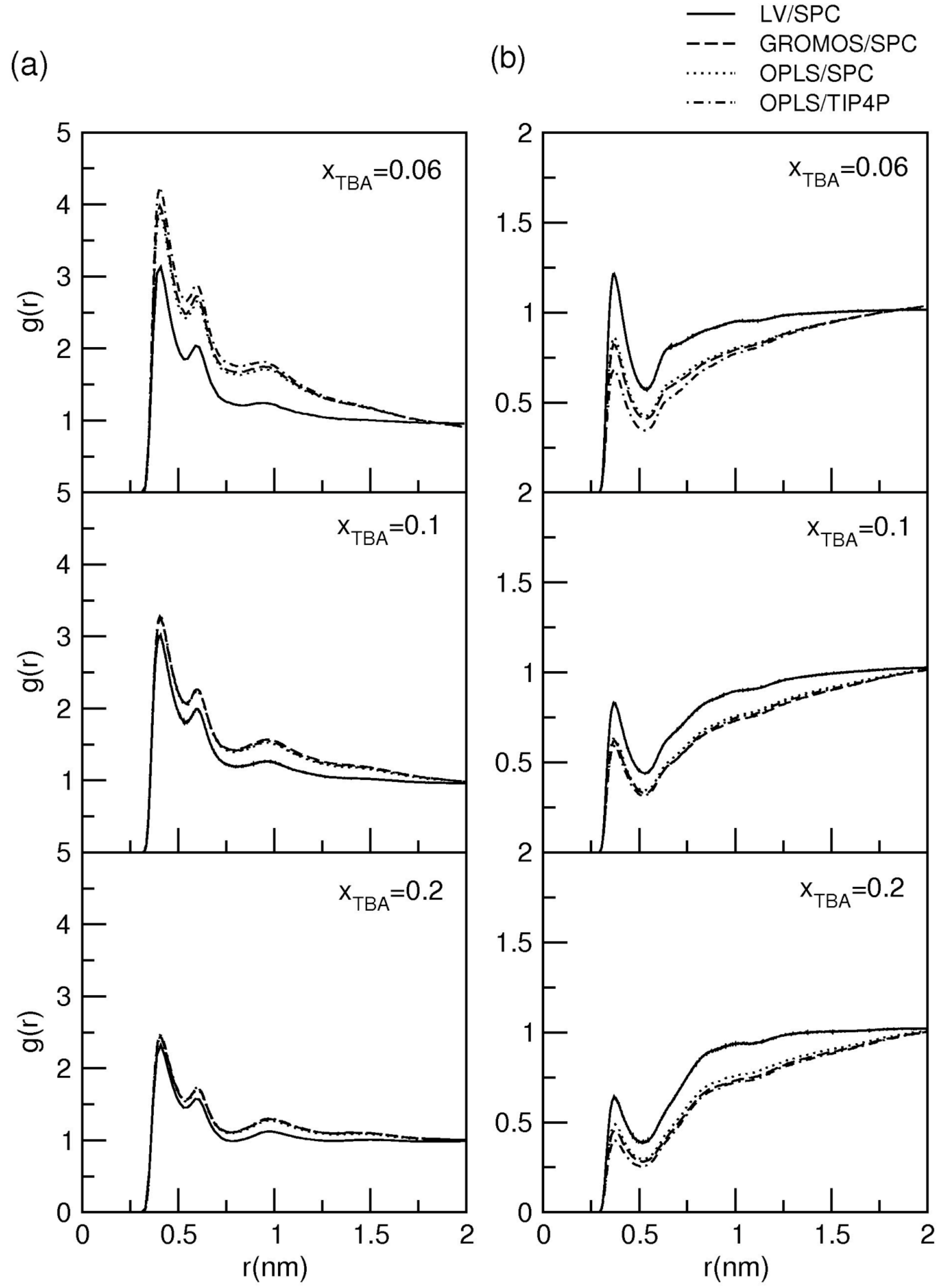


*Figure 4.2 The radial distribution functions versus the distance r at 0.06, 0.1, 0.2 TBA mole fractions; (a) methane-TBA(methyl), (b) methane-water(oxygen)*

*Table 4.2 First shell coordination numbers* $N_j$ *and excess first shell coordination numbers* $N_j^{ex}$ *for methane-TBA(methyl) and methane-water(ow) with first shell distance, 0.53nm.*

| $x_{TBA}$ | Methane-TBA 1st shell $N_j$ ($N_j^{ex}$) | Methane-WATER 1st shell $N_j$ ($N_j^{ex}$) |
|---|---|---|
| | tert-butanol(LV-model)/water(SPC-model) | |
| 0.04 | 3.3(1.0) | 13.7(-3.3) |
| 0.06 | 5.6(2.3) | 9.8(-5.7) |
| 0.1 | 7.9(3.0) | 5.9(-7.2) |
| 0.2 | 9.5(2.1) | 3.3(-5.7) |
| 0.3 | 10.1(1.1) | 2.2(-4.1) |
| 0.5 | 10.7(-0.2) | 1.2(-2.1) |
| 0.7 | 11.0(-0.9) | 0.6(-0.9) |
| 0.9 | 11.1(-1.3) | 0.2(-0.2) |
| | tert-butanol(GROMOS-model)/water(SPC-model) | |
| 0.06 | 7.0(3.8) | 7.2(-9.3) |
| 0.1 | 8.4(3.8) | 4.7(-9.2) |
| 0.2 | 9.7(2.6) | 2.5(-7.0) |
| | tert-butanol(OPLS-model)/water(SPC-model) | |
| 0.06 | 6.8(3.6) | 7.5(-9.0) |
| 0.1 | 8.4(3.8) | 4.8(-9.1) |
| 0.2 | 9.7(2.5) | 2.7(-6.9) |
| | tert-butanol(OPLS-model)/water(TIP4P-model) | |
| 0.06 | 7.5(4.3) | 6.1(-10.7) |
| 0.1 | 9.1(4.1) | 4.8(-10.1) |
| 0.2 | 10.6(2.9) | 2.4(-7.9) |

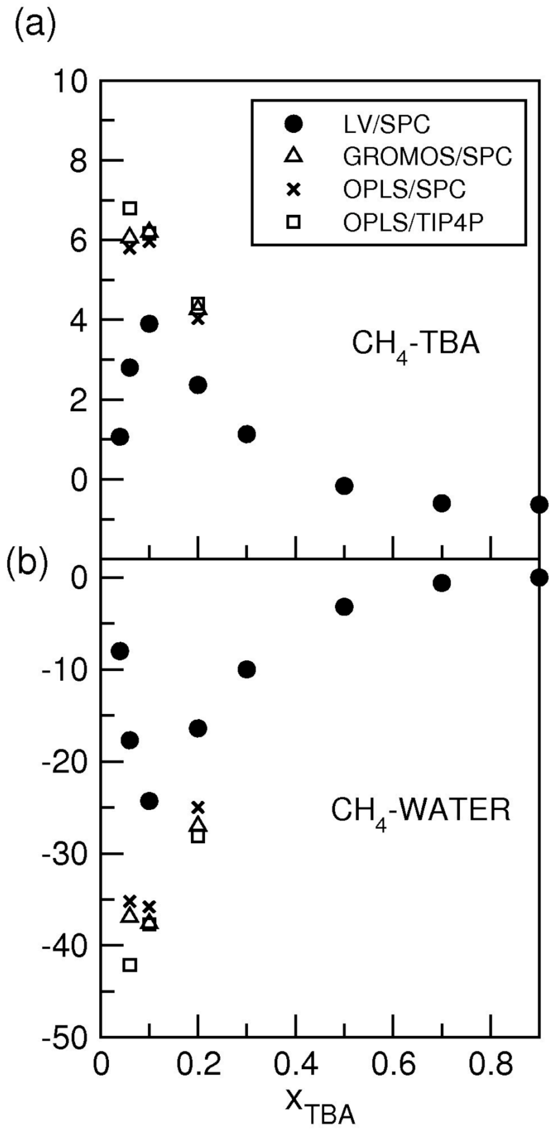


*Figure 4.3 (a) The excess coordination number (R=1nm) for methane and TBA in LV/SPC (black circle), GROMOS/SPC (triangle), OPLS/SPC (cross), and OPLS/TIP4P (square); (b) The excess coordination number (R=1nm) for methane and water (oxygen) in LV/SPC (black circle), GROMOS/SPC(triangle), OPLS/SPC (cross) and OPLS/TIP4P (square).*

### 4.3.3 Methane solvation enthalpy and entropy

Figures 4.4a and 4.4b show the methane solvation enthalpies and entropies versus the alcohol mole fraction of the solution. Interestingly, these curves are non-monotonic in contrast to the solvation free energy (fig. 4.1); *i.e.,* enthalpy and entropy changes are steeply increased in the very dilute alcohol solutions ($0 < x_{TBA} < 0.2$), then the enthalpy is decreased and entropy is fluctuated constantly beyond $x_{TBA}= 0.2$, whereas the solvation free energy is gradually decreased as increasing the alcohol mole fraction. Squares represent the experimental data obtained from solubility data (chapter 4.2.4). And methane solvation enthalpy and entropy in the LV/SPC model were obtained from two ways; *i.e.,* finite difference method (grey circle) and direct calculation (red circle) (see the chapter 4.2.3). Right panel in fig. 4.4 shows enthalpies and entropies obtained from direct calculation of methane solvation in GROMOS/SPC, OPLS/SPC, and OPLS/TIP4P models in the range of 0.06 - 0.2 TBA mole fractions. The experimental data indicate a rapid increase of $\Delta H$ with $x_{TBA}$ with a maximum obtained at $x_{TBA}=0.06$. For larger $x_{TBA}$ the methane solvation enthalpy decreases. The methane solvation entropy rapidly increases up to $x_{TBA} \approx 0.1$ and then remains practically constant. With the LV/SPC model, the experimental enthalpy curve is qualitatively reproduced. The maximum is predicted at the appropriate alcohol concentration and is almost quantitatively reproduced. At higher TBA mole fractions ($x_{TBA}>0.06$) the solvation enthalpies are underestimated while at higher alcohol dilution ($x_{TBA}<0.06$) the solvation enthalpies are overestimated (see also the table 4.3). The same observations apply for the solvation entropies. Note that in pure water ($x_{TBA}=0$), the methane solvation enthalpy and entropy are significantly overestimated. This is caused by a too large thermal expansion coefficient of the SPC water model resulting in a too large solvent reorganization energy.[122, 125]

Based on the GROMOS and OPLS models the solvation enthalpies and entropies are poorly reproduced; although the downward trend of the enthalpy agrees with the experimental data, the sign is wrong, indicating that a significant, positive enthalpy contribution is missing.

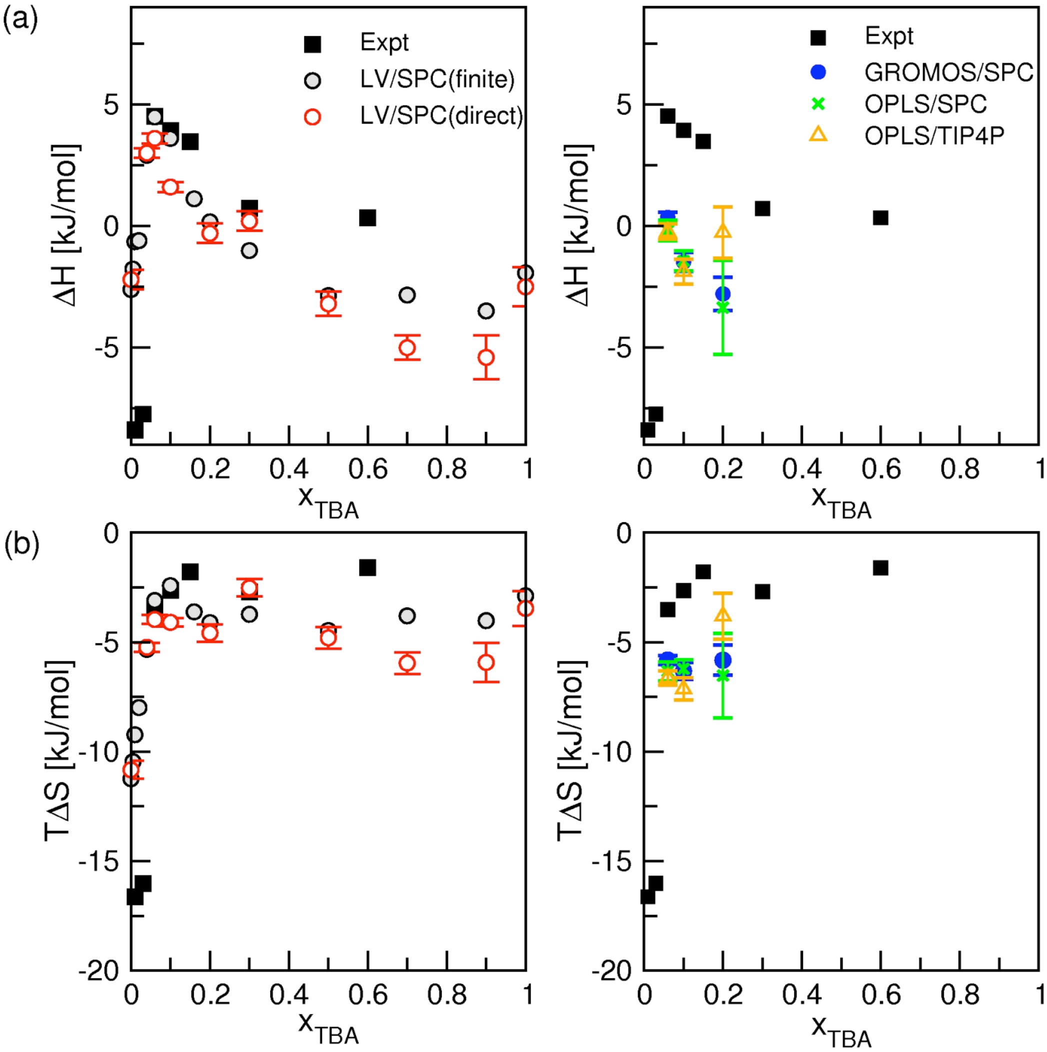


*Figure 4.4 Methane solvation enthalpies (a) and entropies (b) in experimental data (black square), LV/SPC (grey circle: finite difference; red circle: direct), GROMOS/SPC(blue circle: direct), OPLS/SPC (green cross: direct) and OPLS/TIP4P (yellow triangle: direct) with error bar over the TBA mole fractions.*

*Table 4.3 Thermodynamic data for methane solvation (298K, 1atm) in TBA and water mixtures. The experimental data (denoted with subscript exp) were obtained from ref* [120]. *The data calculated based on simulations (denoted with subscript sim) were obtained with the LV/SPC model using test-particle insertions and finite temperature differences (eqs. 4.1-4.3). Units are kJ/mol.*

| $x_{TBA}$ | $\Delta G_{\mathrm{exp}}$ | $\Delta G_{sim}$ | $\Delta H_{\mathrm{exp}}$ | $\Delta H_{sim}$ | $T\Delta S_{\mathrm{exp}}$ | $T\Delta S_{sim}$ |
|---|---|---|---|---|---|---|
| 0.0 | 8.3 | 8.6 | -11.5 | -2.6 | -19.8 | -11.2 |
| 0.01 | 8.4 | 8.6 | -8.4 | -0.7 | -16.6 | -9.2 |
| 0.03 | 8.4 |  | -7.7 |  | -16.0 |  |
| 0.04 |  | 8.2 |  | 2.9 |  | -5.3 |
| 0.06 | 8.1 | 7.6 | 4.6 | 4.5 | -3.5 | -3.1 |
| 0.1 | 6.6 | 5.7 | 3.9 | 3.6 | -2.7 | -2.4 |
| 0.15 | 5.3 | 4.7 | 3.5 | 1.1 | -1.8 | -3.6 |
| 0.2 |  | 4.3 |  | 0.2 |  | -4.1 |
| 0.3 | 3.4 | 2.7 | 0.7 | -1.0 | -2.7 | -3.7 |
| 0.5 |  | 1.6 |  | -2.9 |  | -4.5 |
| 0.6 | 2.0 |  | 0.3 |  | -1.6 |  |
| 0.7 |  | 1.0 |  | -2.8 |  | -3.8 |
| 0.9 |  | 0.5 |  | -3.5 |  | -4.0 |
| 1 |  | 1.0 |  | -1.9 |  | -2.9 |

*Table 4.4 Solvation enthalpy contributions of LV/SPC-model.* $\Delta H(direct)$ *obtained from taking direct energy differences with correction for removing the liberation contribution term,* $RT^2\alpha_P$*, and* $\Delta H(finite)$ *obtained from temperature differences (see the section 4.2.3).* $\Delta H_{VV}$ *and* $\Delta H_{UV}$ *are solvent reorganization energy and solute-solvent binding energy in TBA/water (298K, 1 atm) obtained with LV-model. Units are kJ/mol. Note that* $\Delta H_{UV}$ *depends only weakly on the TBA/water solvent composition while* $\Delta H_{VV}$ *determines the compositional dependence of* $\Delta H$.

| $x_{TBA}$ | $\Delta H(direct)$ | $\Delta H(finite)$ | $\Delta H_{VV}$ | $\Delta H_{UV}$ | $RT^2\alpha_P$ |
|---|---|---|---|---|---|
| 0.0 | -2.2±0.4 | -2.6 | 11.2 | -13.8 | 0.54 |
| 0.01 | | -0.7 | 12.7 | -13.4 | |
| 0.04 | 3.0±0.2 | 2.9 | 16.2 | -13.3 | 0.71 |
| 0.06 | 3.6±0.2 | 4.5 | 17.7 | -13.2 | 0.76 |
| 0.1 | 1.6±0.2 | 3.6 | 16.6 | -13.0 | 0.83 |
| 0.15 | | 1.1 | 14.1 | -13.0 | |
| 0.2 | -0.3±0.4 | 0.2 | 13.1 | -12.9 | 0.86 |
| 0.3 | 0.2±0.4 | -1.0 | 12.0 | -13.0 | 0.88 |
| 0.5 | -3.2±0.5 | -2.9 | 10.2 | -13.1 | 0.86 |
| 0.7 | -5.0±0.5 | -2.8 | 10.2 | -13.0 | 0.87 |
| 0.9 | -5.4±0.9 | -3.5 | 9.5 | -13.0 | 0.86 |
| 1 | -2.5±0.8 | -1.9 | 11.0 | -12.9 | 0.87 |

To better understand the enthalpy change, we decompose this quantity in solute-solvent ($\Delta H_{UV}$) and solvent-solvent enthalpy ($\Delta H_{VV}$) changes, *i.e.*,

$$\Delta H = \Delta H_{UV} + \Delta H_{VV} \tag{4.7}$$

The solute-solvent enthalpy change $\Delta H_{UV}$ is the sum over all solute-solvent pair interactions and was obtained based on 100 ns simulations of the systems summarized in table 4.1 including ten methane solutes. The enthalpy associated with changes of solvent-

solvent interactions $\Delta H_{VV}$ (the solvent reorganization enthalpy) was obtained by subtracting the total potential energy of the solvent boxes in table 4.1 (sampled during 100 ns) from the 'solvent' potential energy of the corresponding boxes containing the additional ten methane molecules. The 'solvent' potential energy of the boxes containing the methane solutes includes only the contributions of solvent-solvent interactions (alcohol-alcohol, alcohol-water and water-water). Table 4.4 and figure 4.5 shows the contributions of solute-solvent enthalpy and solvent reorganization to the solvation enthalpy with solvation enthalpy changes (LV-model). Inspection of these data shows that the shape of the enthalpy curve (fig. 4.5) is entirely determined by the solvent reorganization enthalpy $\Delta H_{VV}$, the solute-solvent enthalpy $\Delta H_{UV}$ changes only little with the solution composition.

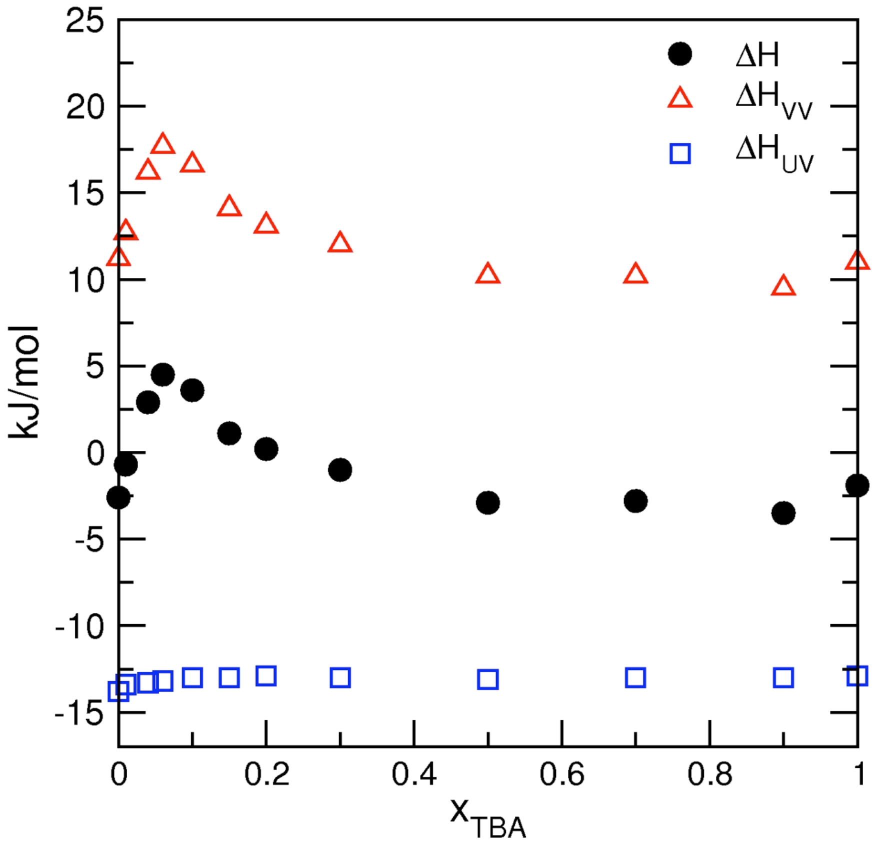


*Figure 4.5 Methane solvation enthalpies $\Delta H$ (black circles) and contributions of solute-solvent interactions $\Delta H_{UV}$ (blue squares) and solvent reorganization enthalpy $\Delta H_{VV}$ (red triangles) for LV/SPC model.*

Figure 4.6 shows snapshots of the methane (grey) solvation in the TBA and water mixtures. The snapshots include the solvents in the first solvation shell at the 0.1 TBA mole fraction. Left snapshot shows methane solvation in the LV/SPC model which is surrounded by TBA and water, whereas right snapshot shows methane first shell solvation in the GROMOS/SPC model which is solvated in an alcohol cluster.

Table 4.5 shows the contribution of water-water, TBA-TBA, and TBA-water interactions to the reorganization energy $\Delta H_{VV}$. The largest contribution to the enthalpy maximum observed in fig. 4.5, results from changes in the TBA-water interaction.

As we have seen in fig. 4.6 and table 4.5, alcohol clusters in water-rich composition region remain sufficiently hydrated to keep the system miscible. Water molecules are however expelled from the alcohol clusters by the introduction of methane.

*Table 4.5 Contributions of water-water, TBA-TBA and TBA-water interactions to the solvent reorganization energy* $\Delta H_{VV}$ *at 298 K and 1 atm. The contribution were obtained by subtracting the potential energy components of a pure solvent box from the respective energy components of boxes including ten methane solutes. Averaging was performed over 100 ns simulations (table 4.1). The error bars were obtained by block averaging. Units are kJ/mol.*

| $x_{TBA}$ | $\Delta H_{H_2O-H_2O}$ | $\Delta H_{TBA-TBA}$ | $\Delta H_{TBA-H_2O}$ |
|---|---|---|---|
| 0.04 | 80 ± 3 | -16 ± 9 | 114 ± 17 |
| 0.06 | -5 ± 13 | -35 ± 8 | 219 ± 19 |
| 0.1 | -41 ±22 | -11 ± 13 | 210 ± 31 |
| 0.2 | -54 ± 42 | 14 ± 29 | 189 ± 64 |
| 0.3 | 13 ± 42 | 91 ± 35 | 31 ± 72 |
| 0.5 | -63 ± 36 | 46 ± 33 | 151 ± 63 |
| 0.7 | -21 ± 37 | 86 ± 36 | 63 ± 68 |
| 0.9 | 0 ± 8 | 131 ± 14 | 2 ± 18 |

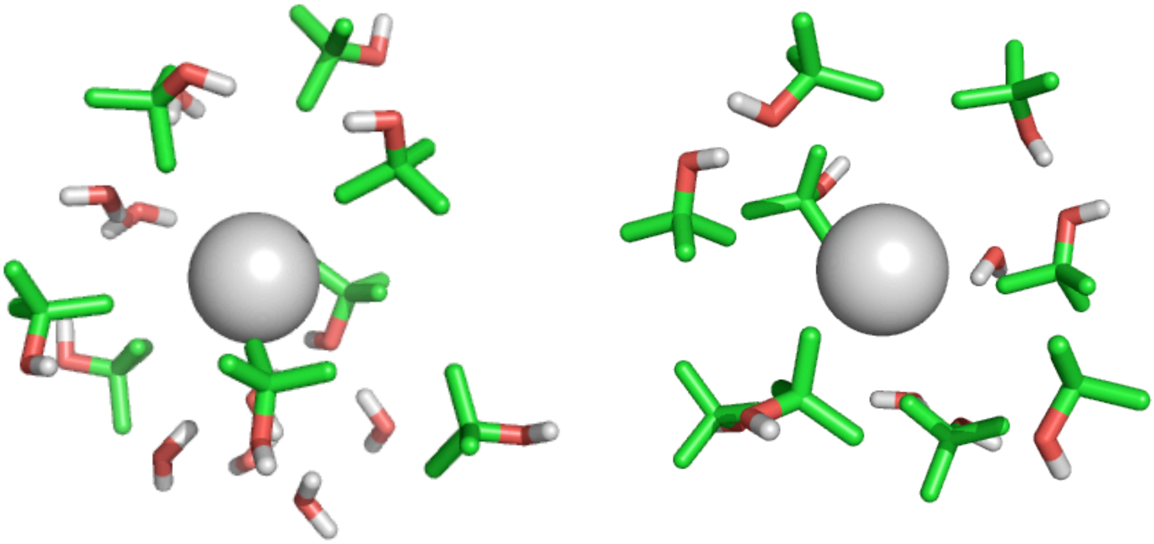


*Figure 4.6 Methane solvation snapshot for the LV/SPC (left) and GROMOS/SPC (right) model. Grey is methane, red is oxygen, and green is carbon and methyl group. Solvent configurations are shown around 6 Å of methane at the 0.1 TBA mole fraction.*

As shown in table 4.5, this lead to a significant enthalpy increase due to disruption of alcohol-water cohesive interactions. The endothermic heat of methane dissolution around $x_A = 0.1$ (fig. 4.4a) is therefore explained by the atomic-scale picture where methane sticks to alcohol clusters and expels hydration water. Water expulsion is less relevant with the GROMOS and OPLS models, since these models produce too strong alcohol aggregation with alcohol clusters being significantly less hydrated (fig. 4.6)

The solvation entropy (fig. 4.4b) follows the trend of the enthalpy (fig. 4.4a), however there is no clear maximum. Again the LV-model reproduces the experimental entropy curve better than do the other models. We note here that the solvent reorganization enthalpy occurs as a contribution $\Delta H_{vv} / T$ in the entropy,[115, 123-125, 132] hence it is not surprising that the discrepancies in $\Delta H$, observed with the GROMOS and OPLS models, occur in the entropy changes too.

### 4.3.4 Discussion

The data in figs. 4.1 and 4.4 clearly indicate that the LV-model performs better in comparison with the other force field models. Moreover, the procedure of validating the solvation enthalpies and entropies (fig. 4.4) seems to provide good insight in the accuracy of the models. The discrepancies observed in these quantities, are larger than in the solvation free energy. This is not surprising if one recognizes that the solvent reorganization enthalpy is always exactly enthalpy-entropy compensating and therefore does not influence the free energy change.[123, 132] The discrepancies between the experimental thermodynamic data and the calculations based on the GROMOS and OPLS models directly correlate with the microscopic solvent structure. The LV-model has been parameterized to reproduce the experimental Kirkwood-Buff integrals which describe the extent of alcohol-alcohol, alcohol-water and water-water aggregation in the solvent mixture (chapter 3). We re-analyzed the solvent-solvent aggregation in the presence of ten methane solutes but did not observe any significant changes in solvent aggregation. Since the thermodynamic solubility data are well - reproduced, we conclude that the solute ‘sees’ the appropriate solvent environment with the Kirkwood-Buff-derived LV-model. With GROMOS and OPLS the extent of alcohol-alcohol and water-water aggregation is significantly larger.[126] As suggested by the data in table 4.2 and fig. 4.3, with these models methane is solvated by larger (but non-physical) alcohol clusters. This leads to lower solvation free energies (because the methane solvation free energy in pure *tert*-butanol is lower than in pure water) and lower solvation enthalpies (because positive contributions of alcohol-water interactions to $\Delta H$ are underestimated). The solvation free energy is influenced mostly by the solute interaction with its solvation shells and in that sense it depends sensitively on a proper description of the preferential solvation by solvent components. The solvation enthalpy and entropy are influenced mostly by the changes in interactions between the solvent molecules and in that sense probe much better the quality of the force field describing the aqueous/organic solvent mixture.

## 4.4 Conclusions

We performed atomistic computer simulation studies of methane solvated in mixtures of *tert*-butanol and water. In addition to aspects of preferential methane solvation in terms of radial distribution functions and excess coordination of solvent components, we calculated the methane solvation free energy, enthalpy and entropy based on four force field models available in the literature. Since preferential solvation of solutes by cosolvent or water molecules is likely to be coupled to aspects of cosolvent and water clustering in the solvent mixture, we used the our LV/SPC force field, which was parameterized to reproduce the Kirkwood-Buff integrals of the solvent mixture in the chapter 3. We also compared the GROMOS and OPLS force fields for *tert*-butanol where the former was combined with the SPC water model and the latter with the TIP4P and SPC water models. We find that the enthalpy and entropy changes of methane insertion are very sensitive to changes of solvent-solvent interactions and therefore accurately probe the quality of the force fields for this solvent mixture. In particular, at dilute alcohol solution, methane sticks to hydrated alcohol clusters and expels alcohol hydration water. This process is responsible for an endothermic heat of methane solvation (as opposed to exothermic solvation in pure water) and an increase of the solvation entropy. Our LV-model provides the good description of the methane solvation free energies, enthalpies and entropies in aqueous *tert*-butanol solution.

# 5. Enthalpy–entropy compensation in the free energy of hydrophobic interactions

## 5.1 Introduction

In the chapter 4, preferential methane solvation in alcohol and water mixtures was discussed in terms of the solvation free energy, enthalpy and entropy. To understand preferential solvation in alcohol and water mixtures, the energy contributions were considered by splitting the contributions of solute-solvent and solvent-solvent interactions.

In the case of methane insertion in TBA/water mixtures, the solvent-solvent energy changes significantly whereas the solute-solvent energy remains constant as the TBA mole fraction is changed. As we have seen in chapter 4, the solvent reorganization energy is a dominant contribution to the methane solvation. A decomposition of free energy into its energetic and entropic components provides useful insights in atomic-scale solvation mechanisms. Moreover, it is helpful in understanding solvation thermodynamics of hydrophobic interactions in liquid mixtures. In particular, it is important to know the exact enthalpy-entropy compensation in the free energy. In this chapter, this process is discussed in detail. We study the implicit contribution of the solvent-solvent reorganization, also understand exact enthalpy-entropy compensation in the free energy, and investigate how important this effect is in solvation processes.

For example, recently Özal and Van der Vegt *et al.* studied the methane solvation in terms of the free energy, enthalpy and entropy with two contributions of solute-solvent interaction and solvent reorganization interaction in various dimethyl sulfoxide and water concentrations.[125] They present that the entropy change of methane transfer from water to aqueous solutions of dimethyl sulfoxide is positive because dimethyl sulfoxide molecules, in preferentially solvating the nonpolar solute, release some of their hydration waters to the bulk. However it is shown that this solvent reorganization process is enthalpy-entropy compensating in the free energy, therefore the entropy change cannot explain the preferential interaction of methane with dimethyl sulfoxide molecules. Van der Vegt *et al.* also presented the transfer of noble gases, methane, ethane, propane, *n*-butane, *neo*-pentane from pure water to various cosolvent (NaCl, urea, dimethyl sulfoxide, and acetone) and water mixtures. They reported that the enthalpy-entropy compensation becomes larger in the order NaCl, urea, dimethyl sulfoxide and acetone.[123] Aliphatic hydrocarbon solvation with increasing solute size in an aqueous urea solution was also studied by considering the solute-solvent and solvent-solvent interactions.[42]

In this chapter we discuss the potential of mean force (PMF; see the chapter 2.3.4) of a *neo*-pentane pair in pure water and an aqueous urea solution with particular emphasis on the energy and entropy contributions. Because urea is well known as a protein denaturant, the hydrophobic interactions in an aqueous urea solution will be essential in protein folding. For example, in water, nonpolar solute molecules associate with each other. This so-called hydrophobic interaction is one of the major driving forces in protein folding. In the simplest, longstanding view, pairing up offers the molecules the advantage of reducing the solvent-exposed surface area, reducing the number of first shell water molecules, and thus reducing the unfavourable entropy at 298 K.[133] This picture has been confirmed by computer simulation studies of the PMFs of bringing together two nonpolar molecules in water, which revealed that the contact pair is stabilized by the entropy while the solvent-separated pair is stabilized by the energy (enthalpy).[134-136]

A comparison is made with the same system in pure water and the observed urea-induced changes of the hydrophobic interaction are discussed in terms of the changes in these two quantities. In particular, effects of solvent reorganization that contribute to the energies and entropies but compensate (cancel) in the PMF are singled out. In aqueous urea these contributions are large and it will be shown that, therefore, the energetic and entropic components of the PMF cannot unambigiously be used to argue if minima of the PMF are stabilized by enthalpy- or entropy-driven molecular-scale processes. The molecular-solvation mechanisms responsible for the large enthalpy-entropy compensation phenomenon occurring in the PMF as well as the mechanisms underlying the changes of the entropy will be discussed and quantified separately.

Herein, we evaluate the free energy, enthalpy, and entropy changes in solvating pair *neo*-pentane in water and aqueous 6.9 M (mole fraction, 0.15) urea solution. In general, protein denaturation is observed in the concentrated urea solution. Experimentally urea solution is saturated beyond the 9 M (mole fraction, 0.25).[137, 138]

Hydrophobic interactions in aqueous urea solution in relation to aspects of protein denaturation, will be discussed in detail in the chapter 6. In this chapter, enthalpic and entropic changes concerning to the PMFs of the pair *neo*-pentane in water and aqueous 6.9 M urea solution are focused.

## 5.2 Enthalpy and entropy compensation

In this section, process of the enthalpy-entropy compensation in the solvation free energy is introduced. A separation of solute-solvent and solvent-solvent contributions to solvation enthalpy, entropy, and exact enthalpy-entropy compensation in the solvation free energy have been referred to in earlier studies.[115, 123, 132, 139-150]

What is the enthalpy-entropy compensation in the free energy? On a particle insertion into the solvent, enthalpy and entropy changes at the same time are increased by inserted particle. In general, the free energy is described in the following as,

$$\Delta G = \Delta H - T\Delta S \tag{5.1}$$

In the eq. 5.1, the minus sign in the free energy gives rise to the compensation of enthalpy and entropy.

In the chapter 2.3 and 2.4, the free energy and chemical potential were addressed using different approach methods. In the case of one solute insertion into the solvent, the Boltzmann factor $\exp(-\beta\psi)$ is evaluated, where $\psi$ equals the interaction energy of the solute surrounding with all other solvent. It can be described in two ways such as,[77]

$$\boldsymbol{B} = \left\langle e^{-\beta\psi} \right\rangle_0 = 1/\left\langle e^{\beta\psi} \right\rangle_1 \tag{5.2}$$

where subscript 0 indicates average over solvent configurations before having inserted solute and subscript 1 stands for average over solvent configurations after having inserted solute. As it has been known, the Boltzmann factor $\boldsymbol{B}$ is closely related to solvation thermodynamics. Thus, the solvation free energy, $\Delta G_S$ is given by

$$\begin{aligned} \Delta G_S &= -k_B T \ln \boldsymbol{B} = k_B T \ln \left\langle e^{\beta\psi} \right\rangle_1 \\ &= \langle \psi \rangle_1 + k_B T \ln \left\langle e^{\beta\left(\psi - \langle\psi\rangle_1\right)} \right\rangle_1 \\ &= \Delta U_{UV} - T\Delta S_{UV} \end{aligned} \tag{5.3}$$

where $\langle \psi \rangle_1$ indicates average interaction energy of an inserted solute with all other solvent, thus $\langle \psi \rangle_1$ is denoted by $\Delta U_{UV}$ which is a direct solute($U$)-solvent($V$) interaction. And $\Delta S_{UV}$ is the solute-solvent entropy change, produced by creating molecular-sized cavity. It quantifies the excluded volume contribution.

Unlike the free energy, the solvation enthalpy change, $\Delta H$ and entropy change, $\Delta S$ are ensemble dependent.[148, 151] In the constant pressure system, the $(\Delta H)_P$ and $(\Delta S)_P$ are affected by local change caused interactions between the solute and its surrounding solvation shell. Thus, solvation enthalpy in the constant pressure ensemble is given by the sum of the solute-solvent binding energy and solvent-solvent reorganization energy.

$$\begin{aligned} (\Delta H)_P &\simeq (\Delta U)_P \\ &= \langle \psi \rangle_1 + \left[ \langle U_N \rangle_{N+1} - \langle U_N \rangle_N \right] \\ &= \Delta U_{UV} + \left( \Delta U_{VV} \right)_P \end{aligned} \tag{5.4}$$

where the term $\left[ \langle U_N \rangle_{N+1} - \langle U_N \rangle_N \right]$ corresponds to the change induced by creating a molecular sized cavity (*i.e.,* larger than inserted solute size). This is called the solvent reorganization energy, $\left( \Delta U_{VV} \right)_P$ in constant pressure condition.

Likewise, solvation entropy at constant pressure is given by

$$(\Delta S)_P = \Delta S_{UV} + \Delta S_{reo} \tag{5.5}$$

where $\Delta S_{UV}$ is the solute-solvent entropy change, and $\Delta S_{reo}$ is solvent reorganization entropy change. However, the solvent reorganization term is implicit contribution in the solvation free energy because $(\Delta U_{VV})_P$ and $\Delta S_{reo} = \Delta U_{VV} / T$ are exactly cancelled out such as the following relationship,

$$\begin{aligned}\Delta G_S &= (\Delta H)_P - T(\Delta S)_P \\ &= \Delta U_{UV} + (\Delta U_{VV})_P - T(\Delta S_{UV} + \Delta S_{reo}) \\ &= \Delta U_{UV} - T\Delta S_{UV}\end{aligned} \tag{5.6}$$

Eq. 5.6 shows the enthalpy-entropy compensation in the solvation free energy. The solute-induced disruptions of cohesive solvent-solvent interactions are always exactly cancel out in the free energy, hence only the solute-solvent interactions explicitly contribute to the free energy. Therefore, it is important to consider implicit solvent reorganization contribution in the solvation free energy. In fact, solvent reorganization is largely contributed in many solvation processes.

## 5.3 Simulation details

All simulations were performed using the SPC/E water model,[96] KBFF urea model[6] and GROMOS96 *neo*-pentane model.[100] An aqueous 6.9 M urea solution with 153 urea and 847 water, and 1000 water for pure water system were designed in the cubic periodic box. All simulations using molecular dynamics were conducted at constant pressure (1 atm) and three different temperatures (273K, 298K, 323K) conditions. Nonbonded parameters and charges for models are shown in table 5.1.

PMFs were calculated by simulating 80 independently equilibrated systems, in each applying a rigid constraint fixing the distance *r* between the mass centers of the two solutes at a preset value, $0.40 \le r \le 1.2$ nm (using equi-distant spacings $\Delta r = 0.01$ nm), and integrating the average constraint force accumulated in each 10-20 ns simulation from 1.2

nm downward. These long runs, which in total require more than $1\mu s$ sampling time per PMF curve, are required to reduce the statistical uncertainty of the first PMF minimum to 0.2 kJ/mol. The PMF was corrected with a term $2k_BT\log(r)$ to ensure that trivial entropy contributions, related to a larger volume element ($\sim r^2$) sampled by the constrained but freely rotating solute pair at larger distance *r*, do not enter the PMF.

To obtain the entropy $\Delta S(r)$ and energy $\Delta H(r)$ as a function of the solute-solute distance *r*, the PMFs (free energy change, $\Delta G(r)$; the $G$ in chapters 4, 5 and 6 denotes free energy, and should not be confused with the Kirkwood-Buff integral *G* in chapters 2 and 3) were calculated at three temperatures (*T*=273 K, 298 K, and 323 K) and the entropy was obtained by taking the numerical finite difference $\Delta S(r) = -\left[\Delta G(r;T+\Delta T) - \Delta G(r;T-\Delta T)\right]/2\Delta T$ with *T*=298 K and Δ*T*=25 K. The energy was finally obtained from $\Delta H(r) = \Delta G(r) + T\Delta S(r)$.

*Table 5.1 Nonbonded parameters and partial charges of urea, water, and neo-pentane models.*

| Model | Atom | $C_6(i,i)(10^{-3}$ kJ mol$^{-1}$ nm$^6)$ | $C_{12}(i,i)(10^{-6}$ kJ mol$^{-1}$ nm$^{12})$ | $q_i(e)$ |
|---|---|---|---|---|
| Urea (KBFF) | O | 1.988E-03 | 1.764E-06 | -0.675 |
| | C | 4.789E-03 | 1.375E-05 | 0.921 |
| | N | 1.810E-03 | 1.637E-06 | -0.693 |
| | H | 5.476E-06 | 8.520E-11 | 0.285 |
| Water (SPC/E) | O | 2.617E-03 | 2.634E-06 | -0.8476 |
| | H | 0.0 | 0.0 | 0.4238 |
| Neopentane (GROMOS) | C | 2.397E-03 | 2.053E-04 | 0.0 |
| | $CH_3$ | 9.614E-03 | 2.665E-05 | 0.0 |

## 5.4 Results

The PMFs of the *neo*-pentane pair in water and aqueous urea solution are discussed with enthalpic and entropic changes in detail. First of all, PMFs in pure water and aqueous urea solution were simulated at the three different temperatures. Figure 5.1 shows the *neo*-pentane pair PMF in water (a) and 6.9 M urea solution (b) at 273 K (dotted), 298 K (full), and 323 K (dashed). The first PMF minimum, further on in this thesis referred to as the contact minimum (CM), becomes deeper with increasing temperature in water as well as in urea solution.

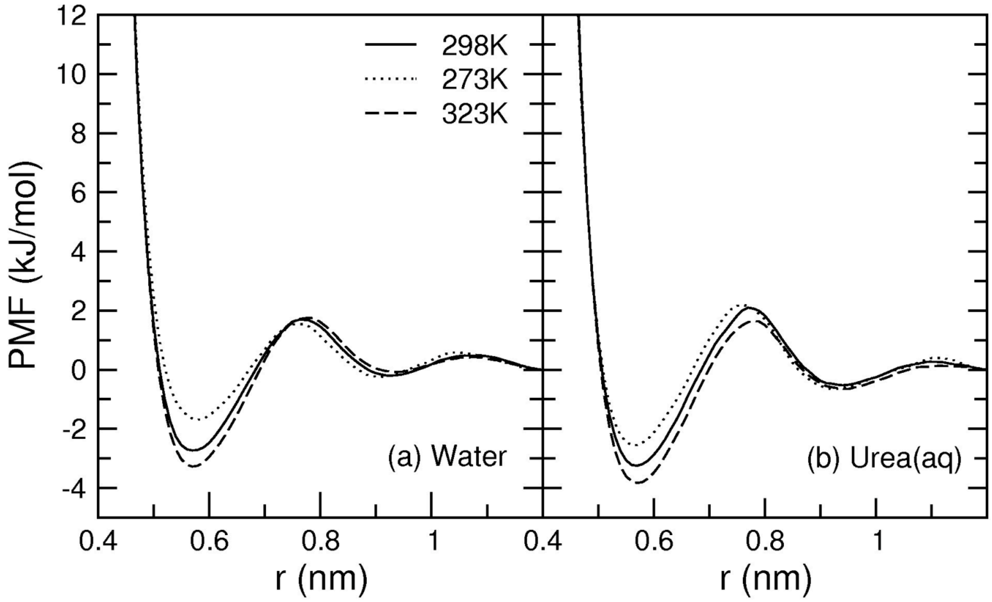


*Figure 5.1 (a) neo-pentane pair PMF in water at 273K (dotted), 298K (full), and 323K (dashed). (b) neo-pentane pair PMF in 6.9 mol/l urea solution at 273K (dotted), 298K (full), 323K (dashed). The statistical uncertainty in the first minima is 0.2 kJ/mol.*

This indicates an increase of entropy in pairing the *neo*-pentane at the contact distance in both systems. This means that *neo*-pentane association at the CM distance is entropically favourable in water and aqueous urea solution.

Southall and Dill,[136] based on calculations with the Mercedes-Benz model of water,[152, 153] showed that the oscillations of the enthalpy and entropy correlate with water hydrogen bonding. They presented that two nonpolar molecules at the contact distance are stabilized by entropy, whereas solute pairing at the solvent separate distance is stabilized by enthalpy in water. Figure 5.2 shows the PMF, enthalpy and entropy related to the *neo*-pentane pairing distance, $r$ at 298K in water (a) and aqueous urea solution (b). In water, the entropy stabilizes the CM, while the enthalpy destabilizes it. Note that these contributions have the same sign and therefore largely cancel out each other in the PMF. The second PMF minimum in fig 5.2a, further on in the thesis referred to as the solvent-separated minimum (SSM), is enthalpically favourable while being entropically unfavourable. Note that the location of the minimum in the $\Delta H(r)$ and $T\Delta S(r)$ curves does not coincide with the SSM of the PMF. This shift is caused by the water-water (hydrogen bonding) interactions being stronger at a slightly smaller distance. We observed a similar shift at the SSM in water and aqueous urea solution in fig. 5.3. Rick[135][154] showed a same trend of the entropy minimum relative to the SSM in his study of methane-methane interactions in water using a polarizable, fluctuating charge (FQ) model for water.

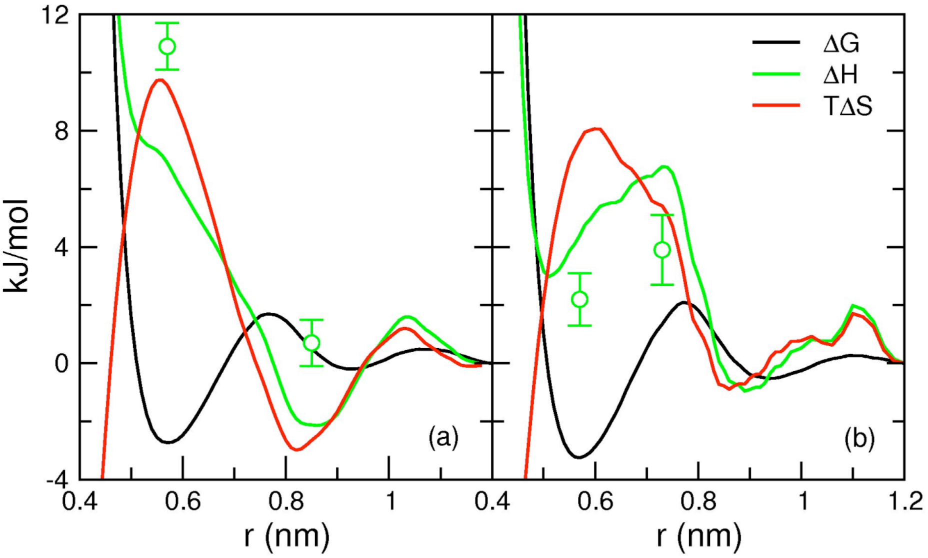


*Figure 5.2 neo-pentane pair PMF (black) and the enthalpy (green) and entropy (red) contributions in water (a) and 6.9 mol/l urea solution (b) at T=298 K and P=1 atm. The data points in (a) and (b) indicated with a circle were obtained by subtracting the average enthalpy of a box with a neo-pentane-neo-pentane constraint distance of 0.57nm (and 0.78nm in urea solution, 0.85nm in water) from the average enthalpy of a box with a neo-pentane-neo-pentane constraint distance of 1.2nm. The average enthalpies were obtained from 100ns NPT simulation at T=298K.*

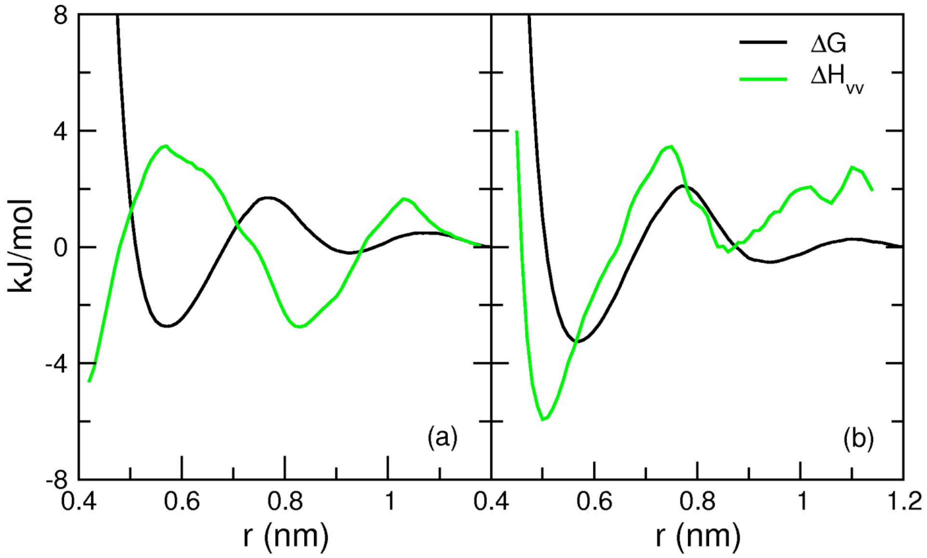


*Figure 5.3 Solvent reorganization energy (green) in water (a) and 6.9 mol/l urea solution (b) at T=298K and P= 1 atm. The PMFs (black) are included for comparison.*

In urea solution (fig. 5.2b), approach of the CM (*r*<0.7 nm) causes an increase of the entropy and a decrease of the enthalpy; hence the enthalpy and entropy changes are reinforcing in the PMF and the enthalpy in contrast to what is observed in water (fig. 5.2a) contributes favourably to the stability of the CM.

The green circles with error bar in fig. 5.2 were obtained by subtracting the average enthalpy at the *neo*-pentane - *neo*-pentane constraint distance (water: 0.57nm, 0.85 nm, urea solution: 0.57nm, 0.78nm)) from the average enthalpy with a *neo*-pentane - *neo*-pentane constraint distance of 1.2nm. The average enthalpies were obtained from 100ns NPT simulation at T=298K. These data points show a similar trend with green line ($\Delta H = PMF + T\Delta S$) although these data are shift up (water) and down (urea solution) of a similar amount relative to green line. The tendency of the enthalpy ($\Delta H$) and entropy ($T\Delta S$) curves to oscillate in-phase with comparable amplitudes means that significant parts will cancel in the PMF ($= \Delta H - T\Delta S$). Statistical-mechanical analyses of solvation processes have previously shown that the solvent reorganization enthalpy, defined as the change in the enthalpy of solvent-solvent interactions, is exactly enthalpy-entropy compensating in the section 5.2. That means that only those enthalpy- and entropy contributions deriving directly from the solute-solvent coupling determine the free energy change. Contributions deriving from solute-induced changes of cohesive solvent-solvent interactions do not affect the change of free energy. In urea solution we expect that compensating solvent-solvent contributions are large.

Early experiments by Wetlaufer *et al.*[34] have shown that enthalpies and entropies of *neo*-pentane transfer from water to aqueous urea solution are positive while the transfer free energy is negative. Neglecting (the obvious) enthalpy-entropy compensation of this process one would conclude the transfer to be entropically driven. Computer simulations3 and theoretical predictions based on scaled-particle-theory[155] however have indicated that the transfer is driven by favourable *neo*-pentane-urea Van der Waals interactions, not by the entropy. The combined experimental and theoretical work thus indicates that contributions of (compensating) solvent reorganization processes to the solvation enthalpy and entropy in aqueous urea are significantly larger than in water and affect the solute transfer thermodynamics in a way that these processes blur signatures of the true molecular driving force.

For the present case of a pairwise hydrophobic interaction, the solute pair (with separation distance *r* fixed) is considered as "one solute" and a decomposition scheme based on solute-solvent and solvent-solvent contributions – as it has been used in solvation studies of a single solute[42, 115, 123, 125] – is applied to the enthalpy and entropy.

The solute-solvent enthalpy change $\Delta H_{UV}(r)$, solvent-solvent enthalpy change $\Delta H_{VV}(r)$ and solute-solvent entropy change $\Delta S_{UV}(r)$ are functions of the constrained distance *r* within the "solute". The *neo*-pentane-*neo*-pentane Van der Waals interactions are included in $\Delta H_{UV}(r)$. The solvent reorganization enthalpy $\Delta H_{VV}(r)$ is obtained by taking the difference between $\Delta H(r)$ and $\Delta H_{UV}(r)$, the solute-solvent entropy change $\Delta S_{UV}(r)$ is obtained from the difference between $\Delta H_{UV}(r)/T$ and $\Delta G(r)/T$. We note that all changes in water-water, urea-water, or urea-urea hydrogen bonding and Van der Waals interactions with changing solute separation *r* will contribute to the solvent reorganization enthalpy.

The solvent reorganization enthalpies $\Delta H_{VV}(r)$ in water and urea solution are shown in fig. 5.3a and fig 5.3b, respectively, together with the PMFs. In water, the reorganization enthalpy oscillates. This is consistent with the observation of Southall and Dill[136] who showed that the enthalpy and entropy oscillations correlate with water-water hydrogen bonding. In the CM, hydrogen bonds are broken causing an increase of the enthalpy and entropy (eq. 5.1). In the SSM, hydrogen bonds are stabilized causing a decrease of, both, enthalpy and entropy. In urea solution, an oscillatory trend is also observed, however a rather deep minimum arises at the CM. Urea preferentially solvates the two nonpolar solutes (see fig. 5.4) and in bringing them into contact urea molecules are expelled from the gap in between the two solutes. The "released" urea molecules form hydrogen bonds with water and/or urea molecules in the bulk solution, causing a large, and negative, solvent reorganization enthalpy as observed in fig. 5.3b. The strengthening of solvent-solvent cohesive interactions in the bulk contributes negatively to the entropy. Because the entropy of the CM in aqueous urea (fig. 5.2b) is only marginally smaller than the entropy of the CM in water, despite solvent-solvent reorganization energy in urea solution is largely change at the CM, an additional process must occur in urea solution that increases the solute-solvent entropy $\Delta S_{UV}$ and compensates for the negative solvent reorganization entropy ($\Delta H_{VV}/T$).

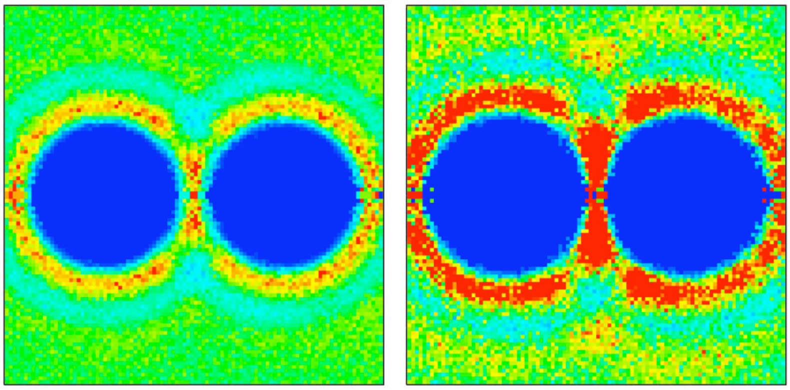


*Figure 5.4 Water (left) and urea (right) number density maps for the neo-pentane at the SSM distance. Blue: zero density, green: average bulk solution density (22.9 water molecules per nm$^3$, 4.13 urea molecules per nm$^3$), red: twice the average bulk density. Notice that urea preferentially bridges between the hydrophobic solutes in the solvent-separated state (right)*

This process will be addressed below. In water (fig. 5.3a) not only the trend of $\Delta H_{VV}$ , but beyond $r$=0.7 nm also the magnitude matches the entropy $T\Delta S$ in fig. 5.2a. The entropy change in water at distances $r$>0.7 nm therefore is completely determined by the reorganization of water-water hydrogen bonds.

Figures 5.5a and 5.5b show the contributions of the solute-solvent enthalpy $\Delta H_{UV}$ and solute-solvent entropy $T\Delta S_{UV}$ to the PMF in water and 6.9 M urea solution, respectively. In water, the CM is stabilized by the solute-solvent entropy but destabilized by the solute-solvent enthalpy. The latter happens because the contact-pair has a smaller solvent accessible area than the solvent-separated pair. In the CM, both, $\Delta H_{UV}$ and $T\Delta S_{UV}$ are positive but smaller in magnitude than $\Delta H$ and $T\Delta S$ (fig. 5.2a), because the former quantities do not include the positive enthalpy and entropy contributions of hydrogen-bond reorganization (~ +3.5 kJ/mol; fig. 5.3a). The SSM (0.92 nm) is stabilized by the solute-solvent entropy. In urea solution (fig. 5.5b) the CM is stabilized by the solute-solvent entropy and destabilized by the solute-solvent enthalpy. In the CM, both, $\Delta H_{UV}$ and $T\Delta S_{UV}$

are positive and larger in magnitude than $\Delta H$ and $T\Delta S$ (fig. 5.2b) because the solvent reorganization enthalpy, contributing to the latter two quantities, is negative (fig. 5.3b). The SSM is stabilized by the solute-solvent enthalpy. Solute-solvent enthalpic stabilization of the SSM in aqueous urea solution is caused by preferential interactions between urea and the aliphatic solutes. This preferential solvation phenomenon is driven by urea-hydrocarbon Van der Waals interactions[3] and stabilizes the SSM because solute-intermediate urea molecules interact with two solutes at the same time (see fig. 5.4). The solute-solvent entropic stabilization of the CM can be understood better after considering the statistical mechanical expression for the solute-solvent entropy upon solute solvation,[125, 147]

$$\Delta S_{UV} = k_B \ln P_{ins} - k_B \left[ \left( \langle \psi^2 \rangle - \langle \psi \rangle^2 \right) \beta^2 / 2 + \cdots \right]_a . \tag{5.7}$$

In eq. 5.7, $k_B$ is the Boltzmann constant and $\beta = 1/k_B T$. The quantity $P_{ins}$ is the probability that a randomly inserted solute molecule (or a solute pair separated at a fixed distance *r*) into a system of only solvent molecules will experience an attractive or zero interaction energy ($\psi \leq 0$). Thus, $k_B \ln P_{ins}$ may be viewed as a solute-solvent excluded volume contribution to the entropy. The second term on the right-hand-side of eq. 5.7 describes an additional entropy penalty related to how strongly attractive solute-solvent interactions ($\psi < 0$; denoted by subscript *a*) bias the solvation shell composition.[125] In the (hypothetical) case that *neo*-pentane-urea dispersion interactions are of the same magnitude as *neo*-pentane-water dispersion interactions, the fluctuations ($\langle \psi^2 \rangle - \langle \psi \rangle^2$) of the solute-solvent interaction energy $\psi$ are small and solute-solvent attractive interactions put little bias on the solvation shell composition. Favourable *neo*-pentane-urea attractive interactions do however cause a strong bias on sampling configurations with excess local urea concentrations. (see fig. 5.4) This biasing process leads to a selected, non-random solvation shell structure and a lowering of entropy which is stronger in the SSM than the CM. The solute-solvent entropy (fig. 5.5b) therefore stabilizes the CM relative to the SSM as well as larger solute-solute separations. On approaching the CM the solute-solvent enthalpy increases (fig. 5.5b) because favourable *neo*-pentane-urea Van der Waals interactions are lost. Figures 5.6a and 5.6b show the contributions of the two terms on the right-hand-side of eq. 5.7 to the solute-solvent entropy in water and urea solution, respectively. The contribution $k_B T \ln P_{ins}$ is very similar in both systems, whereas the second contribution indeed contributes positively to the solute-solvent entropy in passing from the SSM to the CM in aqueous urea solution.

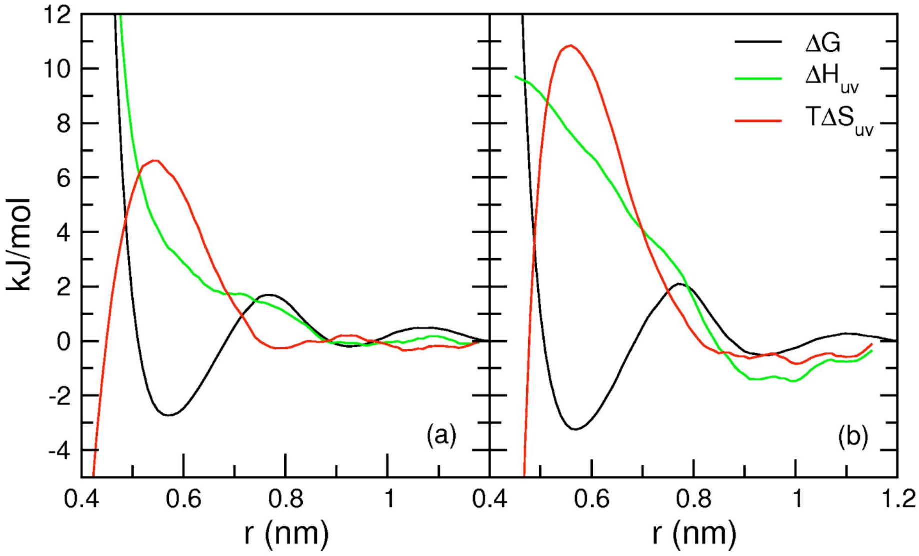


*Figure 5.5 Solute-solvent energy (green), solute-solvent entropy (red) and the PMF (black) in water (a) and 6.9 mol/l urea solution (b) at T=298K and P= 1atm.*

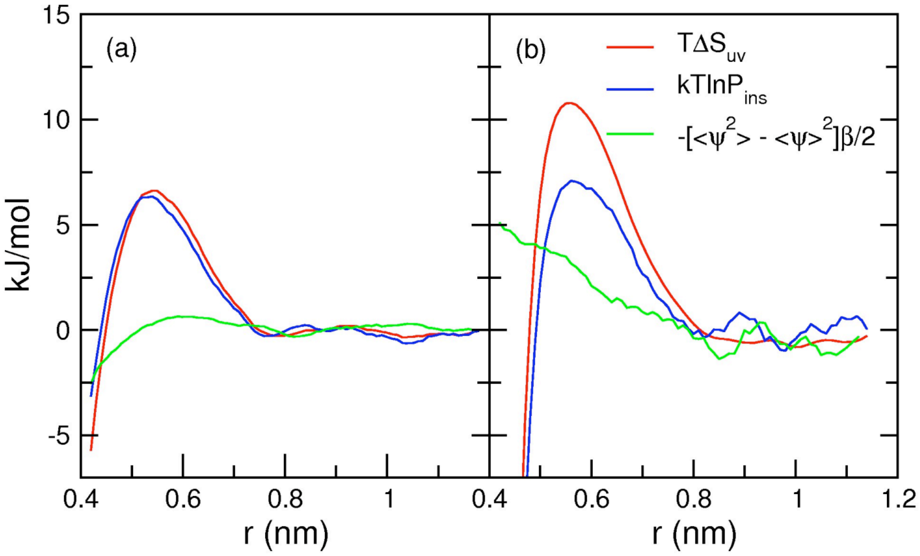


*Figure 5.6 Solute-solvent entropy (red) and contributions (cf. eq. 5.6) of the insertion probability (blue) and solute-solvent energy fluctuations (green) in water (a) and 6.9 mol/l urea solution (b) at 298 K and 1 atm. Solute-solvent energy fluctuations were obtained by extending the MD simulations at each distance r to 50 ns.*

## 5.5 Conclusions

In summary we have used concepts from solvation thermodynamics to argue that changes in the potential energy of cohesive solvent-solvent interactions are enthalpy-entropy compensating in the potentials of mean force of pairwise solute-solute interactions. For hydrophobic interactions in aqueous urea the compensating contribution is large and dominates the features of the enthalpy and entropy change upon solute-solute approach: in the contact minimum of the *neo*-pentane-*neo*-pentane potential of mean force in 6.9 M urea the system's enthalpy (and entropy) is reduced relative to what is observed in pure water because solvent-solvent hydrogen bonds involving urea molecules become stronger. The energy and entropy decomposition turns out to be particularly useful in systems with more than one solvent component[42, 115, 123] and where preferential solvation by one of the solvent components affects the strength of the effective solute-solute pair interaction. Although, through the formation of urea-solvent hydrogen-bonds, the release of urea molecules from the *neo*-pentane solvation shells upon approaching the contact minimum leads to a reduction of the system's enthalpy and entropy, one should think of this solvent reorganization phenomenon merely as a "response effect" rather than a driving force. The molecular driving force stabilizing the contact pair is the solute-solvent entropy, which increases upon solute-solute approach. The latter happens because Van der Waals interactions bias the *neo*-pentane solvation shells to preferentially contain urea molecules. Part of the selected, non-random solvation shell structure is lost in the contact minimum causing an increase of solute-solvent entropy.

# 6. Urea-induced changes of hydrophobic interactions

## 6.1 Introduction

So far, my thesis has been focused on solvation thermodynamics, solvation structures and hydrophobic interaction in binary liquid mixtures and water. In order to describe properly these properties, a force field development for binary liquid mixture has been performed. Additional studies were concerned with solvation thermodynamics and solvation structure of a simple hydrophobic solute (*i.e.,* methane) using our developed model (LV force field). In chapter 5, pair potentials of mean force (PMFs) of nonpolar solutes in pure water and in aqueous urea solution were discussed in terms of free energy, enthalpy and entropy changes. In this chapter, hydrophobic interactions between aliphatic and aromatic solutes in urea solution are calculated. The effects of urea on the strength of these interactions are discussed in the context of protein denaturation in aqueous solutions of urea.

Hydrophobic interactions play an important role in folding of proteins, formation of vesicles, membranes, etc. Because this interaction is mediated through the aqueous solvent, the stability of the above structures is particularly sensitive to external thermodynamic conditions. Hence, temperature, pressure, ionic strength, or the activities of co-solvents and co-solutes affect the magnitude of hydrophobic interactions and the stability of biomolecules.

In the first part of this chapter, proteins are briefly introduced to understand their structure, stability and denaturation. Then, the previous studies related to our current work for hydrophobic interactions in the presence of urea are addressed.

It is generally believed that hydrophobic contacts are dissolved in the presence of urea, thereby contributing to the denaturing process of globular proteins in concentrated urea solutions.[33-38, 117, 156-159] Direct and indirect mechanisms have been discussed in the literature.[33, 38, 40, 160-163] The direct mechanism involves urea H-bonding to the peptide backbone, thereby favouring the denatured state. The indirect mechanism usually adopts a chaotrope argument: urea perturbs the water structure so that hydrophobic groups are more easily solvated. Despite recent indications in favour of a direct mechanism,[160, 162, 163] protein denaturing via both the direct and indirect mechanism has been emphasized as well.[33]

In this work, three types of hydrophobic residues are considered; one is an aliphatic hydrocarbon (*neo*-pentane) and the other two are aromatic hydrocarbon molecules (toluene, *3*-methylindole). In fact, toluene and *3*-methylindole are prototypes for the non-polar, uncharged side groups of the phenylalanine and trpytophan amino acids. Their chemical structures are shown in figure 6.1.

In this chapter, I present potentials of mean force (PMFs) for aliphatic and aromatic pair interactions in water and in an aqueous 6.9 M urea solution at the molecular scale. It will be concluded that hydrophobic clusters do not dissolve entirely in aqueous urea but instead urea acts as a bridging-agent that keeps the hydrophobic pairs together (chapter 5). The implications of this finding for urea-induced protein denaturation will be discussed.

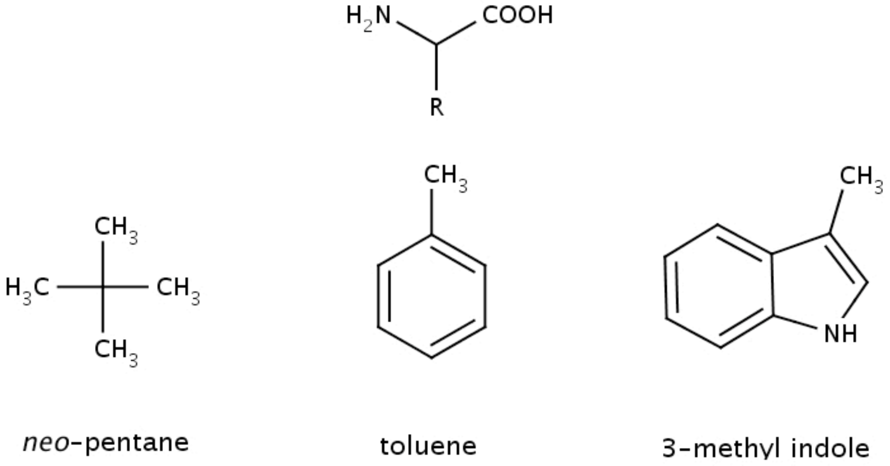


*Figure 6.1 Chemical structures for amino acid with side chain residue R. In this thesis, neo-pentane, toluene, and 3-methylindole were used as prototypes of nonpolar amino acid side groups.*

### 6.1.1 Proteins

The amino acid is the elementary unit in all proteins. Twenty different amino acids exist in nature and all contain the same group $H_2NC_\alpha HRCOOH$. The central carbon (called the *alpha*-carbon, $C_\alpha$) is attached to a side group residue *R*. Amino acids in proteins are joined together by peptide bonds. Thus, the poly-peptide repeat unit $\left[\text{NH - C}_\alpha\text{HR-CO}\right]$. The linear amino acid sequence along the chain is referred as the primary structure.

Folded proteins are characterized by a secondary and tertiary structure. The secondary structure refers to the spatial arrangements of parts of the polypeptide chain; these include $\alpha$ helices, $\beta$ sheets and turns. The way in which the various secondary structured motifs along the polypeptide chain pack together in the overall three-dimensional arrangement of the protein is referred to as the tertiary structure. The latter packing is driven by a variety of intermolecular forces including forces of covalent (*di*-sulfide bonds), electrostatic (hydrogen bonds, ionic interactions, dispersion), and hydrophobic nature. In particular, the latter interaction, driven by the hydrophobicity of nonpolar groups, causes the nonpolar amino acids to mainly pack inside the protein core. Polar and charged amino acids at the protein surface shield the core from its exposure to the aqueous solvent.

Environmental changes may cause changes of the protein's native shape and conformation. These changes are usually caused by heat, change of pH, denaturants (*i.e.,* urea, guanidine hydrochloride, beta-mercaptoethanol, dithiothreitol), inorganic salts (i.e lithium bromide, potassium thiocyanate, sodium iodide), organic solvents (*i.e.,* formamide, dimethylformamide, dichloro- and trichloroacetic acids and their salts), detergents (*i.e.,* sodium dodecyl sulphate), high pressure and ultrasonic homogenization. Protein denaturation is a structural change in biomolecules that leads to the loss of its biological function.

### 6.1.2 Protein denaturation studies using the PMFs

Denaturation mechanisms are poorly understood in spite of extensive studies. In this thesis, based on simple models we discuss protein denaturation induced by urea. In the classical view of urea denaturation mechanism, two aspects are considered. One is that urea weakens hydrophobic interaction with increasing the solubility of hydrocarbons,[34] and the other is that urea weakens the intra-molecular hydrogen bonds of proteins by binding directly to proteins themselves. In general, urea is classified as a structure breaker; *i.e.,* urea alters water structure such that solvent exposure of the protein core becomes thermodynamically favourable.[164]

Numerous computer simulations have been performed to understand the role of urea at the molecular scale. Data on single nonpolar solute in water and urea solution do not suffice to completely understand protein unfolding. The solvent-mediated interactions among two or more nonpolar solutes should be considered. Methane-methane association hydrophobicity has been studied by Wallqvist *et al.* and Ikeguchi *et al.*[40, 160]

Wallqvist *et al.*[160] first simulated potentials of mean force between two methane molecules in aqueous 6 M urea solution and in water. They examined the effect of charges on the effective hydrophobic interaction using "uncharged methane" and "charged methane" because proteins in compact conformation typically include charged residues on the surface. They reported that the urea molecules preferentially adsorb onto the charged hydrophilic residues on the surface while urea stabilized the contact hydrophobic association. The hydrogen bonding distribution of urea-water was found similar to the water-water hydrogen bonding distribution in bulk water. This means that urea does not break a water structure. Urea adsorption on the hydrophilic residues leads to a repulsion between the residues on the surface of proteins and gives rise to expose the hydrophobic residues in the protein core.

Ikeguchi *et al.*[40] studied PMFs between two methane molecules in 7 M aqueous urea and in pure water. The PMFs showed that urea stabilized the methane-methane contact association. However, they calculated the transfer free energies of aliphatic molecules in water to aqueous urea solution to examine the solute-size effect on hydrophobic interaction by urea. Their results showed that hydrophobic interactions between larger hydrocarbons (more than two carbons) were destabilized by urea, while those between small hydrocarbons (methane, ethane) were stabilized by urea. Ikeguchi *et al.* suggested that the model of two methane molecules is not an appropriate approximation of the exposed hydrophobic groups of proteins. Hydrophobic interactions between the exposed larger hydrophobic groups might be weakened in aqueous urea.

Shimizu and Chan[38] investigated pair PMFs of nonpolar solutes (methane - methane, methane - *neo*-pentane, *neo*-pentane - *neo*-pentane) in water and in aqueous urea solution. They reported that urea (de) stabilized hydrophobic interaction depending on the size of hydrophobic solutes. One of their results (the *neo*-pentane pair PMFs) is shown in figure 6.2(d) of this chapter. In fact, Shimizu and Chan[38] simulated *neo*-pentane as a spherical super-atom in an aqueous 6.0 M urea solution, and PMFs were taken to be zero at a 10 Å distance between the two *neo*-pentane particles. As can be seen in figure 6.2(d), *neo*-pentane association is destabilized by urea at the contact distance, while methane - methane and methane - *neo*-pentane associations are not destabilized by urea in their simulation.[38]

Unlike with aliphatic hydrocarbons, it is unclear if urea destabilizes hydrophobic interactions between aromatic hydrocarbons.

Duffy *et al.*[39] used MC simulations to study interactions of urea and guanidinium with aromatic hydrocarbons by computing the free energy profiles of association. They suggested the possibility of direct interaction of urea/guanidinium with protein. Greater exposure of aromatic side chains in unfolding state will give an opportunity for favourable interactions with the urea.

Chipot *et al.*[165] investigated the relative orientational preferences for benzene dimers and toluene dimers in a vacuum and in water. Chelli and Gervasio *et al.*[166]studied the aromatic-aromatic amino acid interaction in various solvents: unbound, bound– phenylalanine-phenylalanine (Phe-Phe), tyrosine-tyrosine (Tyr-Tyr), phenylalanine-tyrosine (Phe-Tyr), histidine-trpytophan (His-Trp) in vacuum, water, carbon tetrachloride, dimethyl sulfoxide and methanol. They reported the competition of the stacked, T-shaped or crossed interactions among the amino acids in different solvent. Stacking interactions are favourable in all solvents with the exception of the Tyr-Tyr interaction in carbon tetrachloride, where T-shaped structures are also important. Polar solvents such as water, methanol, dimethyl sulfoxide, prevent from forming the hydrogen bonding interaction between the two aromatic residues. Meanwhile hydrophobic solvent such as carbon tetrachloride ($CCl_4$) stabilized the hydrogen bonding between two aromatic complexes. Therefore aromatic complex in $CCl_4$ solvent was observed both of stacked and T-shaped structures.

In the case of aromatic residues, hydrophobic interactions are mainly investigated in terms of relative structural arrangements of solute pairs in the various solvents.

## 6.2 Simulation details

As model systems for the influence of urea on pair interactions of typical hydrophobic (nonpolar) amino acid residues, we investigated the potentials of mean force (PMFs, see the chapter 2) of *neo*-pentane, toluene and 3-methylindole in water and in an aqueous solution of 6.9 M urea by molecular dynamics (MD) simulations.

MD simulations for the pair *neo*-pentane in water and in aqueous 6.9 M urea solution were performed based on two nonpolarizable urea models (OPLS[167] and KBFF[6]) and three rigid, nonpolarizable water models (SPC,[82] SPC/E,[96] and TIP4P[97]).

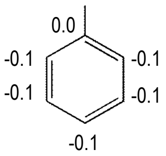
In the case of aromatic-aromatic solutes (toluene, 3-methylindole), the GROMOS force field[89] was used to model the solutes; the aromatic- and indole rings are described in all-atom detail while the methyl groups are treated with a single, united-atom interaction site. The SPC/E water model[96] was used while for urea the KBFF model[6] was used. Table 6.1 shows the nonbonded interactions and partial charges for all models.

In previous MD studies of urea-water mixtures the KBFF/SPC/E force field has been shown to provide an improved description of the solution structure and activity derivatives whereas the OPLS/SPC and OPLS/TIP4P force fields produced too large urea-urea and water-water aggregation.[6] We nonetheless examined PMFs with the different water and urea force fields to get a better idea of model dependencies. All systems contained two solutes (*neo*-pentane, toluene, or 3-methylindole as a solute), 1694 water and 306 urea molecules (2000 waters in case of pure water) in a cubic box with an edge length of 3.9-4.2 nm. PMFs were calculated by simulating a series of 100 independently equilibrated systems, each applying a rigid constraint fixing the distance $r$ between the mass centers of the two solutes at a preset value, $0.4<r<1.6$ nm, and integrating the average constraint force accumulated in each 10-20 ns simulation from 1.6 nm downward. At 1.6 nm all PMFs converge to zero. PMFs were corrected with a term $2k_BT\ln(r)$ to ensure that (trivial) entropy contributions related to a larger volume element ($\sim r^2$) sampled by the constrained but freely rotating solute pair at larger distance $r$ do not enter the PMF. All simulations were conducted at constant pressure (1 atm) and temperature (298 $K$). All simulation settings were identical to those in reference 42.

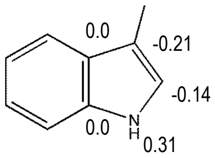
*Table 6.1 nonbonded interaction and partial charges for urea, water, neo-pentane, toluene and 3-methylindole.*

| Model | Atom | $C_6(i,i)(10^{-3}$ kJ mol$^{-1}$ nm$^6)$ | $C_{12}(i,i)(10^{-6}$ kJ mol$^{-1}$ nm$^{12})$ | $q_i(e)$ |
|---|---|---|---|---|
| Urea | | | | |
| OPLS | O | 2.365E-03 | 1.591E-06 | -0.390 |
| | C | 4.890E-03 | 1.360E-05 | 0.142 |
| | N | 3.355E-03 | 3.954E-06 | -0.542 |
| | H | 0.0 | 0.0 | 0.333 |
| KBFF | O | 1.988E-03 | 1.764E-06 | -0.675 |
| | C | 4.789E-03 | 1.375E-05 | 0.921 |
| | N | 1.810E-03 | 1.637E-06 | -0.693 |
| | H | 5.476E-06 | 8.520E-11 | 0.285 |
| Water | | | | |
| SPC | O | 2.617E-03 | 2.634E-06 | -0.82 |
| | H | 0.0 | 0.0 | 0.41 |
| SPC/E | O | 2.617E-03 | 2.634E-06 | -0.8476 |
| | H | 0.0 | 0.0 | 0.4238 |
| TIP4P | O | 2.552E-03 | 2.510E-06 | 0 |
| | D | 0.0 | 0.0 | -1.04 |
| | H | 0.0 | 0.0 | 0.52 |
| *neo*-pentane | | | | |
| GROMOS | C | 2.397E-03 | 2.053E-04 | 0.0 |
| | $CH_3$ | 9.614E-03 | 2.665E-05 | 0.0 |
| LJ | $C_5H_{12}$ | 7.567E-01 | 4.094E-02 | 0.0 |
| Toluene | | | | |
| GROMOS | C | 2.341E-03 | 3.375E-06 | -0.1/0.0[(a)] |
| | HC | 8.464E-05 | 1.513E-08 | 0.1 |
| | $CH_3$ | 9.614E-03 | 2.665E-05 | 0.0 |
| *3*-Methylindole | | | | |
| GROMOS | C | 2.341E-03 | 3.375E-06 | -0.14/-0.21/0[(b)] |
| | HC | 8.464E-05 | 1.513E-08 | 0.14/0.31[(b)] |
| | $CH_3$ | 9.614E-03 | 2.665E-05 | 0.0 |
| | N | 2.436E-03 | 3.389E-06 | -0.1 |

(a)



(b)



## 6.3 Results

First of all, I will discuss *neo*-pentane pair PMFs in water and in aqueous urea. As we have seen in chapter 5, the *neo*-pentane pair PMF has a minimum at the contact distance which is driven by entropy. *Neo*-pentane association at the solvent separated distance is energetically dominated (fig. 5.1). In particular, at the contact distance of a *neo*-pentane pair in aqueous urea solution, the enthalpy and entropy are reduced relative to what is observed in pure water because urea molecules, that are expelled from the nonpolar solvation shells as these solutes approach, form hydrogen bonds with the solvent. At the solvent separated distance, urea preferentially solvates the two *neo*-pentanes and mediates their attraction at that distance by occupying the intra-solute region (fig. 5.4).

Figure 6.2 shows the *neo*-pentane pair PMFs in pure water (solid lines) and in 6.9 M aqueous urea (dashed lines) for different combinations of water and urea models. *Neo*-pentane is modeled with the GROMOS,[89] five-site, united atom model in which $CH_3$-groups are modeled with an effective interaction site [fig. 6.2 (a)-(c)]. We also included in fig. 6.2(d) the *neo*-pentane pair PMF, based on a single-site Lennard-Jones (LJ) model for *neo*-pentane, whose parameters were taken from Kuharski and Rossky,[37] previously studied by Shimizu and Chan.[38]

Comparison of the well depths in the PMFs for water and aqueous urea in figs. 6.2(a)-(c) shows no evidence for urea-induced destabilization of the *neo*-pentane pair interaction. Rather, urea in our simulation stabilizes *neo*-pentane association at the contact distance. However, the first minimum of the PMF in aqueous urea in figure 6.1(d) is slightly shifted upward relative to the PMF in water. At the distance of the second minimum (solvent-separated pair), urea stabilizes the hydrophobic interaction relative to water in all different model combinations. The statistical error, obtained by integrating the error in the mean constraint force from 1.6 nm downward, varies between 0.1 *kJ/mol* in the second minimum and 0.2 *kJ/mol* in the first minimum.

Note however that urea-induced changes of this so-called contact pair (CP) are model dependent and fall just outside the error bar of the calculations. For all models, urea stabilizes the solvent separated pair (SSP) relative to the CP due to two corroborative effects: (1) the free energy basin of the SSP in aqueous urea is broader and is shifted to larger distance than in pure water, (2) the free energy difference between the SSP and CP in urea solution decreases relative to that in pure water (except for the KBFF/SPCE system in figure 6.2b).

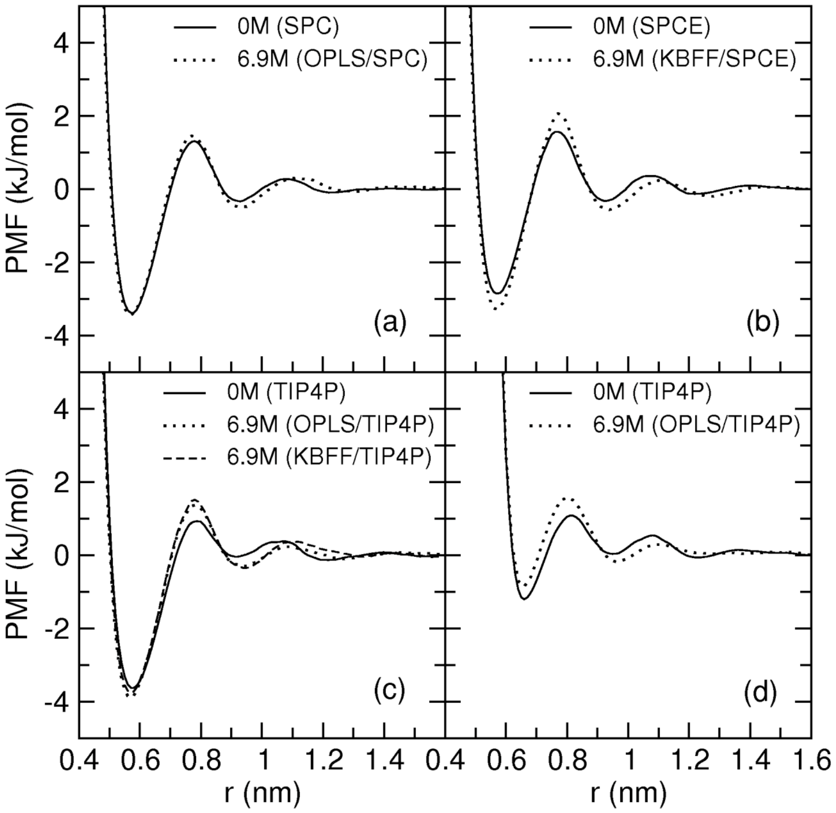


*Figure 6.2 PMFs for neo-pentane pair interactions in pure water (solid lines; water model in parentheses) and 6.9 M aqueous urea (dashed lines; urea/water models in parentheses). In (a)-(c) neo-pentane is modeled with the GROMOS,[89] five-site with united atom model of* $CH_3$*-groups. In (d), neo-pentane is modeled with a spherical Lennard-Jones potential.*[37]

The second effect evidently causes the equilibrium CP↔SSP in aqueous urea to shift in favour of the SSP in comparison to the equilibrium in pure water. The first effect however shifts the equilibrium towards the SSP even further because it results in a larger volume $\left(\sim 4\pi r^2\right)$ available to the SSP. Both effects together cause a significant stabilization of the

SSP relative to the CP, the extent of which we quantified by computing the equilibrium constant $K_{eq}$ defined as

$$K_{eq} = \frac{[SSP]}{[CP]} = \frac{\int_{R_1}^{R_2} r^2 \exp\left[-w(r)/k_BT\right]dr}{\int_0^{R_1} r^2 \exp\left[-w(r)/k_BT\right]dr} \tag{6.1}$$

where [SSP] and [CP] denote the concentrations of SSPs and CPs, respectively, $R_1$=7.8Å, $R_2$=10.8Å (water), $R_2$=11.1Å (aqueous urea), $w(r)$ is the PMF, $k_B$ the Boltzmann constant and $T$=298 $K$ the temperature. With the models studied in figs. 6.2(a)-(d), $K_{eq}$ in 6.9 M urea increases with 19% (OPLS/SPC), 6% (KBFF/SPC/E), 38% (KBFF/TIP4P), 36% (OPLS/TIP4P), and 45% (OPLS/TIP4P; LJ-*neo*-pentane) relative to pure water. Previous simulation studies have reported urea induced stabilization of methane-methane association.[40, 160] In contrast to these small solutes, based on the study in reference 38 where only the (urea-destabilized) CP was investigated and the SSP was neglected, hydrophobic interactions of relatively large nonpolar solutes are believed to be destabilized.[38]

Just like the pair PMFs of aliphatic solutes in water and in aqueous urea solution, PMFs for toluene-toluene and *3*-methylindole – *3*-methylindole in water and 6.9 M urea solution were simulated. The hydrophobicity of aromatic hydrocarbons (*i.e.,* benzene, toluene) has more complex mechanism than the hydrophobicity of aliphatic hydrocarbon.[146] Because the aromatic ring is a weak hydrogen bond acceptor, it is possible to make hydrogen bonds or $\pi - \pi$ interaction with neighboring aromatic ring or polar solvent. Another complexity is the anisotropic shape of the aromatic ring relative to aliphatic hydrocarbon. In earlier works, the relative importance of face-to-face stacked and the T-shaped interaction has been debated. Chipot *et al.*[165] studied the relative orientational preference of benzene-benzene and toluene-toluene dimers in a vacuum and in water. The effects of urea on the stability of aromatic pair interactions in water have not yet been investigated.

Figure 6.3 shows the toluene-toluene and *3*-methylindole – *3*-methylindole pair PMFs in water and 6.9 M urea solution. While urea causes the free energy of the contact pair to shift up slightly for toluene-toluene pair interactions, the contact pair for 3-methylindole – 3-methylindole pair interactions shifts out to slightly larger distance. The second, solvent-separated minimum is stabilized by urea for both systems. The association constant, defined as

$$K_a = 4\pi \int_0^{14\text{Å}} r^2 \exp\left[-w(r) / k_B T\right] dr \tag{6.2}$$

increases with 4 % (toluene) and 3 % (3-methylindole) relative to water. These relatively small changes happen because the contributions of the destabilized CP and stabilized SSP largely compensate. The fraction of solvent-separated states (obtained by using integration limits 0.8–1.1 nm for toluene and 0.9–1.2 nm for 3-methylindole) increases relative to water with 11% for both toluene and 3-methylindole.

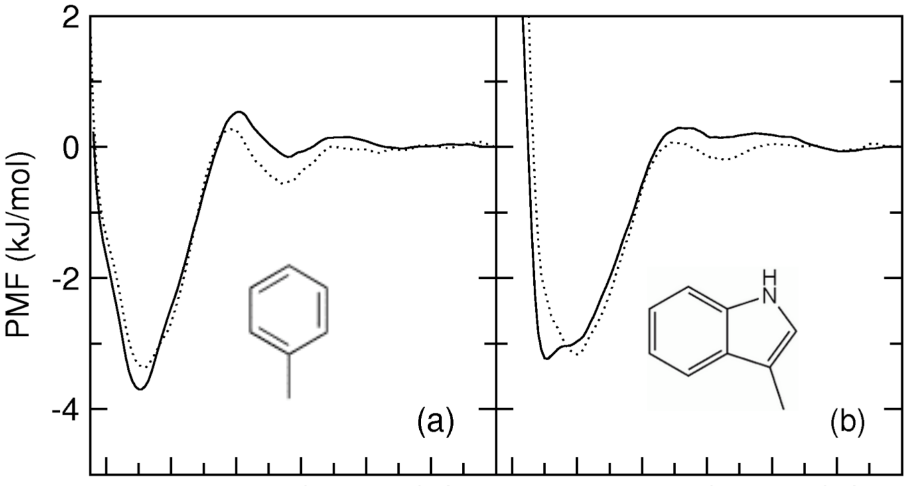


*Figure 6.3 PMFs for toluene-toluene (a) and 3-methylindole – 3-methylindole (b) pair interactions in SPC/E water (solid line) and 6.9M KBFF urea solution (dotted lines) at 298 K and 1 atm.*

Our observations can be explained on the basis of preferential urea-solute interactions. Figure 6.4 shows number density maps of water and urea molecules for the toluene. In the case of *neo*-pentane, 3-methylindole at the CP and SSP, these show a similar trend with toluene for the number density. Figure 6.5 shows snapshots for *neo*-pentane – *neo*-pentane (a), toluene – toluene (b), 3-methylindole – 3- methylindole (c) at the SSP in aqueous urea solution. In the fig. 6.4 and fig. 6.5, urea preferentially interacts with the nonpolar solutes.

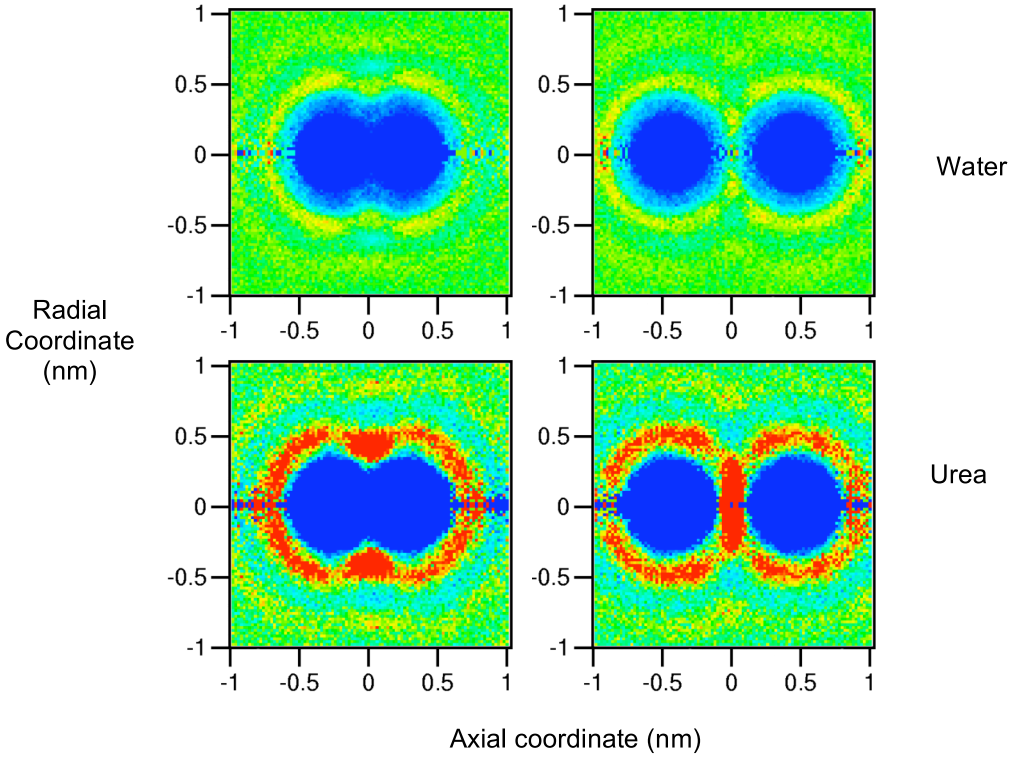


*Figure 6.4 Urea and water number density maps for toluene-toluene at the CP (left) and SSP (right) based on the KBFF/SPC/E model. Blue: zero density, green: average bulk solution density, red: twice the average bulk density. Notice that urea preferentially bridges between the hydrophobic solutes in the solvent-separated state (below panel right).*

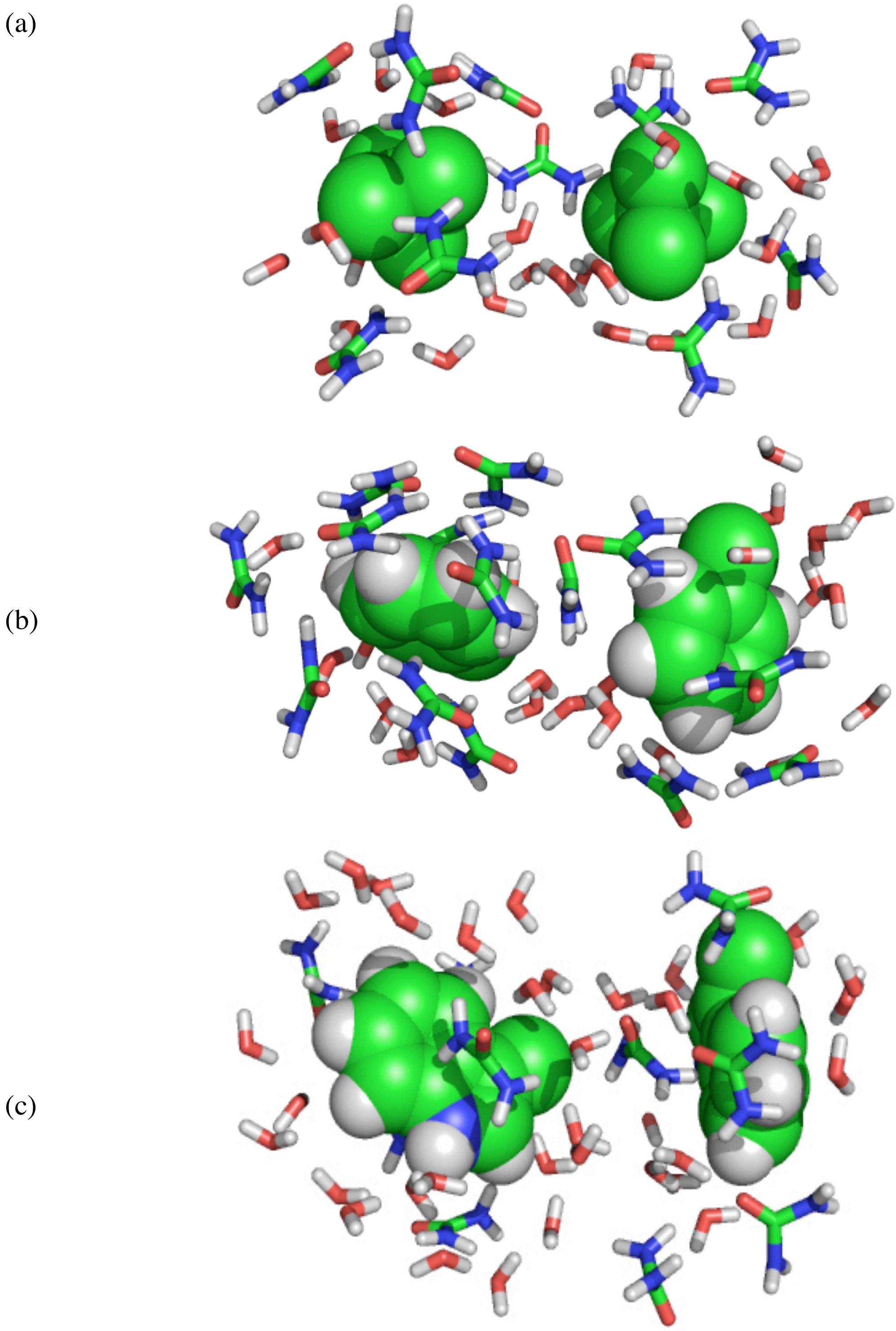


*Figure 6.5 Snapshots for the neo-pentane – neo-pentane (a), toluene – toluene (b), 3-methylindole – 3- methylindole (c) at the solvent separated distance (9.4Å, 9.3Å, 10.5Å) in aqueous urea solution. Each snapshot shows the urea and water configuration around two solutes. All snapshots were obtained from urea (KBFF) and water (SPC/E) model.*

The SSP preferably contains urea molecules interstitial to the solute pair (figs. 6.4 and 6.5). The free energy minimum corresponding to the urea-separated neopentane pair (fig. 6.2) shifts out to larger distances because urea molecules have a larger excluded volume than water molecules. Trzesniak *et al.*[42] argued that preferential urea-hydrocarbon interactions are driven by dispersion energy. The urea-separated pair is stabilized (the free energy minimum is deeper than in water) because the interstitial urea molecules interact through dispersion forces with two nonpolar solutes at the same time.

## 6.4 Implications for protein denaturation

Our results suggest that, upon solvent exposure of the protein core, urea denaturing proceeds by swelling the protein through formation of urea-separated nonpolar contacts. Urea hydrogen bonding with peptide groups likely favours open denatured states, however, thermodynamically stable contact- and urea-separated pairs of nonpolar residues prevent this state of being reached. Hence, relatively compact denatured states with residual hydrophobic clustering may form resulting from the equilibrium between these competing forces. In this study we have not addressed urea-modulated polar- and charged interaction types important for protein folding. Also, we limited our attention to pairwise interactions. Thermodynamic behavior of proteins may well arise from many-body effects and a cooperative interplay of several interactions. Notwithstanding these obvious limitations of the present study, we believe our simulation results may provide a physical reasoning for experimental observations in which the urea denatured states of several proteins have been found to be relatively compact and to contain residual hydrophobic clustering[168, 169] or even elements of native like topology.[170, 171]

## 6.5 Outlook and conclusion

Our results give some important information for protein denaturation induced by urea. We have shown that urea stabilized hydrophobic residues in our simulations. We also found that urea contributes to swell the system by bridging the two molecules.

There are two main conclusions. First, compared to pure water, urea stabilized the hydrophobic interaction. Second, urea molecules play an important role in the interstitial region between nonpolar solutes. Urea molecules enter to the interstitial region and, since urea is bigger than water molecule, lead to an increased inter-molecular distance. However, the system is not fully swollen but rather becomes more stable due to the presence of urea as a bridge.

The current work is limited to the simple nonpolar residues which is just a starting point for a protein denaturation study. In the future, it will be worth investigating the polar and charged residue interactions in various solvent environments and further, it can be extended to the studies of residues attached to polypeptide chains.

## Chapter 7

# Conclusions

This thesis describes the thermodynamics and structures of a selection of liquid mixtures. In this context, we performed a force field development for *tert*-butanol (TBA) and water mixtures, and validated this force field against other force fields and experiments for its thermodynamic properties. Using this new force field, we studied the molecular aggregation properties in TBA/water mixtures (298 *K*, 1 atm.), and discussed the aggregation in relation to the non-ideal thermodynamic properties of this mixture. Based on the validated TBA/water force field and the established knowledge on this cosolvent/water mixture, we could also study the hydrophobic solvation of gaseous solutes. In this context we were able to reproduce the experimental methane solvation thermodynamics and provide a detailed analysis of the solvation of methane in this cosolvent/water mixture. In addition to hydrophobic solvation we examined hydrophobic interactions (*i.e.,* association of two nonpolar solutes in an aqueous environment) in pure water and an urea/water mixture with the aim to better understand the effects of the urea cosolvent on the strength of the hydrophobic interaction.

The first part of this thesis deals with TBA/water mixtures and starts with the development of a new nonpolarizable tertiary-butanol (TBA) force field based on Kirkwood-Buff integrals (chapter 3). This model, combined with the SPC water model, performs very well over the full range of TBA mole fractions. Using our new force field (LV model), we have investigated thermodynamics and structures of TBA/water mixtures over a wide range of

TBA mole fractions, as mentioned in chapter 3. TBA and water mixtures are an important test system for understanding the aqueous solvation properties of small amphiphiles, as TBA contains the largest hydrophobic moiety in the series of alcohols remaining fully miscible with water. However, TBA and water mixtures show highly nonideal solution behavior in dilute TBA solutions. In the water-rich region (TBA mole fractions smaller than 0.2), strong alcohol-alcohol aggregation is observed, while water-water aggregation occurs over the full range of compositions. The hydrophobic interactions between TBA molecules are the main cause of aggregation of these alcohol molecules. The LV model (TBA) developed in this work correctly predicts the aggregation behavior of the aqueous mixtures and reproduces the experimental solution activities, as shown in figs. 3.7 and 3.8. Results obtained for other properties (*e.g.,* partial molar volumes, isothermal compressibilities) reproduce the experimental trends of solution composition. However, the solution densities and static dielectric permittivities are underestimated in our model.

The TBA model parameterized in this work was next used in a study of hydrophobic solvation in various TBA/water mixtures (chapter 4). Here, aspects of preferential interaction of the TBA cosolvent molecules with the hydrophobic solute (methane) were examined. Experimental data for the methane solvation free energy, enthalpy, and entropy were compared with our simulation results based on three nonpolarizable TBA models (LV, GROMOS, OPLS) and two rigid nonpolarizable water models (SPC, TIP4P). Among the different combinations of TBA and water models, the LV/SPC had the best agreement with experimental data for solvation free energy, enthalpy and entropy. We concluded that the observed methane solubility is mainly due to entropic contributions in the dilute TBA mole fractions and energetic contributions in the larger TBA mole fractions (fig. 4.4). We also predicted the preferential solvation of methane by TBA clusters (figs. 4.2 and 4.3). TBA clusters of LV/SPC model in the dilute TBA concentration remain sufficiently hydrated to keep the system miscible. However, water molecules are expelled from the TBA clusters by the insertion of methane (table 4.5 and fig. 4.6). Moreover, we showed that details of solvent-solvent interactions are more accurately probed in the solvation entropies and enthalpies than in the free energies. The latter is important for testing and developing force fields that describe liquid solvents.

A further goal of this thesis was to understand association of hydrophobic molecules and their thermodynamics in aqueous urea solution. We simulated the potentials of mean force (PMFs) for *neo*-pentane - *neo*-pentane in 6.9 M aqueous urea solution and in pure water using molecular dynamics in chapter 5 and 6. We presented a study of the enthalpic and entropic contributions to these PMFs. In order to understand the main driving forces for thermodynamically favourable *neo*-pentane contact distances and solvent-separated distances in urea solution, we discussed solvation free energy, enthalpy and entropy change of solute-solvent and solvent-solvent in chapter 5. For hydrophobic interactions in aqueous urea, the enthalpy-entropy compensating contribution of the solvent reorganization energy is large and dominates the enthalpy and entropy change upon decreasing the solute-solute distance. In the contact distance of the *neo*-pentane – *neo*-pentane potential of mean force in 6.9 M urea, the enthalpy (and entropy) is reduced relative to what is observed in pure water, because solvent-solvent hydrogen bonds involving urea molecules become stronger. The *neo*-pentane solvation shells preferentially contain urea molecules due to the dominance of the Van der Waals interactions (figs. 5.4 and 6.4).

Although we investigated interactions between pairs of small hydrophobic solutes only, the above results give some important information on the mechanisms of protein denaturation. Chapter 6 was focused on the protein denaturation induced by urea. By comparing two nonpolar solutes (*neo*-pentane – *neo*-pentane, toluene-toluene, 3-methylindole – 3-methylindole) potentials of mean force (PMFs) in pure water and in aqueous 6.9 M urea solution, we found that the urea molecules bias the PMF minima in an unexpected way. While the first PMF minimum (contact pair) in urea solution has the same shape and depth as the corresponding minimum in water, the second minimum (solvent-separated pair) in urea solution is broader, deeper and shifted out to a slightly larger distance (figs. 6.2 and 6.3). This free energy minimum, characterized at the molecular scale by the solute pair being bridged by a urea molecule, is due to preferential urea - solute interaction. We found evidence that urea molecules act as a bridge of hydrophobic molecule pairs (fig. 6.5) and stabilize the interaction of two hydrophobic molecules. This observation is in contrast with the widely accepted view that the urea cosolvent causes solubilization of hydrophobic contacts.

The fact that urea preferentially interacts with the nonpolar solutes and keeps them together at a slightly larger spatial separation than in pure water, provides an indication that urea denatures proteins by swelling their nonpolar interior.

# Bibliography

[1] E. Matteoli, L. Lepori, *Journal of Chemical Physics* **1984**, *80*, 2856.

[2] M. J. Blandamer, J. Burgess, A. Cooney, H. J. Cowles, I. M. Horn, K. J. Martin, K. W. Morcom, P. Warrick, *Journal of the Chemical Society-Faraday Transactions* **1990**, *86*, 2209.

[3] J. G. Kirkwood, F. P. Buff, *Journal of Chemical Physics* **1951**, *19*, 774.

[4] S. Weerasinghe, P. E. Smith, *Journal of Chemical Physics* **2003**, *118*, 10663.

[5] S. Weerasinghe, P. E. Smith, *Journal of Physical Chemistry B* **2005**, *109*, 15080.

[6] S. Weerasinghe, P. E. Smith, *Journal of Physical Chemistry B* **2003**, *107*, 3891.

[7] S. Weerasinghe, P. E. Smith, *Journal of Chemical Physics* **2003**, *119*, 11342.

[8] S. Weerasinghe, P. E. Smith, *Journal of Chemical Physics* **2004**, *121*, 2180.

[9] A. Ben-Naim, *Journal of Chemical Physics* **1977**, *67*, 4884.

[10] M. C. A. Donkersloot, *Journal of Solution Chemistry* **1979**, *8*, 293.

[11] K. J. Patil, *Journal of Solution Chemistry* **1981**, *10*, 315.

[12] E. Matteoli, *Journal of Physical Chemistry B* **1997**, *101*, 9800.

[13] K. Nishikawa, *Chemical Physics Letters* **1986**, *132*, 50.

[14] H. Hayashi, K. Nishikawa, T. Iijima, *Journal of Physical Chemistry* **1990**, *94*, 8334.

[15] M. Misawa, K. Yoshida, *Journal of the Physical Society of Japan* **2000**, *69*, 3308.

[16] L. Almásy, L. Cser, G. Jancso, *Journal of Molecular Liquids* **2002**, *101*, 89.

[17] S. Dixit, J. Crain, W. C. K. Poon, J. L. Finney, A. K. Soper, *Nature* **2002**, *416*, 829.

[18] R. D. Mountain, *Journal of Physical Chemistry A* **1999**, *103*, 10744.

[19] A. Perera, F. Sokolic, *Journal of Chemical Physics* **2004**, *121*, 11272.

[20] A. Perera, F. Sokolic, L. Almasy, Y. Koga, *Journal of Chemical Physics* **2006**, *124*.

[21] A. Perera, F. Sokolic, L. Almasy, P. Westh, Y. Koga, *Journal of Chemical Physics* **2005**, *123*.

[22] M. E. Lee, N. F. A. van der Vegt, *Journal of the American Chemical Society* **2006**, *128*, 4948.

[23] M. E. Lee, N. F. A. van der Vegt, *Journal of Chemical Theory and Computation* **2006**, Accepted.

[24] A. Ben-Naim, S. Baer, *Transactions of the Faraday Society* **1964**, *60*, 1736.

[25] A. Ben-Naim, *Journal of Physical Chemistry* **1967**, *71*, 4002.

[26] A. Ben-Naim, G. Moran, *Transactions of the Faraday Society* **1965**, *61*, 821.

[27] A. Ben-Naim, *Journal of Physical Chemistry* **1968**, *72*, 2998.

[28] K. Nakanishi, K. Ikari, S. Okazaki, H. Touhara, *Journal of Chemical Physics* **1984**, *80*, 1656.

[29] H. Tanaka, K. Nakanishi, *Fluid Phase Equilibria* **1993**, *83*, 77.

[30] H. Tanaka, K. Nakanishi, H. Touhara, *Journal of Chemical Physics* **1984**, *81*, 4065.

[31] R. Noto, V. Martorana, A. Emanuele, S. L. Fornili, *Journal of the Chemical Society-Faraday Transactions* **1995**, *91*, 3803.

[32] P. G. Kusalik, A. P. Lyubartsev, D. L. Bergman, A. Laaksonen, *Journal of Physical Chemistry B* **2000**, *104*, 9533.

[33] B. J. Bennion, V. Daggett, *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 5142.

[34] D. B. Wetlaufer, R. L. Coffin, S. K. Malik, L. Stoller, *Journal of the American Chemical Society* **1964**, *86*, 508.

[35] J. A. Rupley, *Journal of Physical Chemistry* **1964**, *68*, 2002.

[36] G. G. Hammes, P. R. Schimmel, *Journal of the American Chemical Society* **1967**, *89*, 442.

[37] R. A. Kuharski, P. J. Rossky, *Journal of the American Chemical Society* **1984**, *106*, 5794.

[38] S. Shimizu, H. S. Chan, *Proteins-Structure Function and Genetics* **2002**, *49*, 560.

[39] E. M. Duffy, P. J. Kowalczyk, W. L. Jorgensen, *Journal of the American Chemical Society* **1993**, *115*, 9271.

[40] M. Ikeguchi, S. Nakamura, K. Shimizu, *Journal of the American Chemical Society* **2001**, *123*, 677.

[41] A. Caballero-Herrera, K. Nordstrand, K. D. Berndt, L. Nilsson, *Biophysical Journal* **2005**, *89*, 842.

[42] D. Trzesniak, N. F. A. van der Vegt, W. F. van Gunsteren, *Physical Chemistry Chemical Physics* **2004**, *6*, 697.

[43] A. K. Soper, E. W. Castner, A. Luzar, *Biophysical Chemistry* **2003**, *105*, 649.

[44] A. Idrissi, *Spectrochimica Acta Part a-Molecular and Biomolecular Spectroscopy* **2005**, *61*, 1.

[45] T. Ishida, P. J. Rossky, E. W. Castner, *Journal of Physical Chemistry B* **2004**, *108*, 17583.

[46] R. D. Mountain, D. Thirumalai, *Journal of Physical Chemistry B* **2004**, *108*, 6826.

[47] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller, *Journal of Chemical Physics* **1953**, *21*, 1087.

[48] M. P. Allen, D. J. Tildesley, *Computer Simulation of Liquids*, Oxford University Press Ins, New York, **1987**.

[49] I. Cacelli, G. Cinacchi, G. Prampolini, A. Tani, *Journal of the American Chemical Society* **2004**, *126*, 14278.

[50] P. Linse, S. Engstrom, B. Jonsson, *Chemical Physics Letters* **1985**, *115*, 95.

[51] P. Linse, *Journal of the American Chemical Society* **1984**, *106*, 5425.

[52] M. J. Hwang, T. P. Stockfisch, A. T. Hagler, *Journal of the American Chemical Society* **1994**, *116*, 2515.

[53] A. T. Hagler, E. Huler, S. Lifson, *Journal of the American Chemical Society* **1974**, *96*, 5319.

[54] A. T. Hagler, S. Lifson, *Acta Crystallographica Section B-Structural Science* **1974**, *B 30*, 1336.

[55] A. T. Hagler, P. Dauber, S. Lifson, *Journal of the American Chemical Society* **1979**, *101*, 5131.

[56] A. Warshel, S. Lifson, *Journal of Chemical Physics* **1970**, *53*, 582.

[57] J. L. M. Dillen, *Journal of Computational Chemistry* **1990**, *11*, 1125.

[58] L. Verlet, *Physical Review* **1967**, *159*, 98.

[59] R. W. Hockney, *Methods in Computational Physics, Vol. 9*, **1970**.

[60] W. C. Swope, H. C. Andersen, P. H. Berens, K. R. Wilson, *Journal of Chemical Physics* **1982**, *76*, 637.

[61] D. Beeman, *Journal of Computational Physics* **1976**, *20*, 130.

[62] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, NJ, **1971**.

[63] P. P. Ewald, *Annalen Der Physik* **1921**, *64*, 253.

[64] S. W. de Leeuw, J. W. Perram, E. R. Smith, *Proc. R. Soc. London.* **1980**, *A373*, 27.

[65] P. E. Smith, W. F. van Gunsteren, in *Computer Simulation of Biomolecular Systems* (Eds.: W. F. van Gunsteren, P. K. Weiner, A. J. Wilkinson), ESCOM, Leiden, **1993**, pp. 182.

[66] T. Darden, D. York, L. Pedersen, *Journal of Chemical Physics* **1993**, *98*, 10089.

[67] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, L. G. Pedersen, *Journal of Chemical Physics* **1995**, *103*, 8577.

[68] L. Onsager, *Journal of the American Chemical Society* **1936**, *58*, 1486.

[69] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, J. R. Haak, *Journal of Chemical Physics* **1984**, *81*, 3684.

[70] S. Nose, M. L. Klein, *Molecular Physics* **1983**, *50*, 1055.

[71] S. Nose, *Molecular Physics* **1984**, *52*, 255.

[72] M. Parrinello, A. Rahman, *Physical Review Letters* **1980**, *45*, 1196.

[73] M. Parrinello, A. Rahman, *Journal of Applied Physics* **1981**, *52*, 7182.

[74] H. C. Andersen, *Journal of Chemical Physics* **1980**, *72*, 2384.

[75] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *Journal of Computational Physics* **1977**, *23*, 327.

[76] B. Widom, *Journal of Chemical Physics* **1963**, *39*, 2808.

[77] B. Widom, *Journal of Physical Chemistry* **1982**, *86*, 869.

[78] D. Frenkel, B. Smit, *Understanding Molecular Simulation*, Academic Press, New York, **1996**.

[79] M. Mezei, D. L. Beveridge, *Ann. N. Y. Acad. Sci.* **1986**, *482*, 1.

[80] D. A. Pearlman, P. A. Kollman, *Journal of Chemical Physics* **1989**, *90*, 2460.

[81] W. L. Jorgensen, *Journal of Physical Chemistry* **1986**, *90*, 1276.

[82] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, H. J., in *Intermolecular Forces* (Ed.: B. Pullman), Reidel, Dordrecht, **1981**, pp. 331.

[83] A. J. Cross, *Chemical Physics Letters* **1986**, *128*, 198.

[84] D. Chandler, *Introduction to Modern Statistical Mechanics*, Oxford University press, New York, **1987**.

[85] A. Ben-Naim, *Statistical Thermodynamics for Chemists and Biochemists*, Plenum Press, New York, **1992**.

[86] P. W. Atkins, *Physical Chemistry*, fifth ed., Oxford University Press, **1994**.

[87] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, *Journal of the American Chemical Society* **1995**, *117*, 5179.

[88] W. L. Jorgensen, D. S. Maxwell, J. TiradoRives, *Journal of the American Chemical Society* **1996**, *118*, 11225.

[89] C. Oostenbrink, A. Villa, A. E. Mark, W. F. Van Gunsteren, *Journal of Computational Chemistry* **2004**, *25*, 1656.

[90] M. G. Martin, J. I. Siepmann, *Journal of Physical Chemistry B* **1998**, *102*, 2569.

[91] M. Fioroni, K. Burger, A. E. Mark, D. Roccatano, *Journal of Physical Chemistry B* **2001**, *105*, 10967.

[92] D. P. Geerke, C. Oostenbrink, N. F. A. van der Vegt, W. F. van Gunsteren, *Journal of Physical Chemistry B* **2004**, *108*, 1436.

[93] G. Milano, F. Muller-Plathe, *Journal of Physical Chemistry B* **2004**, *108*, 7415.

[94] B. Chen, J. J. Potoff, J. I. Siepmann, *Journal of Physical Chemistry B* **2001**, *105*, 3093.

[95] I. Brovehenko, B. Guillot, *Fluid Phase Equilibria* **2001**, *183*, 311.

[96] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *Journal of Physical Chemistry* **1987**, *91*, 6269.

[97] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *Journal of Chemical Physics* **1983**, *79*, 926.

[98] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, I. G. Tironi, *Biomolecular Simulation: The GROMOS96 Manual and User Guide*, vdf Hochschulverslag, ETH Zürich, Switzerland, **1996**.

[99] W. R. P. Scott, P. H. Hunenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Kruger, W. F. van Gunsteren, *Journal of Physical Chemistry A* **1999**, *103*, 3596.

[100] L. D. Schuler, X. Daura, W. F. Van Gunsteren, *Journal of Computational Chemistry* **2001**, *22*, 1205.

[101] R. Walser, A. E. Mark, W. F. van Gunsteren, M. Lauterbach, G. Wipff, *Journal of Chemical Physics* **2000**, *112*, 10450.

[102] K. Nishikawa, Y. Kodera, T. Iijima, *Journal of Physical Chemistry* **1987**, *91*, 3694.

[103] Y. Koga, W. W. Y. Siu, T. Y. H. Wong, *Journal of Physical Chemistry* **1990**, *94*, 7700.

[104] K. Nakanish, N. Kato, M. Maruyama, *Journal of Physical Chemistry* **1967**, *71*, 814.

[105] A. Ben-Naim, Y. Marcus, *Journal of Chemical Physics* **1984**, *81*, 2016.

[106] J. G. Kirkwood, *Journal of Chemical Physics* **1935**, *3*, 300.

[107] J. W. Smith, *Electric Dipole Moments*, Butterworths Scientific, London, **1955**.

[108] J. M. Prausnutz, R. N. Lichtenthaler, E. Gomes de Azevedo, *Molecular Thermodynamics of Fluid Phase Equilibria, Vol. Prentice Hall*, New Jersey, **1999**.

[109] H. Kubota, Y. Tanaka, T. Makita, *International Journal of Thermophysics* **1987**, *8*, 47.

[110] A. C. Brown, D. J. G. Ives, *Journal of the Chemical Society* **1962**, 1608.

[111] A. Glättli, X. Daura, W. F. van Gunsteren, *Journal of Chemical Physics* **2002**, *116*, 9811.

[112] M. Neumann, *Molecular Physics* **1983**, *50*, 841.

[113] D. T. Bowron, J. L. Finney, A. K. Soper, *Journal of Physical Chemistry B* **1998**, *102*, 3551.

[114] D. T. Bowron, S. D. Moreno, *Journal of Chemical Physics* **2002**, *117*, 3753.

[115] N. F. A. van der Vegt, W. F. van Gunsteren, *Journal of Physical Chemistry B* **2004**, *108*, 1056.

[116] T. Rispens, M. F. Lensink, H. J. C. Berendsen, J. Engberts, *Journal of Physical Chemistry B* **2004**, *108*, 5483.

[117] C. Tanford, *Advances in Protein Chemistry* **1970**, *24*, 1.

[118] J. F. Carpenter, J. H. Crowe, *Cryobiology* **1988**, *25*, 459.

[119] G. F. Xie, S. N. Timasheff, *Biophysical Chemistry* **1997**, *64*, 25.

[120] Y. L. Wang, B. X. Han, H. K. Yan, R. L. Liu, *Thermochimica Acta* **1995**, *253*, 327.

[121] R. D. Lins, C. S. Pereira, P. H. Hunenberger, *Proteins-Structure Function and Bioinformatics* **2004**, *55*, 177.

[122] B. Hess, C. Holm, N. van der Vegt, *Journal of Chemical Physics* **2006**, *124*.

[123] N. F. A. van der Vegt, D. Trzesniak, B. Kasumaj, W. F. van Gunsteren, *Chemphyschem* **2004**, *5*, 144.

[124] N. F. A. van der Vegt, M. E. Lee, D. Trzesniak, W. F. van Gunsteren, *Journal of Physical Chemistry B* **2006**, *110*, 12852.

[125] T. A. Özal, N. F. A. van der Vegt, *Journal of Physical Chemistry B* **2006**, *110*, 12104.

[126] M. E. Lee, N. F. A. van der Vegt, *Journal of Chemical Physics* **2005**, *122*.

[127] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *Journal of Computational Chemistry* **2005**, *26*, 1701.

[128] W. G. Hoover, *Physical Review A* **1985**, *31*, 1695.

[129] C. A. Wick, J. I. Siepmann, M. R. Schure, *Journal of Physical Chemistry B* **2003**, *107*, 10623.

[130] A. Ben-Naim, *Solvation Thermodynamics*, Plenum Press, New York, **1987**.

[131] G. I. Makhatadze, P. L. Privalov, *Journal of Molecular Biology* **1993**, *232*, 639.

[132] H. A. Yu, M. Karplus, *Journal of Chemical Physics* **1988**, *89*, 2366.

[133] K. A. Dill, S. Bromberg, *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*, Garland Science, New York, **2003**.

[134] D. E. Smith, A. D. J. Haymet, *Journal of Chemical Physics* **1993**, *98*, 6445.

[135] S. W. Rick, B. J. Berne, *Journal of Physical Chemistry B* **1997**, *101*, 10488.

[136] N. T. Southall, K. A. Dill, *Biophysical Chemistry* **2002**, *101*, 295.

[137] J. F. Brandts, L. J. Hunt, *Journal of the American Chemical Society* **1967**, *89*, 4826.

[138] Landolt-Börnstein, *New Series, Group IV, Springer-Verlag* Spinger-Verlag, New York, **1977**.

[139] A. Ben-Naim, *Journal of Physical Chemistry* **1978**, *82*.

[140] L. R. Pratt, D. Chandler, *Journal of Chemical Physics* **1977**, *67*, 3683.

[141] B. Guillot, Y. Guissani, *Journal of Chemical Physics* **1993**, *99*, 8075.

[142] H. Qian, J. Hopfield, *Journal of Chemical Physics* **1996**, *105*, 9292.

[143] E. Gallicchio, M. M. Kubo, R. M. Levy, *Journal of physical Chemistry B* **2000**, *104*, 6271.

[144] T. Lazaridis, *Journal of Physical Chemistry B* **2000**, *104*, 4964.

[145] M. T. Stone, P. J. in't Veld, I. C. Sanchez, *Molecular Physics* **2002**, *100*, 2773.

[146] P. Schravendijk, N. F. A. van der vegt, *Journal of Chemical Theory and Computation* **2005**, *1*, 643.

[147] I. C. Sanchez, T. M. Truskett, P. J. in't Veld, *Journal of Physical Chemistry B* **1999**, *103*, 5106.

[148] D. Ben-Amotz, F. O. Raineri, G. Stell, *Journal of Physical Chemistry B* **2005**, *109*, 6866.

[149] C. Peter, C. Oostenbrink, A. van Dorp, W. F. van Gunsteren, *Journal of Chemical Physics* **2004**, *120*, 2652.

[150] N. F. A. van der Vegt, *Journal of Membrane Science* **2002**, *205*, 125.

[151] N. Matubayasi, L. H. Reed, R. M. Levy, *Journal of Physical Chemistry* **1994**, *98*, 10640.

[152] A. Ben-Naim, *Journal of Chemical Physics* **1971**, *54*, 3682.

[153] K. A. T. Silverstein, A. D. J. Haymet, K. A. Dill, *Journal of the American Chemical Society* **1998**, *120*, 3166.

[154] S. W. Rick, *Journal of Physical Chemistry B* **2003**, *107*, 9853.

[155] G. Graziano, *Journal of Physical Chemistry B* **2001**, *105*, 2632.

[156] G. Barone, E. Rizzo, Vitaglia.V, *Journal of Physical Chemistry* **1970**, *74*, 2230.

[157] M. J. Mastroia, M. J. Pikal, Lindenba.S, *Journal of Physical Chemistry* **1972**, *76*, 3050.

[158] F. Franks, M. D. Pedley, *Journal of the Chemical Society-Faraday Transactions I* **1981**, *77*, 1341.

[159] D. O. V. Alonso, K. A. Dill, *Biochemistry* **1991**, *30*, 5974.

[160] A. Wallqvist, D. G. Covell, D. Thirumalai, *Journal of the American Chemical Society* **1998**, *120*, 427.

[161] F. Vanzi, B. Madan, K. Sharp, *Journal of the American Chemical Society* **1998**, *120*, 10748.

[162] R. D. Mountain, D. Thirumalai, *Journal of the American Chemical Society* **2003**, *125*, 1950.

[163] D. Tobi, R. Elber, D. Thirumalai, *Biopolymers* **2003**, *68*, 359.

[164] H. S. Frank, F. J. Franks, *Journal of Chemical Physics* **1968**, *48*, 4746.

[165] C. Chipot, R. Jaffe, B. Maigret, D. A. Pearlman, P. A. Kollman, *Journal of the American Chemical Society* **1996**, *118*, 11217.

[166] R. Chelli, F. L. Gervasio, P. Procacci, V. Schettino, *Journal of the American Chemical Society* **2002**, *124*, 6133.

[167] E. M. Duffy, D. L. Severance, W. L. Jorgensen, *Israel Journal of Chemistry* **1993**, *33*, 323.

[168] D. Neri, M. Billeter, G. Wider, K. Wuthrich, *Science* **1992**, *257*, 1559.

[169] J. Klein-Seetharaman, M. Oikawa, S. B. Grimshaw, J. Wirmer, E. Duchardt, T. Ueda, T. Imoto, L. J. Smith, C. M. Dobson, H. Schwalbe, *Science* **2002**, *295*, 1719.

[170] D. Shortle, M. S. Ackerman, *Science* **2001**, *293*, 487.

[171] K. W. Plaxco, M. Gross, *Nature Structural Biology* **2001**, *8*, 659.