Interactions in aqueous salt solutions

Atomistic modelling versus experiment

Aspelin, Vidar

2020

Document Version:
Publisher's PDF, also known as Version of record

Interactions in aqueous salt solutions

Atomistic modelling versus experiment

VIDAR ASPELIN | DIVISION OF THEORETICAL CHEMISTRY | LUND UNIVERSITY

Interactions in aqueous salt solutions

Atomistic modelling versus experiment

Vidar Aspelin

LUND
UNIVERSITY

LICENTIATE THESIS

bye due permission of the Faculty of Science, Lund University, Sweden.
To be defended on May 26 2020 at 10:15 at Lund Institute of advanced Neutron and X-ray Science, Delta 5, 5th floor, IDEON building, Scheelevägen 19, 223 70 Lund.

Faculty opponent
Professor Jiří Kolafa, University of Chemistry and Technology, Prague, Czechia

Organization	Document name
LUND UNIVERSITY Division of Theoretical Chemistry Department of Chemistry Box 124 SE–221 00 Lund, Sweden	**LICENTIATE THESIS** Date of issue 2020-05-26 Sponsoring organization
Author(s) Vidar Aspelin	

Title and subtitle

Interactions in aqueous salt solutions: Atomistic modelling versus experiment

Abstract

The research in this thesis investigates how atoms and molecules constituting aqueous salt solutions interact using computational approaches, namely Monte Carlo simulations and molecular dynamics simulations.

In the first paper, an atomistic model was developed for aqueous solutions of sodium thiocyanate and potassium thiocyanate (NaSCN and KSCN). The model reproduced several experimentally measured thermodynamic properties in bulk solution and at the air-water interface. The model further gave insight into cation specific effects on the thiocyanate anion. K^+ was found to show a preferential attraction to the S atom of SCN^-, forming a diffuse first coordination shell around the atom. Na^+, on the other hand, showed a relatively stronger preferential interaction with the N atom of SCN^-, resulting in a more distinct first coordination shell. At high salt concentration, the difference in cation-anion interactions resulted in NaSCN forming larger and more closely packed clusters than KSCN.

In the second paper, a method handling long-range electrostatic interactions was developed, using a short-ranged potential. The method is mathematically exact and has physical foundation in that it cancels an arbitrary number of electrostatic moments at the cutoff. With an appropriate choice of how many moments that are cancelled, the method was shown to produce accurate results compared to Ewald and particle mesh Ewald. The method is advantageous in that it scales with $\mathcal{O}(N)$, compared with Ewald and particle mesh Ewald, scaling with $\mathcal{O}(N^{3/2})$ and $\mathcal{O}(N\log(N))$, respectively.

Key words

statistical thermodynamics, molecular simulation, MD simulations, MC simulations, Kirkwood-Buff solution theory, Hofmeister series

Classification system and/or index terms (if any)

Supplementary bibliographical information	Language
	English
ISSN and key title	ISBN
	978-91-7422-743-7 (print) 978-91-7422-744-4 (pdf)
Recipient's notes	Number of pages: 138 / Price
	Security classification

Interactions in aqueous salt solutions

Atomistic modelling versus experiment

Vidar Aspelin

LUND
UNIVERSITY

LICENTIATE THESIS

Interactions in aqueous salt solutions: Atomistic modelling versus experiment

Cover illustration front: Sea waves. Drawing by Vidar Aspelin.

Funding information: The thesis work was financially supported by The Swedish Research Council.

Division of Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University

ISBN: 978-91-7422-743-7 (print)
ISBN: 978-91-7422-744-4 (pdf)

Printed in Sweden by Media-Tryck, Lund University, Lund 2020

Media-Tryck is a Nordic Swan Ecolabel certified provider of printed material. Read more about our environmental work at www.mediatryck.lu.se

MADE IN SWEDEN

Contents

List of papers

This thesis is based on the following papers, referred to by their Roman numerals:

I **Specific Cation Effects on SCN^- in Bulk Solution and at the Air-Water Interface**
G. Tesei, **V. Aspelin**, M. Lund
The Journal of Physical Chemistry B, 2018, 122(19), pp. 5094-5105

II **Generalized Moment Correction for Long-Ranged Electrostatics**
B. Stenqvist, **V. Aspelin**, M. Lund
Accepted for publication, *Journal of Chemical Theory and Computation*, 2020

Acknowledgements

First I would like to thank my supervisor Mikael, who introduced me to the field of statistical thermodynamics by taking me in as a student for a summer project. Thank you for all the guidance and fruitful discussions throughout these years. I also want to thank all the members in our group, as well as all colleagues at the division of Theoretical Chemistry and the division of Physical Chemistry for providing a friendly and creative working environment. A special thanks to my co-supervisor Björn for all the guidance and for proofreading my thesis, to Richard for giving advice on simulation-related problems, and to Samuel and Stefan for all the inspiring discussions.

In addition, I would like to thank my friends and family for always being supportive and for making me practice on explaining my research in a comprehensible way.

Lastly, I would like to thank Ule for always encouraging me and for proofreading my popular summary.

Populärvetenskaplig sammanfattning på svenska

Joner är atomer eller molekyler som har en elektrisk laddning. Utan att kanske direkt tänka på det använder vi joner till allt möjligt i vår vardag, till exempel när vi borstar tänderna eller när vi saltar vår mat. Koksalt innehåller nämligen saltet natriumklorid. Saltet i sig är kristaller i vilka de positivt laddade natriumjonerna och de negativt laddade kloridjonerna har ordnat sig i perfekt symmetri, med lika långt avstånd mellan varje jon. Kristallerna hålls ihop av starka jonbindningar, så starka att det krävs en temperatur på tusentals grader för att smälta kristallerna. Men hur kommer det sig då att vi i vanlig rumstemperatur kan slänga några saltkristaller i lite vatten och få jonerna att dela på sig? Svaret ligger i att vattenmolekylerna bildar nya starka bindningar med jonerna som frigör energi som är större än energin som krävs för att bryta bindningarna mellan jonerna i kristallen.

Syftet med den här avhandlingen är att beskriva de bindningar, eller växelverkningar, som sker mellan atomer och molekyler i kemiska system som innehåller joner och vatten. För detta ändamål har jag använt mig av något som kallas molekylsimulering, där varje atom ges parametrar som beskriver hur den växelverkar med andra atomer. Genom denna typ av simulering får man en detaljerad bild av hur atomerna eller molekylerna som utgörs av atomerna beter sig, till exempel vilka andra atomer eller molekyler de helst binder till eller hur snabbt de sprider sig. Givet att man har en verklighetstrogen modell av atomerna är simuleringarna viktiga för att på molekylnivå kunna förklara utfallet av redan utförda experiment, eller förutspå utfallet från framtida, planerade experiment. I den första artikeln i denna avhandling utvecklade vi en modell för tiocyanatjonen (SCN^-). Tiocyanatjonen är speciell då den ökar proteiners löslighet i vatten genom att binda till proteinet, och den används därför ofta i experiment som involverar proteiner. Förutom att vi utformade en robust och validerad modell, använde vi den även för att ge insikt i hur två salter som innehåller tiocyanatjonen, natrium- och kaliumtiocyanat, beter sig när de är upplösta i vatten; till exempel hur lösligheterna av de två salterna kan förklaras genom de skillnader vi observerade i växelverkningar mellan å ena sidan natrium och tiocyanat, och å andra sidan kalium och tiocyanat.

I den andra artikeln i denna avhandling utvecklade vi en ny metod för att beskriva hur atomers laddningar växelverkar på långa avstånd. Vi applicerade metoden dels på system som endast bestod av vatten, och dels på system som bestod av vatten och joner, och den visade sig ge bra resultat jämfört med andra, vedertagna metoder. Fördelen med modellen vi utvecklade är att den är mindre kostsam, i bemärkelsen att den kräver mindre datortid, än tidigare, vedertagna modeller.

För närvarande studerar vi inbindningen av kaliumjoner till kronetrar. Kronetrar är

ringformade strukturer som har visat sig kunna binda in olika metalljoner, bland vilka kaliumjonen binder in särskilt starkt. När en jon binder in ser kronetern ut som en krona på ett huvud, vilket har gett upphov till dess namn. Kronetern har sedan dess upptäckt (vilket gav ett nobelpris) varit väldigt uppmärksammad på grund av dess erkända och potentiella applikationsområden. Till exempel kan de användas för att lösa upp salter, som själva inte är lösliga, i organiska ämnen. Den molekyl vi studerade består av två kronetrar som är sammankopplade, och experiment har visat att när en av kronetrarna binder in en kaliumjon, binder en andra kaliumjon in ännu starkare till den andra kronetern. Inbindningen av den första jonen gynnar alltså inbindningen av den andra! Syftet med studien är dels att utveckla en precis modell, för att sedan kunna förklara denna observation från ett molekylärt perspektiv.

Chapter 1

Introduction

Ions are charged atoms or molecules that exist everywhere around us. They play an essential role in chemistry occurring in a wide range of environments, from the human body to marine aerosols in the atmosphere. Marine aerosols are microdroplets of seawater, which is an aqueous solution of ions, formed by the action of sea waves. The ion chemistry within these aerosols is important for reactions occuring in the atmosphere, *e.g.* ozone depletion.[1] Although the relative concentration of ions is nearly identical in seawater from different parts of the world,[2] the total concentration or the *salinity*, which measures the weight of dissolved ions per weight of seawater, can vary significantly. For example, relatively dilute seawater is found close to the North Pole whereas the Dead Sea shows exceptionally high salinity ($\sim$ 340 g/kg compared to 34.7 g/kg which is the global mean).[3] The salinity sets the conditions for life and underlying chemical processes, and to understand the chemistry of seawater is thus of importance in environmental chemistry. The purpose of this work has been to create a toolbox which can be used to describe the thermodynamics of aqueous solutions of ions on an atomistic scale, with emphasis on inorganic ions commonly found in nature. The long-term goal is to use this toolbox to study problems in environmental chemistry, for example carbon dioxide/mineral chemistry in seawater, and chemistry occurring at the air/water interface of marine aerosols. To begin with, let us define in which setting this toolbox is developed and why it becomes important.

When describing a chemical system on the atomistic level, the complexity of the problem increases rapidly with the number of atoms in the system. For example, studying the motion of particles obeying the laws of Newtonian mechanics, an exact solution can only be obtained for systems with no more than two particles.[4] Thus, trying to obtain accurate solutions for larger systems becomes an essentially impossible task if we rely on analytical solutions carried out with pen and paper. However, through

the help of computers we can simulate larger systems with approximate models and, given that the model is sufficiently accurate, get valuable information about the system. This is the idea behind molecular simulation which started to develop in the 1950's when the foundation for modern Monte Carlo simulation was laid[5] and the first molecular dynamics simulations were carried out.[6,7] Since then, the methods have been refined,[8] and due to the almost exponential increase in computer power with time, they have advanced to handle larger, and more complex systems. An advantage with studying a chemical system on the atomistic level is that we get detailed information such as interactions between single atoms, and thereby we can explain macroscopic observables with more detailed information. Yet, disadvantages are that the system size is still limited by computer power, and the model needs to be validated against experiments. It is within this framework of molecular simulation techniques that the toolbox for describing thermodynamics of ions is developed, with the purpose of providing insight into chemical phenomena on the atomic scale, helping us to understand what we observe on the macroscopic scale.

In Paper I, we develop a model, or force field, of potassium and sodium thiocyanate (NaSCN, KSCN). The thiocyanate anion, SCN^-, is of interest due to its prominent position at the extreme end of the Hofmeister series,[9] a series in which ions are ranked according to their ability to increase solubility of proteins. In other words, the thiocyanate anion has an unusually high ability to dissolve proteins. The force field is developed using Kirkwood-Buff theory (see Chapter 7), and reproduces properties measured experimentally on both the air/water-interface and in the bulk, for a range of salt concentrations. The model is further used to give insight at the molecular scale into the experimentally observed differences between the sodium and the potassium salt.

In Paper II, we develop a method for handling long-range electrostatic interactions using a short-range potential. The method is tested on systems containing only water and on systems containing water and salt, and it yields results in agreement with other, conventional methods for a range of thermodynamic properties. The advantage with the method is that it has a lower computational cost than previous, conventional methods, providing a possibility to expand the limits on time and space in simulated systems.

Chapter 2

Statistical mechanics

Statistical mechanics provides a link between the equilibrium properties of a macroscopic system, such as the temperature and the pressure, to its microscopic properties, *i.e*. its constituent particles and their interactions. Whenever it is used to describe thermodynamic properties, it is usually referred to as statistical *thermodynamics*. Whereas thermodynamics provides well-known equations relating one macroscopic property to another, statistical thermodynamics lays the theoretical foundation for describing *why* these relations are observed.[10] This chapter is meant to provide the basic theory behind statistical thermodynamics and how it is used in molecular simulation, and the interested reader is referred elsewhere[4,10] for a more thorough description.

1 Statistical thermodynamics

To understand the link between micro- and macroscopic properties, let us start with describing one of the key concepts used in statistical thermodynamics, namely the concept of *ensembles*. An ensemble is a collection of a large number of microstates, where each state is macroscopically identical to the thermodynamic system we wish to investigate, *i.e.*, they are in the same thermodynamic state. The microstates can be either quantum or classical states, and in the following they will be referred to simply as states. Even though the states are macroscopically identical, they can vary significantly on the molecular level. This stems from the fact that in general, there is an *extremely* large number of states constistent with a given thermodynamic state. Since each state is only one of all possible, it would not be very fruitful to calculate properties based on them individually. Instead, the average microscopic property among all states, the *ensemble* average, yields a measure of the macroscopic property that we are interested

in.[10] This provides the link between the micro- and macroscopic properties, so that with a *large* number of microscopic states belonging to the same thermodynamic state, we can describe and explain properties of the corresponding thermodynamic system from a molecular point of view.

In the *microcanonical* ensemble, the number of particles (N), the volume (V), and the total energy (E) are kept constant. Since the energy is kept constant, the ensemble is a representative of an *isolated* thermodynamic system, which leads us to a key postulate within statistical thermodynamics known as *equal a priori probabilities*. The postulate says that the states representing an isolated thermodynamics system are uniformly distributed. This means that in the microcanonical ensemble, all states are equally probable. The total number of states is given by the microcanonical *partition function*, $\Omega_{N,V,E}$, and it is related to the entropy, S, through the well-known formula

$$S = k_B \ln(\Omega_{N,V,E}), \tag{2.1}$$

where k_B is the Boltzmann constant. The entropy is of special relevance in the microcanonical ensemble in the sense that isolated thermodynamic systems strives to maximize the entropy. In other words, entropy is at its maximum when an isolated system is at thermodynamic equilibrium.

In the context of thermodynamics, it turns out to be more suitable to employ other types of ensembles than the microcanonical. For example, when conducting experiments it is more convenient to keep the temperature (T) constant rather than the energy. Ensemble representatives of such a thermodynamic system are the canonical ensemble in which N, V, and T are constant, and the *isothermal-isobaric* ensemble where N, the pressure P, and T are held constant. Another useful ensemble is the grand canonical ensemble in which the chemical potential μ, V, and T are fixed. In the canonical ensemble, the partition function is given by[11]

$$Q_{N,V,T} = \sum_l \Omega(E_l) \exp(-\beta E_l), \tag{2.2}$$

where $Q_{N,V,T}$ is denoting the canonical partition function, $\beta = 1/k_B T$, and the sum is over all energy levels l. $\Omega(E_l)$ introduced above is the degeneracy of the l'th energy level, *i.e.* the number of states being at energy level E_l. The exponential in the sum is called the Boltzmann factor, weighting the contribution of states to the partition function according to their energies. The states are thus no longer equally probable, since low-energy states are favored over high-energy states. Similarly to the entropy being maximized in the equilibrated microcanonical ensemble, the *Helmholtz* free

energy, A, is minimized at equilibrium in the canonical ensemble, and it is related to the canonical partition function through

$$A = E - TS = -k_B T \ln Q_{N,V,T}. \tag{2.3}$$

In the isothermal-isobaric ensemble, the partition function is given by

$$\Delta_{N,P,T} = \sum_v Q(V_v) \exp(-\beta P V_v), \tag{2.4}$$

where the contribution from states are weighted according to their energy but also according to their volume. Here, the system strives to minimize the *Gibbs* free energy, defined as

$$G = E - TS + PV = -k_B T \ln \Delta_{N,P,T}. \tag{2.5}$$

In the grand canonical ensemble, the number of particles can travel in and out of the system, which makes it an *open* system. This makes it unique compared to all previously defined ensembles which are *closed* systems since the number of particles is kept constant. The grand canonical partition function, $\Xi_{\mu,V,T}$, is defined as

$$\Xi_{\mu,V,T} = \sum_n Q(N_n) \exp(-\beta \mu N_n), \tag{2.6}$$

where it is seen to include the canonical partition function, but in addition weighs the states according to the number of particles present. This system strives to minimize the grand potential, defined as

$$\Phi = E - TS - \mu N = -k_B T \ln \Xi_{\mu,V,T}. \tag{2.7}$$

Now that we have defined the most commonly used ensembles within statistical thermodynamics, let us declare the importance of the partition function. The partition function can be seen as a weighted sum of all states that a certain thermodynamic system can be in, where different constraints on the system leads to different weights. Thus, if we know the partition function, we know every microscopic detail of the system, and can in principle calculate *any* thermodynamic property. This is possible since, if we take the canonical ensemble as an example, the probability of being in a state i is given by

$$P_i = \frac{\exp(-\beta E_i)}{Q_{N,V,T}}. \tag{2.8}$$

If we then want to calculate an arbitrary thermodynamic property X, we know that we can obtain it as

$$\langle X \rangle = \sum_i P_i X_i = \sum_i \frac{X_i \exp(-\beta E_i)}{Q_{N,V,T}} = \frac{\sum_i X_i \exp(-\beta E_i)}{Q_{N,V,T}}, \tag{2.9}$$

where the angular brackets indicates that it is the average value of X, and the sum is over all states. The partition function is a constant, and can thus be put outside the sum.

2 Classical statistical thermodynamics

Up to this point we have expressed the partition functions for all ensembles as sums over discrete energy levels, which is the quantum mechanical representation. However, in molecular simulation one most often use the classical mechanical representation, in which the states are so close in energy such that they can be considered a continuum. In a classical system, the total energy E can be replaced by the so-called Hamiltonian, $\mathcal{H}$, defined as

$$\mathcal{H} = U + \mathcal{K}, \tag{2.10}$$

where U is the potential energy, due to particles interacting, and $\mathcal{K}$ is the kinetic energy, caused by motion of the particles. The partition function is then expressed as an integral over all possible classical states, *i.e.* all possible particle positions $\mathbf{r}^N$ and momenta $\mathbf{p}^N$. The classical canonical partition function then becomes

$$Q_{class} = \frac{1}{N! h^{3N}} \int \exp[-\beta \mathcal{H}(\mathbf{r}^N, \mathbf{p}^N)] \mathrm{d}\mathbf{r}^N \mathrm{d}\mathbf{p}^N, \tag{2.11}$$

where h is Planck's constant. The kinetic part can be integrated out directly, and we can express Q_{class} as an integral over the particle positions only,

$$Q_{class} = \frac{1}{N! \Lambda^{3N}} \int \exp[-\beta U(\mathbf{r}^N)] \mathrm{d}\mathbf{r}^N, \tag{2.12}$$

where the Hamiltonian is reduced to the potential energy, which is the only energy term that depends on the particle positions. The resulting integral, $\int \exp[-\beta U(\mathbf{r}^N)] \mathrm{d}\mathbf{r}^N$,

is the *configurational integral*, in the following denoted Z. The name originates from the fact that the states integrated over are now configations, *i.e.* particle positions. The term $\Lambda = h/(2\pi m k_B T)^{1/2}$ is the de-Broglie wavelength resulting from the kinetic part of the partition function, where m is the mass.

The probability of finding a system in a certain configuration $\mathbf{r}^N$ with potential energy $U(\mathbf{r}^N)$ is thus, in analogy with Equation 2.8,

$$P(\mathbf{r}^N) = \frac{\exp[-\beta U(\mathbf{r}^N)]}{Z}, \tag{2.13}$$

and average thermodynamic properties are obtained by rewriting Equation 2.9 as

$$\langle X \rangle = \frac{\int X(\mathbf{r}^N) \exp[-\beta U(\mathbf{r}^N)] \mathrm{d}\mathbf{r}^N}{Z}. \tag{2.14}$$

Chapter 3

Intermolecular interactions

When applying statistical thermodynamics to describe properties of chemical systems, the particles in the system are generally atoms or molecules constituting several atoms. The interactions present in the system is typically divided into two parts: *intra*-molecular interactions describing covalent bonds within molecules and *inter*-molecular interactions acting between atoms in different molecules. Intermolecular interactions are generally weak compared to covalent bonds, but still important to describe the behavior of a chemical system. For example, in a system composed of water and ions, the relative strength of water-water, water-ion, and ion-ion interactions is determining how well the ions are solvated.

1 Pairwise additivity

Before digging deeper into the details of intermolecular interactions, one key approximation commonly used in molecular simulation needs to be declared. When we evaluate intermolecular interactions in a simulated system, we assume that we can describe the total potential energy as a sum of pair-interactions, *i.e.* they are *pairwise additive*. This means that we assume every interaction between two atoms to be independent of all other atoms. Accordingly, we can model intermolecular interactions using pair potentials, only depending on the distance between the two considered atoms. This assumption greatly simplifies the problem to describe a chemical system through simulation since each pair interaction can be evaluated separately and then combined to obtain the total potential energy. Nevertheless, the computational complexity is still $\mathcal{O}(N^2)$, and to further reduce it, other methods are required which will be described in Chapter 4.

2 van der Waals interaction

Three types of interactions contribute to what is collectively referred to as van der Waals (vdW) interaction, namely the dipole-dipole (Kesom), the dipole-induced dipole (Debye), and the induced dipole-induced dipole (London) interaction.[14] Common for all three is that they decay as r^{-6}, and the total vdW interaction can be expressed as

$$\begin{aligned} u^{\mathrm{vdW}}(r) &= u^{\mathrm{Keesom}}(r) + u^{\mathrm{Debye}}(r) + u^{\mathrm{London}}(r) \\ &= -\left(\frac{C^{\mathrm{Keesom}}}{r^6} + \frac{C^{\mathrm{Debye}}}{r^6} + \frac{C^{\mathrm{London}}}{r^6}\right) = -\frac{C^{\mathrm{vdW}}}{r^6}. \end{aligned} \tag{3.1}$$

Here, each C is a constant depending on the strength of the corresponding interaction, and the sum of them can be expressed as a single constant, C^{vdW}, depending on the strength of the total vdW interaction.

3 Exchange repulsion

The Pauli exclusion principle[12] states that two identical fermions cannot occupy the same quantum state. Since the constituent parts of atoms, including electrons, are all fermions, the principle implies that atoms that are coming too close to one another will experience repulsion. This is the origin of a short-range type of interaction called exchange repulsion, which, in its simplest form, can be modelled by the following pair potential,

$$u^{\mathrm{hs}}(r) = \begin{cases} \infty, & \text{if } r < \sigma \\ 0, & \text{otherwise.} \end{cases} \tag{3.2}$$

This is known as the hard-sphere (HS) potential, where r is the distance between two atoms and σ is a parameter depending on the size of the two atoms, most often calculated as the *average* diameter. The potential completely prevents atoms from approaching each other closer than their average diameter by returning infinite energy, whereas at any larger distance, the potential is zero resulting in the atoms not interacting at all (much like billiard balls on a pool table).

A pair potential that models both exchange repulsion and vdW interaction is the Lennard-Jones (LJ) potential[15]

$$u^{\mathrm{lj}}(r) = 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right]. \tag{3.3}$$

Here, the constants ϵ and σ determine the shape of the potential, where ϵ sets the strength of the interaction and σ, which depends on the size of the atoms, acts to shift the potential laterally. The first term within the parenthesis is repulsive and models exchange repulsion, whereas the second term is attractive and models vdW interaction, or a subset of the terms (Keesom, Debye, London) of the vdW interaction, depending on how one chooses the constants. The LJ is used heavily in simulations, mainly due to its computational simplicity.

It is also possible to model exchange repulsion by including the repulsive part of the LJ potential only, providing an alternative to the HS potential. One such potential is the Weeks-Chandler-Andersen reference system pair potential (WCA),[13] defined as

$$u^{\mathrm{wca}}(r) = \begin{cases} 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon, & \text{if } r < 2^{1/6}\sigma \\ 0, & \text{otherwise.} \end{cases} \tag{3.4}$$

This potential is a truncated and shifted form of the LJ potential, where the parameter ϵ sets the strength of the repulsion. In contrast to the HS potential, the WCA potential and its derivative (the negative force) are continuous functions of r. In molecular dynamics simulations, this is beneficial since discontinuities in the potential or force can introduce instability.

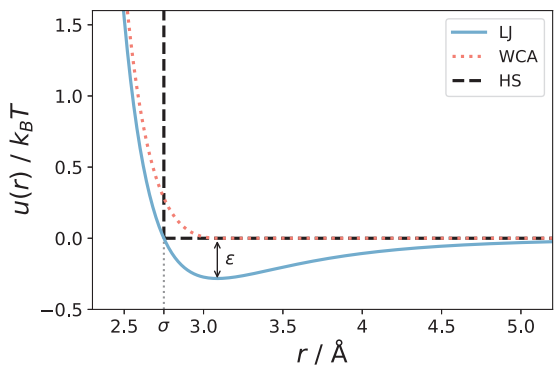

Figure 3.1: Different pair potentials plotted for the same set of parameters. The hard-sphere potential (HS) and the Weeks-Chandler-Andersen reference system pair potential (WCA), represented by the black, dashed line and the orange, dotted line, respectively, describe exchange repulsion by being everywhere repulsive. The Lennard-Jones potential (LJ) is represented by the solid, blue line and includes a repulsive ($u > 0$) and an attractive ($u < 0$) part to model both exchange repulsion and vdW interaction. The constants ϵ and σ are indicated in order to illustrate how they determine the shapes of the potentials.

Figure 3.1 depicts the HS, WCA, and LJ potential for the same set of parameters ϵ and σ.

4 Electrostatic interaction

The interaction between two point charges, q_i and q_j, is described by Coulomb's law according to

$$u^{\text{coul}}(r) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r}, \tag{3.5}$$

where q_i and q_j are the charges on two atoms i and j, ϵ_0 is the vacuum permittivity and ϵ_r is the relative permittivity of the dielectric medium, *e.g.* a solvent. In contrast to exchange repulsion and vdW interaction, the electrostatic interaction between two charges scales as r^{-1}, and due to its long-ranged nature, it imposes implications when applied in simulations, which will be discussed in the next chapter.

Chapter 4

Molecular simulation

As we saw in Chapter 2, statistical thermodynamics provides the theory for how we can retrieve information about a macroscopic sample on the microscopic scale. It would be fortunate if we could calculate the partition function, because then we could calculate any thermodynamic property. However, for all but the very simplest model systems (*e.g.* a single particle moving on a well-defined potential in one dimension), the partition function cannot be obtained exactly. To obtain even an approximate estimate of the partition function by analytical means would soon become an impossible task if we were to add more particles, dimensions, and introduce interactions to the system. However, with the use of numerical methods we can avoid a lot of simplifications and approximations and give the system a more realistic representation. Then, we can sample the partition function without knowing it beforehand, and calculate thermodynamic properties based on the sampled configurations. Molecular simulation is a numerical method where we specifically model molecular systems. When simulating large molecules, it is common to represent parts of the molecules with entities that are larger than their constituent atoms, *i.e. coarse-grained* models. However, for the studies in this thesis we have entirely employed atomistic or *all-atom* models, where each atom in the system is considered an individual entity and assigned parameters determining how it interacts with the other atoms. There are two main techniques used within molecular simulation, namely molecular dynamics (MD) simulations and Monte Carlo (MC) simulations, which will be the focus of the remainder of this chapter.

1 The simulation box

Common for MD and MC simulations is that the system is defined by a simulation box of a certain volume. The box is then filled with the molecules that one wishes to simulate, and constraints are specified to the system according to which type of ensemble that is representing the real, thermodynamic system. The problem with a box serving as our model is that the boundaries of the box will influence the system.[4] When studying bulk properties, which has been the main focus in this thesis, a simulation box with hard walls would not be suitable since the walls would have an effect on the simulated molecules, and one would thus no longer simulate bulk behavior. One could of course define an extremely large box, more mimicking a macroscopic sample, diminishing the effect from the hard walls. However, this would require an unreasonable amount of computational time and power, not feasible in present-day. The most common way to deal with the problem is to introduce periodic boundary conditions (PBC), where the simulation box is replicated in all directions to form an infinite lattice. Here, a molecule that exits the box on one side will simply enter it again on the opposite side. PBC thus have the effect of removing the hard walls of the simulation box, resulting in a more realistic model of bulk systems. However, it should be noted that PBC introduce periodicity in the replicated system, since all periodic boxes are identical. Applying PBC on very small systems thus causes errors, yet the error diminishes with the size of the simulation box and is most often smaller than the errors caused by using hard walls.

2 Pair potentials in simulations

As we stated in Chapter 3, intermolecular interactions are generally described as pair potentials. When transforming the simulation box from a system of finite size to an infinite lattice of periodic boxes by applying PBC, we would thus need to consider all pair interactions, including the molecules in the original simulation box as well as the ones in the periodic boxes. These would be infinitely many, and the approach is thus computationally infeasible. A way around this is to truncate the interactions after some distance by defining what is called a cutoff, denoted R_c. Every molecule in the simulation box is then only interacting with molecules within this (usually spherical) cutoff region. The approach will be discussed in the following, specifically considering the potentials used in this thesis, namely the LJ potential and the Coulomb potential.

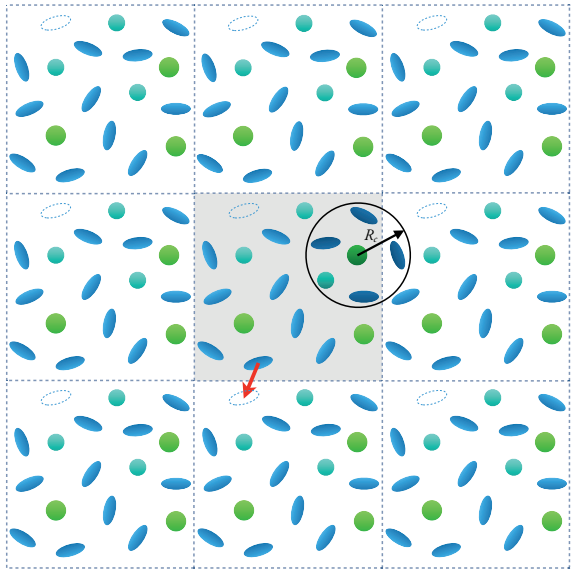

Figure 4.1: Schematic illustration of how periodic boundary conditions are applied in two dimensions. As one molecule exits the box from one side, it enters at the opposite side, as is shown by the red arrow. The black circle is the cutoff region, determined by the cutoff distance R_c (black arrow), and molecules within this region are shown in a darker color. Considering the atom in dark green, LJ interactions are only evaluated for pairs including itself and other molecules within this region.

2.1 LJ potential

For the LJ potential, truncation at a carefully chosen distance R_c is reasonable since the potential decays rapidly with r ($u(r) \propto r^{-6}$). Thus, by choosing a sufficiently large R_c, the error resulting from the truncation can be made arbitrarily small.[4] If the intermolecular potential is non-zero at the cutoff it will still result in a systematic error in the total potential. This is typically compensated for by adding a tail contribution to the long-range ($r > R_c$) part of the potential. For an arbitrary pair potential in three dimensions, the tail correction to the total potential energy U is[4]

$$U^{\text{tail}} = \frac{N\rho}{2} \int_{R_c}^{\infty} 4\pi r^2 u(r) \mathrm{d}r, \tag{4.1}$$

where N is the number of particles in the system. Here it is assumed that the number density of particles, $\rho(r)$, at any distance $r > R_c$ from any particle is equal to the average density ρ, why it can be moved out of the integral. The term $4\pi r^2 \, \mathrm{d}r$ is the volume element at distance r from any particle, in three dimensions. The tail correction converges for any potential decaying faster than r^{-3} which accordingly makes it applicable to the LJ potential.

Although the above procedure is an alternative in MC simulations, discontinuities

in the potential cause impulsive forces in MD simulations which in turn result in unstable systems. To simply truncate the potential is thus not a preferred option in MD simulations. The discontinuity at R_c is typically handled by shifting the potential so that it vanishes at the cutoff. For the LJ potential, the truncation and shift results in the following pair potential

$$u^{\mathrm{lj,tr-sh}}(r) = \begin{cases} u^{\mathrm{lj}}(r) - u^{\mathrm{lj}}(R_c), & \text{if } r < R_c \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

The shift clearly results in a potential different from the true LJ potential, with a resulting difference in the potential energy. The difference is compensated for by adding an additional term to the tail contribution in Equation 4.1.[4]

2.2 Coulomb potential

As previously discussed, the tail correction in Equation 4.1 diverges if it is applied to any pair potential decaying slower than r^{-3}; hence, it cannot be applied to the Coulomb potential. Instead, one has to resort to other methods than simply truncating the potential at some cutoff distance. Some of these methods will be described below.

Ewald summation

The Ewald method[16] makes use of PBC and treats the electrostatic potential by splitting it into a short-range part calculated in real space, and a long-range part which is Fourier-transformed and calculated in reciprocal space. In practice, to make the summation computationally feasible, both a real space and a reciprocal space cutoff need to be specified in order to neglect the smallest interactions in each space. Even with cutoffs limiting the extent of the summation, the computational cost of the method scales as $\mathcal{O}(N^2)$, where N is the number of atoms, although with optimal parameters it scales as $\mathcal{O}(N^{3/2})$. The method is exact for PBC systems, but the high computational cost compared to other methods is a drawback. Yet, there are more efficient variants of Ewald like particle mesh Ewald (PME)[17] which scales as $\mathcal{O}(N \log N)$.

Wolf formalism

Summing over all atom charges within a cutoff sphere or a sub-system of the simulation box often yields a net charge. It has been shown that a corresponding summation

of all Coulomb interactions within a cutoff sphere yields an error proportional to this net charge.[18] This was the motivation for the development of the Wolf method, in which the error was minimized by pairing every atom with an oppositely polarized image particle at the cutoff, thus effectively neutralizing the net charge of the sub-system. In systems where the net charge within a given cutoff is close to zero, the method was shown to give accurate results. The convergence of the method was later improved by adding damping parameters converting the point-charge distribution to a Gaussian charge distribution. The advantage of this method compared to Ewald-summation is that electrostatics are treated using a short-ranged potential by truncation at the cutoff, and thus the computational cost is greatly reduced, scaling with $\mathcal{O}(N)$.

3 Molecular dynamics

A molecular dynamics (MD) simulation can be seen as a computer experiment mimicking a physical experiment in the sense that the motion of molecules is evolved over time. The molecular motion is described by Newton's equations of motion, where Newton's second law of motion states that $\mathbf{F} = m\mathbf{a}$. Here $\mathbf{F}$ is the force acting on a molecule, whereas m and $\mathbf{a}$ are the the mass and acceleration of the molecule. Assuming pairwise additivity and intermolecular interactions described by pair potentials, the force acting on a molecule i at time t is

$$\mathbf{F}_i(t) = -\sum_{\text{pairs}} \nabla u[r(t)] \tag{4.3}$$

where the sum is carried out over all pairs involving molecule i, evaluating the potential for each pair resulting from the intermolecular distance r at time t. To integrate the molecular motion over time as to track the positions and velocities of the molecules, numerical integration is needed. A common numerical integrator used in MD is the *Velocity Verlet* integrator,[19] resulting from a Taylor expansion around the position of the molecule at time t, combined with Newton's second law of motion. The scheme used in simulation is as follows,

1. Calculate the position of the molecule at time $t + \delta t$:

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{\mathbf{F}(t)}{m}\delta t^2, \tag{4.4}$$

   where $\mathbf{r}$ and $\mathbf{v}$ are the position and velocity of the molecule, and δt is the timestep.

2. Calculate the force, $\mathbf{F}(t+\delta t)$, at the new position $\mathbf{r}(t+\delta t)$ using Equation 4.3.

3. Calculate the velocity of the molecule at time $t+\delta t$:

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{\mathbf{F}(t)+\mathbf{F}(t+\delta t)}{2m}\delta t \tag{4.5}$$

4. Assign $t = t+\delta t$ and start over from step 1.

The scheme illustrates how molecule positions and velocities are evolved over time. When starting a simulation, it is crucial to assign the molecules initial positions and velocities as their positions and velocities are needed in the first integration step. The Velocity Verlet algorithm natively preserves the energy of the system, given a sufficiently small timestep. Thus, when simulating other ensembles than the microcanonical, it is not enough to apply the Velocity Verlet algorithm only. For instance, simulating the canonical ensemble requires the use of a thermostat, keeping the temperature in the system constant. Simulating the isothermal-isobaric ensemble requires both a thermostat and a barostat to ensure constant temperature and pressure. There are many different thermostats [20,21,22] and barostats [23,24] available, and while these will not be discussed further in this thesis, the interested reader is referred to the references above.

4 Monte Carlo

In contrast to physical experiments and MD simulations where thermodynamic properties are obtained as averages over time, Monte Carlo (MC) simulations instead more directly rely on what was presented in Chapter 2, *i.e.* that a macroscopic sample can be represented by an ensemble, or a collection of states. As a result, associated thermodynamic properties are obtained as ensemble averages, rather than time-averages. The connection between the two methods is grounded in the *ergodic hypothesis* stating that the average value of some property, measured over a long time, is equal to the ensemble average. Since time is excluded in MC simulations, concepts such as velocity and acceleration do not exist, and only the positions of atoms, the configurations, are considered. The exclusion of time entails the disadvantage that time-dependent properties cannot be obtained like in MD simulations. On the other hand, in contrast to MD simulations where forces are evaluated, we only need to evaluate potential energies in MC simulations. This is advantageous in the sense that discontinuous pair potentials do no longer entail impulsive forces, with the result that pair potentials that would be devastating to use in an MD simulation, *e.g.* the HS pair potential, can be

used successfully in an MC simulation. Another advantage with evaluating potential energies only and iterating over states in the ensemble rather than over time is that the system is less prone to get trapped in local energy minima. In Figure 4.2, a schematic illustration of how a particle might get trapped in a local minimum of a one dimensional potential is provided. The particle is initially in the local minimum which is separated by a large energy barrier from the global minimum. When simulating the particle using MD, it has an initial velocity to the right (lowermost particle denoted MD). After two timesteps, the barrier has exerted a force strong enough to invert the direction of the particle velocity (uppermost particle denoted MD), and thus it cannot cross the barrier. When the particle is simulated using MC, a trial move (described in Subsection 4.1) is performed attempting to move the particle from the local minimum (leftmost particle denoted MC) to the global minimum (rightmost particle denoted MC). Since the potential energy in the global minimum is smaller than that in the local minimum, the particle crosses the barrier without noticing it.

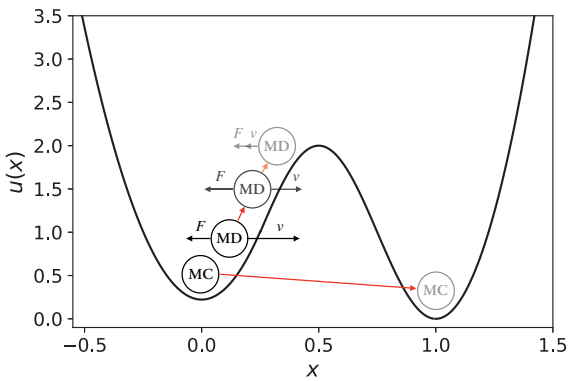

Figure 4.2: Schematic illustration of a system where a molecule moves on a 1D potential. When simulated in MD, the molecule does not possess enough velocity to cross the energy barrier and reach the global minima. When the same molecule is simulated in MC, only the difference in energy between the two positions is considered. Even though the positions are separated by a large energy barrier, the energy of the new position is smaller than that of the initial position, and the molecule crosses the barrier without any "knowledge" of it.

In the most brute-force kind of MC simulation, called random sampling, one randomly generates a set of states which are weighted by the corresponding Boltzmann factor, $e^{-\beta U}$. The weighting is functioning as to mirror the probability of the state to occur, so that high-energy states contribute less to the ensemble average than states with lower energy. Random sampling however results in a large proportion of high-energy states generated, which is inefficient since their contribution to the ensemble average is negligible. A more efficient way to sample the ensemble using MC simulation is by using the so-called *Metropolis method*,[5] which will be described next.

4.1 Metropolis method

The Metropolis method utilizes what is called *importance sampling* to preferentially sample low-energy over high-energy states. In the context of MC simulation, the sampling of different states is carried out by performing trial moves. To illustrate the method, let us consider an initial state, or configuration, m. A trial move is performed, for instance by displacing one molecule in the system by a random displacement $\delta\mathbf{r}$, resulting in a new configuration n. In equilibrium, the transition probability to go from configuration m to n, π_{mn}, must satisfy the following condition[4]

$$\mathcal{N}_m \pi_{mn} = \mathcal{N}_n \pi_{nm}, \tag{4.6}$$

where $\mathcal{N}_m$ and $\mathcal{N}_n$ are the probabilities of finding the system in configuration m and n, respectively, and π_{nm} is the transition probability of going from n to m. The condition is called *detailed balance*, and it ensures that the equilibrium distribution of the simulated system is maintained. The transition probability to go from one configuration to another can also be written in terms of the acceptance probability of the trial move, acc_{mn}

$$\pi_{mn} = \alpha_{mn} \text{acc}_{mn}. \tag{4.7}$$

Here, acc_{mn} is the transition matrix determining the probability of performing a trial move from m to n. Usually, this matrix is chosen to be symmetric, $\alpha_{mn} = \alpha_{nm}$, and Equation 4.6 can thus be rewritten as

$$\mathcal{N}_m \text{acc}_{mn} = \mathcal{N}_n \text{acc}_{nm}. \tag{4.8}$$

The equation can be rearranged to express the ratio between the acceptance probability of going from m and n and vice versa. Also, as we saw in Equation 2.13, the probability of finding the system in a certain configuration is given by the ratio between the Boltzmann factor and the configurational integral, and combining Equation 2.13 and 4.8 gives

$$\frac{\text{acc}_{mn}}{\text{acc}_{nm}} = \frac{\mathcal{N}_n}{\mathcal{N}_m} = \frac{\frac{\exp(-\beta U_n)}{Z}}{\frac{\exp(-\beta U_m)}{Z}}. \tag{4.9}$$

The configurational integral, Z, is unknown yet we know it is a constant. Therefore, the Z's in the above equation cancel and we get

$$\frac{\text{acc}_{mn}}{\text{acc}_{nm}} = \frac{\exp(-\beta U_n)}{\exp(-\beta U_m)} = \exp(-\beta[U_n - U_m]). \tag{4.10}$$

There are many choices for acc_{mn} that satisfies Equation 4.10. In the original paper,[5] the authors chose the following, known as the Metropolis criteria,

$$\text{acc}_{mn} = \begin{cases} \mathcal{N}_n/\mathcal{N}_m, & \text{if } \mathcal{N}_n < \mathcal{N}_m \\ 1, & \text{if } \mathcal{N}_n \geq \mathcal{N}_m. \end{cases} \tag{4.11}$$

Expressed in terms of the Boltzmann factors, utilizing Equation 4.9 and 4.10, Equation 4.11 can be rewritten as

$$\text{acc}_{mn} = \begin{cases} \exp(-\beta[U_n - U_m]), & \text{if } \exp(-\beta[U_n - U_m]) < 1 \\ 1, & \text{if } \exp(-\beta[U_n - U_m]) \geq 1. \end{cases} \tag{4.12}$$

According to the Metropolis criteria, a trial move is thus always accepted if it results in a decrease in the total potential energy of the system. However, if the potential energy increases, the trial move is accepted with probability $\exp(-\beta[U_n - U_m])$. The scheme used in a MC simulation is given in the following:[4]

1. Randomly choose a molecule, and calculate its energy $U(\mathbf{r}^N)$
2. Make a trial move by assigning the molecule a random displacement $\delta\mathbf{r}$, and calculate the energy of the new configuration, $U(\mathbf{r'}^N)$, where $\mathbf{r'} = \mathbf{r} + \delta\mathbf{r}$.

3a. If $\exp(-\beta[U(\mathbf{r'}^N) - U(\mathbf{r}^N)]) \geq 1$, accept the move.

3b. If $\exp(-\beta[U(\mathbf{r'}^N) - U(\mathbf{r}^N)]) < 1$, generate a uniform random number between 0 and 1. If $\exp(-\beta[U(\mathbf{r'}^N) - U(\mathbf{r}^N)])$ is greater than the random number, accept the move. Otherwise, reject the move and restore the old configuration.

4. Return to step 1.

Different kinds of trial moves can be performed, including translational (illustrated in the above scheme) and rotational moves. In addition to configurational moves it is also common to introduce unphysical moves like altering the charge of an atom or molecule (a technique used in this thesis and illustrated in Chapter 5, Subsection 2.1). The use of the Metropolis criteria for the evolution of the system makes moves like this much more straight-forward to implement in MC than in MD, which makes MC advantageous in many cases.

Chapter 5

Biased sampling

In molecular simulation, convergence of properties is often very time consuming since in principle the entire configuration space needs to be sampled. To bias the sampling in some way often leads to faster convergence, since we then can force the simulation to sample regions of interest in configuration space. For example, we can force molecules to overcome high energy barriers, sampling regions of configuration space that otherwise would be practically inaccessible. Or we can bias the simulation so that it more frequently samples configurations that contribute most to the property of interest. Through biased sampling techniques it is also possible to calculate the free energy associated with some process. In the following, some of these techniques will be presented.

1 Smarter Monte Carlo

In conventional MC, all molecules are selected for trial moves with equal probability. However, in many cases this is not the most efficient way of sampling. Instead, we might wish to select some molecules more frequently than others, or to move molecules in preferred directions. This is the idea behind a series of methods referred to as Smarter Monte Carlo,[25] of which the preferential sampling method will be described next.

1.1 Preferential sampling

Imagine that we want to study properties of a solute in a solvent, for example an ion dissolved in water. In this case, ion-water interactions and water-water interactions close to the ion are most important for determining the properties we are interested in whereas water molecules far away from the ion do not play an as important role. Thus, to bias the selection criteria so that water molecules close to the ion are selected more often for trial moves than remote ones is a more efficient way to sample than to randomly select molecules. To decide how to distinguish "close" from "remote" molecules, one usually defines a region, R_{sol}, around the solute. Molecules within this region are considered close and in the following denoted "in", whereas molecules outside are considered remote and denoted "out". In addition, a parameter p is specified, giving the probability of selecting an "out" molecule relative to an "in" one. The closer p is to zero, the more frequently "in" molecules will be selected and moved. Following the procedure by Owicki and Scheraga,[26] the sampling scheme proceeds as follows:

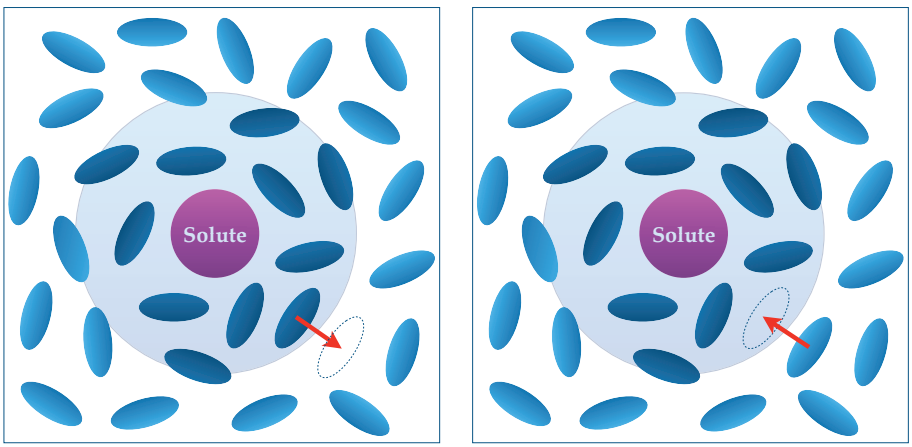

Figure 5.1: Attempts to move a molecule across the boundary of R_{sol} which is illustrated as the blue circle. The blue molecules surrounding the solute are solvent molecules, and are assigned a darker or brighter color depending on if they are "in" or "out" molecules.

1. Randomly choose a molecule.

2a. If it is "in", make a trial move.

2b. If it is "out", generate a uniform random number between 0 and 1. If p is greater than the random number then make a trial move. If not, then return to step 1.

Due to the last condition, attempts to move "out" molecules will sometimes be omitted, with the result of "in" molecules being selected and moved more frequently. Applying this scheme directly would however violate the condition for detailed balance. Imagine that we are moving an "in" molecule from a state m inside the region to a state n outside the region. Then, due to the last condition in the sampling scheme, the probability to attempt the reverse move from state n to state m would have a lower probability, given that $p < 1$. This would result in our simulation being driven out of equilibrium, since we would have a non-zero net flux of particles across the boundary of R_{sol}, from "in" to "out". To meet the requirement for detailed balance, a bias must be applied which exactly compensates for the fact that we are preferentially trying to move "in" over "out" molecules. To illustrate how this is achieved,[25] we consider the probability of accepting a trial move from state m to n, $\min(1, \mathcal{N}_n\alpha_{nm}/\mathcal{N}_m\alpha_{mn})$, where $\mathcal{N}_n$ is the probability density in state n, and α_{nm} is the probability to attempt moving a molecule from state n to m. In conventional MC, the probability of attempting to move a molecule to a specific point in the box is the same as the probability of attempting to move the same molecule from that point back to its original position, *i.e.* $\alpha_{nm} = \alpha_{mn}$, which reduces the probability of accepting a trial move to $\min(1, \mathcal{N}_n/\mathcal{N}_m)$. The problem we are left with is thus to calculate α_{nm}/α_{mn}. To do this, the authors first identified the probability of selecting an "in" molecule as

$$p_{\text{in}} = \frac{N_{\text{in}}}{N} + (1-p)\frac{N_{\text{out}}}{N}\frac{N_{\text{in}}}{N} + \left((1-p)\frac{N_{\text{out}}}{N}\right)^2 \frac{N_{\text{in}}}{N} + \ldots = \frac{N_{\text{in}}}{N'}, \qquad (5.1)$$

where N_{in} and N_{out} are the numbers of "in" and "out" molecules, respectively, and N is the total number of particles, *i.e* $N = N_{\text{in}} + N_{\text{out}}$. The first term in the equation is the probability to select an "in" molecule at first, the second term is the probability of first selecting an "out" molecule, but omitting the trial move and returning to step 1 in the sampling scheme, and then selecting an "in" molecule, and so on. The n'th term is thus the probability of selecting an "in" molecule at the n'th try to move a molecule, after trial moves on "out" molecules have been omitted $n-1$ times. The probability of selecting an "in" molecule is obtained as an infinite sum ($n \to \infty$) of these probabilities, equal to the term on the right hand side, where $N' = pN + (1-p)N_{\text{in}}$.

Now we consider a box R with N_R distinct positions (see Figure 5.2), in which we attempt to move between two different configurations m and n, with the only difference that one molecule is moved from its original position inside R_{sol} to a specific position outside R_{sol}. Then, the probability to attempt that move is

$$\alpha_{mn} = \frac{N_{\text{in}}}{N'}\frac{1}{N_{\text{in}}}\frac{1}{N_R} = \frac{1}{N'N_R}. \qquad (5.2)$$

Evaluating the expression in the middle, the first term is the probability to select an "in" molecule, the second term the probability to select a specific "in" molecule, and the third term the probability to attempt to move the molecule to one specific site in R. Considering the reverse case, where the same molecule is now an "out" molecule (among now $N_{\text{out}}+1$ "out" molecules), the probability to select an "out" molecule is

$$\begin{aligned} p_{\text{out}} =& p\frac{N_{\text{out}}+1}{N} + (1-p)\frac{N_{\text{out}}+1}{N}p\frac{N_{\text{out}}+1}{N} + \\ & \left((1-p)\frac{N_{\text{out}}+1}{N}\right)^2 p\frac{N_{\text{out}}+1}{N} + \ldots = \frac{N_{\text{out}}+1}{pN+(1-p)(N_{\text{in}}-1)}. \end{aligned} \tag{5.3}$$

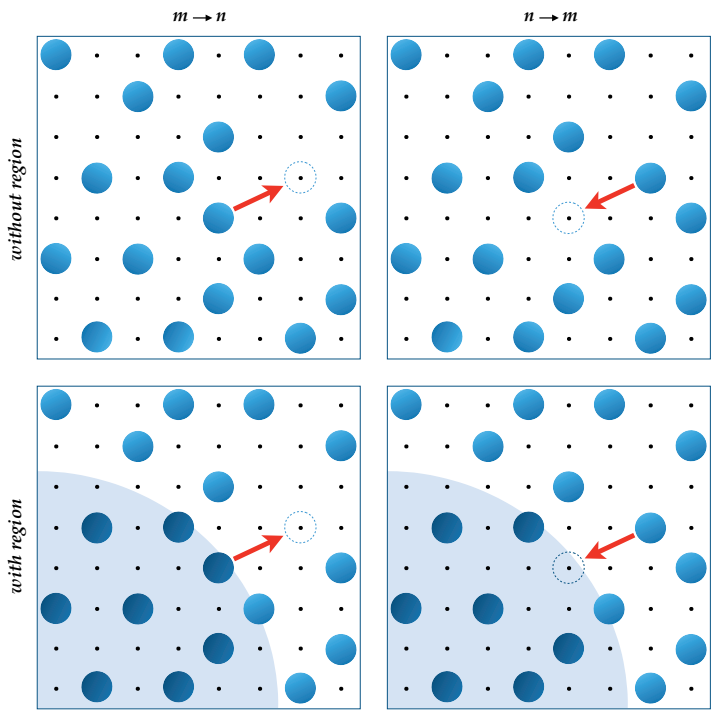

Figure 5.2: A simple system with $N = 18$ and $N_R = 64$. The left and right columns show configurations m and n, respectively. The images in the left column show a scenario when a move from state $m \to n$ is attempted, whereas the ones in the right column show the reverse scenario. The bottom and top rows show systems with and without preferential sampling. For the system without preferential sampling, the probabilities of attempting to move between m and n are equal in both directions, $\alpha_{mn} = \alpha_{nm} = \frac{1}{N}\frac{1}{N_R} = 8.7 \cdot 10^{-4}$. However, for the system with preferential sampling, α_{mn} and α_{nm} are no longer equal. For example, with a parameter $p = 0.3$, and with $N_{\text{in}} = 8$ and $N_{\text{out}} = 10$ as in this example, $\alpha_{mn} = \frac{1}{N'N_R} = 1.4 \cdot 10^{-3}$ whereas $\alpha_{nm} = \frac{p}{N'N_R}\left(1-\frac{1-p}{N'}\right)^{-1} = 4.0 \cdot 10^{-4}$

By expressing N_{out} as $N_{\text{out}} = N - N_{\text{in}}$ we get for the probability to attempt the reverse move, from $n \to m$,

$$\alpha_{nm} = \frac{N_{\text{out}} + 1}{pN + (1-p)(N_{\text{in}} - 1)} \frac{1}{N - N_{\text{in}} + 1} \frac{1}{N_R} = \frac{p}{N' N_R} \left(1 - \frac{1-p}{N'}\right)^{-1}. \tag{5.4}$$

The ratio, α_{nm}/α_{mn}, can then be determined for all four cases,

Table 5.1: The ratio, α_{nm}/α_{mn}, in four different cases.

$m \to n$	α_{nm}/α_{mn}
in $\to$ out	$p\left(1 - (1-p)/N'\right)^{-1}$
in $\to$ in	1
out $\to$ out	1
out $\to$ in	$\left[p\left(1 + (1-p)/N'\right)\right]^{-1}$

When molecules cross the boundary of the specified region, the corresponding ratio has to be added as a bias to the difference in energy between the new and old state in order to satisfy the conditions of detailed balance.

2 Wang-Landau methods

The Wang-Landau (WL) algorithm[27,28] falls under the category of flat histogram methods, and is an efficient tool for directly determining the density of states and the corresponding free energy landscape. Even though the WL algorithm is one of the most well-known flat histogram methods, the idea appears even earlier.[29,30] Originally, the WL algorithm was applied to classical spin systems, successfully determining first and second order phase transitions, but since then it has found application in more complex systems involving polymer films,[31] Lennard-Jones fluids,[32] and proteins.[33] The probability density is obtained by performing random walks in energy space. In the theoretical formulation, as different states were visited a modification factor, $f > 1$, was added to the existing density of states, $g(E)$, as $g(E) \to g(E)f$. In practice, however, it is more convenient to work with $\ln g(E)$, and update according to $\ln g(E) \to \ln g(E) + \ln f$, since numbers then are greatly reduced and can be better handled by a computer. As energy space is explored, a histogram $H(E)$ is constructed by adding integers as $H(E) \to H(E) + 1$. When the histogram is considered sufficiently flat, it is reset to zero and the modification factor f is made finer accordingly,

in the original formulation according to $f_{new} = \sqrt{f_{old}}$. Also, the density of states is modified by subtracting the smallest energy that has been accumulated in any state from all states such that $g(E)$, immediately after an update, is the difference relative to the state that has the lowest density. The simulation is then continued, using f_{new}, and when the histogram is again considered sufficiently flat, the modification factor is reduced according to the same recipe, and the simulation is continued, and so on. The idea behind reducing the modification factor is to successively reduce the error on the flatness of the histogram. When the histogram is "flat", the density of states has converged to the true value with an error proportional to $\ln f$. Thus, every time f is reduced and the histogram again becomes "flat", the estimate of the density of states has improved. The criterion to decide whether the histogram is flat can be constructed in different ways. To illustrate how the method can be used, a simple example is provided below using the method as it is currently implemented in the MC simulation software Faunus.[34]

2.1 Implementation in Faunus

For the purpose of illustrating how the method is implemented in Faunus, the Wang-Landau algorithm is applied on a very simple system consisting of two point charges, denoted i and j, in a dielectric continuum resembling water ($\epsilon_r = 78$) at 298 K. The point charges are held fixed at a separation of 7Å. They both start at $q = 0e$, where e is the elementary charge, but during the simulation the point charges are allowed to exchange charge. The system is kept electroneutral by constructing the charge exchange so that if one point charge gains $0.1e$ in charge, the other point charge simultaneously loses $0.1e$. Limits are set up so that the charge of point charge i can only visit values in the range $[-1, 0]e$, whereas the charge of point charge j is limited to explore the range $[0, +1]e$. The density of states is then projected onto the charge of point charge j, which we define as our *reaction coordinate*. Our reaction coordinate is assigned 21 bins with an increment of 0.05. Projections onto a reaction coordinate like this is often done if one wants to calculate the free energy change of a certain process, for example the free energy of binding of a ligand to a receptor, or the free energy of translating a molecule across some energy barrier, like a lipid membrane. Since we project the density of states onto the charge of point charge j and simultaneously alter the charge of both i and j, the Wang-Landau algorithm will determine the density of states, or free energy, for the process of charging point charges i and j, *i.e.* going from $q_i = q_j = 0e$ to $q_i = -1e; q_j = +1e$.

After N steps, where N is specified from input, it is checked whether all bins are sampled at least n times (n also specified from input). This works as the criterion for checking if the histogram is "flat". If all bins have been sampled at least n times, the

histogram is set to zero and the modification factor is reduced according to $f_{new} = cf_{old}$, where c is a constant between 0 and 1, given as input. The density of states, projected onto our reaction coordinate, is then updated by subtracting the energy contained in the bin that has been sampled least from all bins. Then we continue the simulation to build up a new histogram.

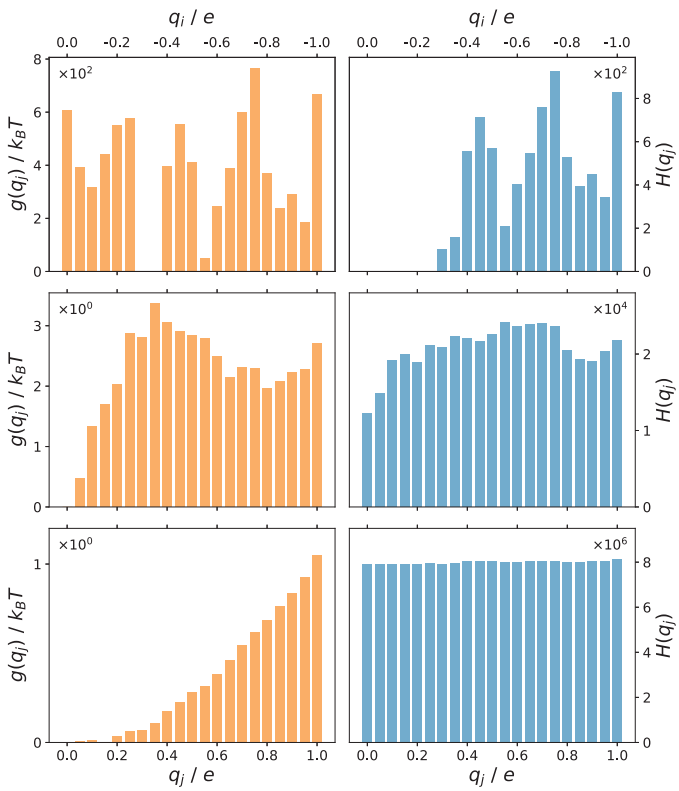

Figure 5.3: The density of states and histogram after different number of MC steps (top: 10^4 steps, middle: 10^6 steps, and bottom: $2.4 \cdot 10^9$ steps.)

In Figure 5.3, three sets of plots are depicted, where the left and right columns show the density of states and the histogram, respectively, as a function of our reaction coordinate. The top x-axis shows the charge of point charge i, to illustrate that the two point charges exchange charge. The rows of plots show results as we prolong the simulation. The top, middle, and bottom row depict the results after 10^4, 10^6, and $2.4 \cdot 10^9$ steps corresponding to 0, 78, and 128 updates, respectively. The histogram in the bottom row of plots is relatively flat with a smooth surface, and the corresponding

density of states resembles a smooth curve, monotonically increasing with q_j. At this point in the simulation, the histogram and density of states have been updated several times, and for every update, the modification factor has been made finer. Here, the modification factor started at $1k_BT$, and for every update, is was scaled with a factor 0.9. Hence, in the bottom row, it has reached the value $f_{128} = 1k_BT \cdot 0.9^{128} \approx 1.4 \cdot 10^{-6}k_BT$.

To obtain the free energy of charging point charge i and j, the density of states is inverted. This is done to produce the plot in Figure 5.4, where the free energy from the simulation is shown as orange circles. The circles show the free energy difference for each bin with respect to the bin that has been sampled the least number of times, since this bin is subtracted from all other bins when the density of states is updated. The free energy of charging is obtained by taking the difference between the value at $q_i = -1e; q_j = +1e$ and at $q_i = q_j = 0e$ according to: $\Delta G_{charging} = \Delta G(q_{i/j} = \pm 1e) - \Delta G(q_{i/j} = 0e)$ which gives $-1.05k_BT$. This is close to the analytic solution obtained with Coulomb's law ($-1.02k_BT$).

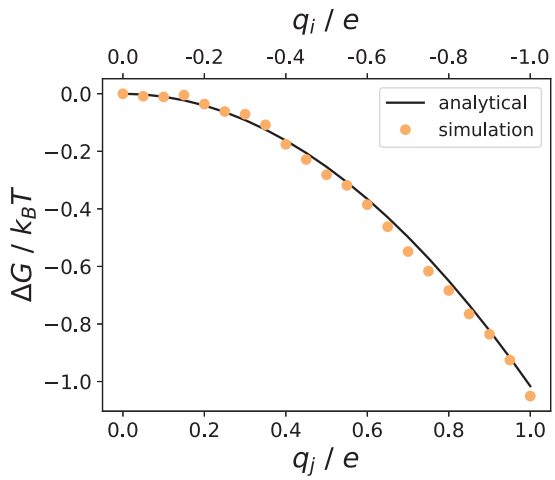

Figure 5.4: The free energy profile when charging point charge i and j. The orange circles are simulated data, whereas the black curve is obtained analytically through Coulomb's law.

Chapter 6

Bennett acceptance ratio

In addition to the WL algoritgm, other free energy calculation methods have been derived. One of them is the Bennett acceptance ratio method (BAR),[35] being one of the earliest methods for calculating free energy differences. To calculate the free energy difference between two states, *i* and *j*, the method requires data from configurations at each state. The BAR method relies on the idea that for the same configuration, *e.g.* a configuration sampled from state *i*, $\mathbf{r}_i^N$, there is a pathway connecting two states with potentials U_i and U_j. As the states are drawn from the same configuration, it is possible to calculate the potential difference between the states, $\Delta U_{ij} = U_j - U_i$, exactly. Likewise, one can calculate the potential difference $\Delta U_{ji} = U_i - U_j$ for a configuration sampled from state *j*, $\mathbf{r}_j^N$. Linking the distributions of potential energy differences, ΔU_{ij} and ΔU_{ji}, calculated from configurations belonging to each state, and finding the optimal way to use the data sampled from the two states was the starting point for the derivation of the method. The aim of this chapter is to present a condensed version of the original derivation of the method, ending with an illustrative section to further explain the concept.

1 Derivation

The Metropolis acceptance probability of going from state *i* to *j* with potentials U_i and U_j is,

$$\mathrm{acc}(U_j - U_i) = \min(1, \exp\{-\beta[U_j - U_i]\}). \tag{6.1}$$

Taking the ratio between acceptance probabilities when going from state i to j and vice versa gives,

$$\frac{\mathrm{acc}(U_j - U_i)}{\mathrm{acc}(U_i - U_j)} = \frac{\min(1, \exp\{-\beta[U_j - U_i]\})}{\min(1, \exp\{-\beta[U_i - U_j]\})} = \exp(-\beta[U_j - U_i]). \tag{6.2}$$

Rearranging this equation gives,

$$\mathrm{acc}(U_j - U_i)\exp(-\beta U_i) = \mathrm{acc}(U_i - U_j)\exp(-\beta U_j). \tag{6.3}$$

To derive the BAR method, the author considered the special kind of trial move referred to in the start of this section - one that goes from one state to another, while keeping the configuration belonging to the first state. Integrating over configuration space, and multiplying with the (trivial) ratios of configuration integrals Z_i/Z_i and Z_j/Z_j, he expressed Equation 6.3 as,

$$Z_i \frac{\int \mathrm{acc}(U_j - U_i)\exp(-\beta U_i)\mathrm{d}\mathbf{r}^N}{Z_i} = Z_j \frac{\int \mathrm{acc}(U_i - U_j)\exp(-\beta U_j)\mathrm{d}\mathbf{r}^N}{Z_j}. \tag{6.4}$$

The fractions are ensemble averages of the acceptance probabilities $\mathrm{acc}(U_j - U_i)$ and $\mathrm{acc}(U_i - U_j)$ sampled in state i and j, respectively, and the equality can be rearranged to,

$$\frac{Z_i}{Z_j} = \frac{\langle \mathrm{acc}(U_i - U_j)\rangle_j}{\langle \mathrm{acc}(U_j - U_i)\rangle_i}. \tag{6.5}$$

The physical interpretation of Equation 6.5 is that the ratio between the configuration integrals for ensemble i and j equals the ratio between the ensemble-averaged acceptance probabilities when switching the potential from U_i to U_j and vice versa, sampling in ensemble j and i, respectively. Since only one ensemble is sampled to get one of the ensemble-averaged acceptance probabilities, *e.g.* only ensemble j is sampled to obtain $\langle \mathrm{acc}(U_i - U_j)\rangle_j$, the potential-switching trial moves do not need to be carried out. Instead, $\langle \mathrm{acc}(U_i - U_j)\rangle_j$ and $\langle \mathrm{acc}(U_j - U_i)\rangle_i$ can be obtained by calculating the acceptance probabilities on the fly during sampling of ensemble j and i, respectively.

To obtain a reliable estimate of the ratio in Equation 6.5, and thus the free energy difference between state i and j, the two ensembles i and j must possess sufficient overlap in their configuration space. With negligible overlap, the main argument that one can

calculate the free energy difference between two states i to j by only sampling in ensemble i falls apart since ensemble j is then *not similar enough* to ensemble i. Another important requirement, of practical relevance, is that the ensemble-averaged acceptance probabilities of switching the potential from U_i to U_j and vice versa must both be large enough to allow them to be determined with reasonable statistical accuracy during the course of any simulation of reasonable length. If both ensemble-averaged acceptance probabilities are too small to be sampled properly, it indicates that there is an insufficient overlap between the configuration space of ensemble i and j. One solution to this is to add intermediate states between i and j, acting as a chain connecting the two end-states with sufficient overlap between each neighboring state, increasing at least one of the ensemble-averaged acceptance probabilities. If only one of the ensemble-average acceptance probabilities is too small, it is possible to increase it, at the cost of the other, by adding a constant to the corresponding potential function in order to shift its origin, bringing it closer to the potential function of the other state. Intuitively, there is an optimal choice of this constant which acts to equalize the two ensemble-averaged acceptance probabilities. In his original paper, the author showed that this optimal choice also improves the estimate of the ratio in Equation 6.5, and the theory behind the idea is presented in the following.

To derive the ratio of configuration integrals, the Metropolis acceptance probability was used to arrive at Equation 6.5. However, a more general formula expressing this ratio was derived by including an arbitrary weighting function, $w(\mathbf{r}^N)$, which is a function of the coordinates and finite everywhere,

$$\frac{Z_i}{Z_j} = \frac{Z_i}{Z_j}\frac{\int w(\mathbf{r}^N)\exp(-U_i - U_j)\mathrm{d}\mathbf{r}^N}{\int w(\mathbf{r}^N)\exp(-U_j - U_i)\mathrm{d}\mathbf{r}^N} = \frac{\langle w\exp(-\beta U_i)\rangle_j}{\langle w\exp(-\beta U_j)\rangle_i}. \tag{6.6}$$

Equation 6.6 is valid for an arbitrary w, and Equation 6.5 can be considered a special case of Equation 6.6 with $w = \exp(\min(U_i, U_j))$. The challenge then lies in finding the weighting function which minimizes the statistical error in $\beta\Delta F = \ln(Z_i/Z_j)$. In the original paper, this optimal weighting function was found to be,

$$w = \frac{\text{constant}}{(Z_i/n_i)\exp(-U_j) + (Z_j/n_j)\exp(-U_i)}, \tag{6.7}$$

where n_i and n_j are the numbers of statistically independent samples in ensemble i and j, respectively. Substituting w with this expression in Equation 6.6 yields,

$$\frac{Z_i}{Z_j} = \frac{\langle(1 + \exp\{U_i - U_j + C\})^{-1}\rangle_j}{\langle(1 + \exp\{U_j - U_i - C\})^{-1}\rangle_i}\exp(C), \tag{6.8}$$

where the constant C acts to shift the potentials. The expressions inside the ensemble averages can be recognized as Fermi-Dirac functions, $f(x) = 1/(1 + \exp(x))$, and Equation 6.8 can thus be rewritten as,

$$\frac{Z_i}{Z_j} = \frac{\langle f(U_i - U_j + C)\rangle_j}{\langle f(U_j - U_i - C)\rangle_i} \exp(C). \tag{6.9}$$

Equation 6.9 is true for any value of C. However, in the original paper it was shown that there is an optimal choice, $C = \ln\left((Z_i n_j)/(Z_j n_i)\right)$, which minimizes the statistical error in $\beta\Delta F$. This expression poses an obstacle since it requires the ratio Z_i/Z_j, which is exactly what we want to compute. Fortunately, it is possible to calculate the optimal value of C in a self-consistent manner, which is described in the following.

Expressing the free energy using Equation 6.9, where the ensemble averages are replaced by sums over finite sample series, which is what we deal with when using simulations, yields,

$$\begin{aligned} \beta\Delta F = \ln\frac{Z_i}{Z_j} &= \ln\left(\frac{\sum_j f(U_i - U_j + C)}{\sum_i f(U_j - U_i - C)}\frac{n_i}{n_j}\exp(C)\right) \\ &= \ln\frac{\sum_j f(U_i - U_j + C)}{\sum_i f(U_j - U_i - C)} - \ln\frac{n_j}{n_i} + C. \end{aligned} \tag{6.10}$$

But, as was shown in the original paper, for the optimal value of C the error in the free energy difference is minimized by equalizing the sums of Fermi-Dirac functions in each ensemble, making the first term in the right hand side of Equation 6.10 equal to zero, and we have,

$$\beta\Delta F = -\ln\frac{n_j}{n_i} + C. \tag{6.11}$$

Thus, if we know the potentials U_i and U_j for a sufficient number of configurations in both ensemble i and j, we can compute $\sum_j f(U_i - U_j + C)$ and $\sum_i f(U_j - U_i - C)$ for a range of values of C. When the sums are equal, we know that we have found the optimal value of C, and the free energy difference can be calculated according to Equation 6.11.

2 Illustration

To calculate the solvation free energy of potassium chloride (KCl), a series of simulations is performed. In each simulation, the ions are assigned a coupling parameter λ. The parameter functions as to couple the interactions, in this case LJ and Coulomb interactions, between the ions and the rest of the system. This is achieved by altering the interaction parameters on the ions so that for LJ interactions, we have $\lambda\sigma$ and $\lambda\epsilon$, whereas for Coulomb interactions, we have λq. As stated in the previous section, we are only interested in the end-states. In principle, only two simulations, one with $\lambda = 0$ and one with $\lambda = 1$, would thus be needed in order to calculate the free energy of solvation. The former case corresponds to no interactions at all between the ions and the rest of the system, mimicking the scenario where K^+ and Cl^- are in vacuum. The latter case corresponds to full interaction between the ions and the rest of the system, which is the case when the ions are solvated in the aqueous phase. In practice though, a set of only two simulations generally does not suffice to obtain a reliable estimate of the free energy associated with some process since the overlap in configuration space between the two states, *i.e.* $\lambda = 0$ and $\lambda = 1$, is poor. Hence, intermediate states corresponding to intermediate λ's are needed in order to connect the end-states, enabling a greater overlap in configuration space between each neighboring state (λ).

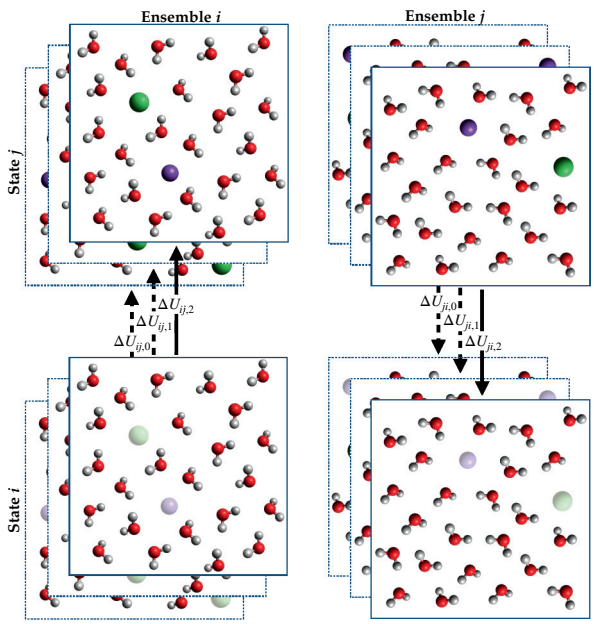

Figure 6.1: Schematic illustration of how values of $\Delta U_{ij,n}$ and $\Delta U_{ji,n}$ are calculated during the course of two separate simulations, where one simulation is carried out in state *i* (left column, denoted Ensemble *i*) and the other in state *j* (right column, denoted Ensemble *j*).

Figure 6.1 illustrates the first step towards calculating the free energy of solvation of KCl, where the purple and green spheres represent K^+ and Cl^-, respectively. The left and right columns in the figure represent simulations in two neighboring states, denoted "Ensemble *i*" and "Ensemble *j*". The transparent colors of the spheres in the bottom row of images ("State *i*") indicate that we are calculating the potential energy for state *i*, with λ_i corresponding to a state where the interactions between the ions and the rest of the system are slightly decoupled, *i.e.* $\lambda_i < 1$. The full colors in the top row of images ("State *j*") correspond to the calculation of the potential energy in state *j*, where the ions are fully interacting, *i.e.* $\lambda_j = 1$. The stacking of 2D simulation boxes is meant to illustrate that $\Delta U_{ij,n}$ and $\Delta U_{ji,n}$, being the potential energy difference between state *j* and *i* and vice versa, are calculated for different configurations during a simulation run, where *n* is the *n*'th configuration of the simulation trajectory. To obtain the free energy of solvation of KCl, additional simulations are performed, connecting the end-state $\lambda = 1$ to the lower end-state $\lambda = 0$, and the free energy differences between all neighboring states are summed together.

The code snipped below is a minimal example of how the BAR method might be implemented in a computer code. Here, the number of independent configurations in 'Ensemble *i*" and "Ensemble *j*" are equal, meaning that the logarithmic term in 6.11 dissappears, and we obtain the free energy difference simply as

$$\beta \Delta F = C. \tag{6.12}$$

The code illustrates how the optimal value of C is found by equalizing the difference between the sums of Fermi-Dirac functions (see Equation 6.10) for two neighboring states.

```
import numpy as np
from scipy.optimize import minimize

RT = 0.008314*300 # convert energies from kJ/mol to unitless

def fermiDiracFunctions(C):
    # Reading files with histograms of independent samples of:
    # Potential energy diff. btw state i and j simulated in state j (dU_ji)
    # Potential energy diff. btw state j and i simulated in state i (dU_ij)
    dU_ji = np.loadtxt("dU_ji.dat", usecols=[0,1])
    dU_ij = np.loadtxt("dU_ij.dat", usecols=[0,1])

    # Initializing sums of Fermi-Dirac functions
    sum_f_ji = 0; sum_f_ij = 0

    # Adding Fermi-Dirac function of each potential energy diff. encountered (dU[0]), weighted by the number of occurrences (dU[1]) of that diff.
    for dU in dU_ji:
        sum_f_ji += dU[1]*(1/(1+np.exp(dU[0]/RT+C)))
    for dU in dU_ij:
        sum_f_ij += dU[1]*(1/(1+np.exp(dU[0]/RT-C)))

    # Return squared diff. btw sums of Fermi-Dirac functions
    return( (sum_f_ji - sum_f_ij) ** 2 )

# Finding the C which equalizes the sums of Fermi-Dirac functions
result = minimize(fermiDiracFunctions, 10)

# Printing the found value of C
print("C =", np.round(result.x[0], decimals=2), "kT")

# Printing diff. btw sums of Fermi-Dirac functions
print("Difference, sums of Fermi Dirac functions\t = ", np.round(np.sqrt(result.fun), decimals = 6))

# Printing free energy diff. in kJ/mol, equal to C*RT
print("Free energy difference =", result.x[0]*RT, "kJ/mol")
```

Chapter 7

Kirkwood-Buff solution theory

The Kirkwood-Buff (KB) theory of solutions[36] is a statistical mechanical theory that provides a pathway to obtain several observable, macroscopic properties of a system from its microscopic properties. Examples of such macroscopic properties are derivatives of chemical potentials with respect to concentrations, partial molar volumes, and compressibility. These are obtained by integrating the radial distribution functions (RDFs) between components in the grand canonical, μVT, ensemble. The theory is exact and valid for any shape of particles, not necessarily spherical ones. The theory does not assume pairwise additivity of particle interactions, which brings an advantage over less general theories that rely on just this.[37] The central property in KB theory is the Kirkwood-Buff integral (KBI), defined as

$$G_{\alpha\beta} = \frac{1}{V}\int_V\int_V (g_{\alpha\beta}^{\mu VT}(r_{12}) - 1)\mathbf{dr}_1\mathbf{dr}_2, \tag{7.1}$$

where subscripts α and β denotes two components and $g_{\alpha\beta}^{\mu VT}(r_{12})$ is the RDF between those components in the grand canonical ensemble, $\mathbf{r}_1$ and $\mathbf{r}_2$ are the coordinates of all pairs of particles in the box, with interparticle distance $r_{12} = \mid \mathbf{r}_1 - \mathbf{r}_2 \mid$. The double integral thus runs over all particle pairs in the volume V, but only pairs including both component α and β contribute to the integral. In the thermodynamic limit, *i.e.* when $V \to \infty$, the double integral in Equation 7.1 can be reduced to the following single integral using the coordinate transformation $\mathbf{r}_2 \to \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$,[38]

$$G_{\alpha\beta} = \int_0^\infty (g_{\alpha\beta}^{\mu VT}(r) - 1)4\pi r^2\mathbf{dr}. \tag{7.2}$$

The integral runs over an infinite volume, meaning that a system of infinite size is required to obtain the exact KBI. The other requirement is that the RDF is sampled in the μVT ensemble, which is an open system allowing for particle fluctuations across the system boundaries while keeping the chemical potential constant.

1 Application in simulations

In simulations, the size of the system of interest cannot be infinite, due to the limits of computational power. Also, KB solution theory requires an open system, and we cannot apply it directly on closed systems like the canonical, *NVT*, or the isothermal-isobaric, *NPT*, ensembles.

1.1 Approximations

To circumvent these issues, the usual procedure is to truncate the KBI obtained from the simulated, closed system at some truncation distance R,

$$G_{\alpha\beta} \approx G^R_{\alpha\beta}(R) = \int_0^R (g_{\alpha\beta}(r) - 1)4\pi r^2 \mathrm{d}r, \tag{7.3}$$

where $G^R_{\alpha\beta}(R)$ and $g_{\alpha\beta}(r)$ denote the KBI and the RDF in the closed system, respectively. The integral is then evaluated at this truncation distance, where it is assumed that bulk properties are reached. In other words, when going further out from the central component from which the RDF is calculated we will not see a change in the composition of the solution. The sub-region of the now considered system (from here on denoted the KB region), with a radius equal to the distance from the center component to R, now effectively becomes an open system since particles can move freely across its boundary. In this way, we can apply the KB solution theory to closed systems like the canonical, *NVT*, and the isothermal-isobaric, *NPT*, ensemble.

Applying the KB solution theory to a finite system, truncating the integral like above is an approximation, and with it errors are introduced. To analyze these errors another property becomes useful, namely the excess coordination number, defined as,

$$\Delta N_{\alpha\beta} = \rho_\beta G_{\alpha\beta}. \tag{7.4}$$

The subscripts, α followed by β, means that $\Delta N_{\alpha\beta}$ is the excess coordination number of component β around component α, *i.e.* the excess number of β particles in the

KB region, compared to the number of β particles in a uniformly distributed system. Hence, in any system that deviates from an ideal gas, $\Delta N_{\alpha\beta}$ will be non-zero; positive or negative depending on whether there is an excess or deficit of component β around component α. In a finite, closed system, an excess of particles β in the KB region must be compensated for by a deficit in the remainder of the system. Since RDFs are computed as $g_{\alpha\beta}(r) = \rho_{\alpha\beta}(r)/\rho_\beta$ where $\rho_{\alpha\beta}(r)$ is the number density of component β at a distance r from the center component α, and ρ_β is the average number density of component β in the whole simulation box, the deficit of component β in the remainder of the box will make the RDF converge to a value smaller than one. Likewise, a deficit of particles β in the KB-region would lead to an excess in the remainder of the box, overestimating the tail of the RDF.

1.2 Correction factors

To correct for this finite size error, the following correction for RDFs has been proposed,[39]

$$g^*_{\alpha\beta}(r) = g_{\alpha\beta}(r)\frac{N_\beta\left(1-\frac{V(r)}{V_{box}}\right)}{N_\beta\left(1-\frac{V(r)}{V_{box}}\right)-\Delta N_{\alpha\beta}(r)-\delta_{\alpha\beta}}, \tag{7.5}$$

where $g_{\alpha\beta}(r)$ is the RDF obtained from simulation and $g^*_{\alpha\beta}(r)$ is the finite size-corrected RDF. In the correction factor, N_β is the total number of particles of type β, $V(r)$ is the volume of a sphere of radius r, V_{box} is the volume of the simulation box, $\Delta N_{\alpha\beta}$ is the excess coordination number of component β around component α, and $\delta_{\alpha\beta}$ is the Kronecker delta, equal to one if $\alpha = \beta$ and zero otherwise. The numerator of the correction factor is the number of particles of type β that would be present outside the sphere of radius r if they were uniformly distributed throughout the box. The denominator is the actual number of particles of type β present outside this sphere, based on $\Delta N_{\alpha\beta}$ obtained from simulation. The Kronecker delta is included to account for the the central particle if $\alpha = \beta$, since it is not included in $\Delta N_{\alpha\beta}$. The correction thus have the effect of accounting for the true bulk density rather than the average density which is used when RDFs are calculated from simulation.

In another study,[38] the authors proposed a correction factor to account for errors introduced when reducing the double integral in Equation 7.1 to the single, finite integral in Equation 7.3. For the coordinate transformation $\mathbf{r}_2 \rightarrow \mathbf{r} = \mathbf{r}_1 - \mathbf{r}_2$ to be valid, $\mathbf{r}$ must be independent of $\mathbf{r}_1$, which is only true for an infinite volume. However, in a simulated system the volume is always finite, which makes $\mathbf{r}$ dependent on $\mathbf{r}_1$. To account for this, the authors suggested the replacement of $4\pi r^2$ in Equation

7.3 with a geometrical function $w(x)$,

$$G^{w}_{\alpha\beta}(R) = \int_0^R (\, g_{\alpha\beta}(r) - 1) w(x) \mathrm{d}r, \tag{7.6}$$

where the superscript w indicates that it is the geometry-corrected KBI. The geometrical function depends on $x = r/R$, and the resulting expression depends on the dimensionality of the system. In three dimensions, it is

$$w(x) = 4\pi r^2 \left(1 - \frac{3x}{2} + \frac{x^3}{2}\right). \tag{7.7}$$

The authors showed that, for sufficiently large R, $G^{w}_{\alpha\beta}(R)$ is proportional to $1/R$. Thus, by plotting $G^{w}_{\alpha\beta}(R)$ for different truncation distances R, and then linearly extrapolating $G^{w}_{\alpha\beta}(R)$ to $1/R \to 0$, a precise estimate of the exact KBI, $G_{\alpha\beta}$, is obtained. As an alternative, to avoid the extrapolation step, a corresponding extrapolated expression was presented. The expression was obtained by performing a first-order Taylor expansion of $G^{w}_{\alpha\beta}(R)$ in $1/R$ around zero: $G_{\alpha\beta} \approx G^{w}_{\alpha\beta}(R) - (1/R)(\mathrm{d}G^{w}_{\alpha\beta}(R)/\mathrm{d}(1/R)) \equiv \hat{G}^{w}_{\alpha\beta}(R)$, resulting in,

$$\hat{G}^{w}_{\alpha\beta}(R) = \int_0^R (\, g_{\alpha\beta}(r) - 1) \left(w - x\frac{\partial w}{\partial x}\right) \mathrm{d}r, \tag{7.8}$$

referred to as the *extrapolated* expression. Evaluating the last term in the integral gives,

$$w - x\frac{\partial w}{\partial x} = 4\pi r^2(1 - x^3) = 4\pi r^2 \left(1 - \left(\frac{r}{R}\right)^3\right). \tag{7.9}$$

The authors pointed out an important difference between the directly truncated KBI, $G^{R}_{\alpha\beta}(R)$, and the geometry-corrected KBIs, $G^{w}_{\alpha\beta}(R)$ and $\hat{G}^{w}_{\alpha\beta}(R)$. In the expression for $G^{R}_{\alpha\beta}(R)$, a central particle is considered, meaning that the symmetry of the solution is only respected in the limit where $R \to \infty$. This means that for any finite volume $4\pi R^3/3$ considered, the average density inside this volume is not equal to the macroscopic value. However, in the expressions for $G^{w}_{\alpha\beta}(R)$ and $\hat{G}^{w}_{\alpha\beta}(R)$, the geometrical function $w(r, R)$ corrects for this so that the average density inside the finite volume considered equals the macroscopic value.

In a third study,[40] the authors showed that applying both correction factors presented above performs better than applying either of them solely. Thus, we follow their approach in Paper I and II and simultaneously apply the finite size-corrected RDFs[39] with the extrapolated geometrical function[38] according to,

$$G^*_{\alpha\beta}(R) = \int_0^R (g^*_{\alpha\beta}(r) - 1) 4\pi r^2 \left(1 - \left(\frac{r}{R}\right)^3\right) \mathrm{d}r. \tag{7.10}$$

2 Thermodynamic properties

Although the original derivation of KB solution theory started from the grand canonical ensemble, it is possible to recast the resulting expressions for average thermodynamic properties into constant pressure form.[36,37,41] In a two-component solution at constant pressure and temperature, the derivative of the chemical potential of species α with respect to its average number density is given by[37]

$$\frac{\rho_\alpha}{k_B T}\left(\frac{\partial \mu_\alpha}{\partial \rho_\alpha}\right)_{p,T} = \frac{1}{1 + \rho_\alpha(G_{\alpha\alpha} - G_{\alpha\beta})}, \tag{7.11}$$

where μ_α is the chemical potential of component α and ρ_α is the average number density of component α. The derivative can also be expressed in terms of the activity of component α, and is equivalent to,

$$\left(\frac{\partial \ln a_\alpha}{\partial \ln \rho_\alpha}\right)_{p,T} = \frac{1}{1 + \rho_\alpha(G_{\alpha\alpha} - G_{\alpha\beta})}. \tag{7.12}$$

In Paper I and II, we used Equation 7.12 to calculate activity derivatives for different salts and concentrations, but replaced the KBIs with the corresponding corrected ones, as given in Equation 7.10.

Chapter 8

On-going research

Currently, we are investigating binding of potassium ions to a compound called the bis(18-crown-6) analogue of Tröger's base (BCETB). Crownethers have since their discovery[42] been known to form stable complexes with several different metal cations, where the potassium ion is one of them. Whereas numerous studies have been conducted on single crownethers such as 18-crown-6,[43,44,45,46] the BCETB has not been studied to the same extent. The BCETB consists of two 18-crown-6 ethers connected by the Tröger's base, and hence it has the ability to bind two cations, one in each crownether. The compound is of interest since calorimetry measurements have shown that binding of a second potassium is enthalpically stabilized by the binding of the first. This is evident from the experimental results showing that the enthalpy of binding of the second potassium ion is more negative than that of the first.[47] The first binding event is considered as the binding of a potassium ion to one of the two crownethers when nothing is bound, whereas the second binding event is the binding of a second potassium ion when the first is already bound.

To describe the binding events in simulation we have to somehow transfer the potassium ions from the solution bulk to the binding sites in the BCETB. The usual way to treat the problem is to divide it up into two separate processes. In the first process, the potassium ion is transferred from vacuum to the solution bulk giving the free energy of solvation, ΔG_{solv}. In the second process, the potassium ion is transferred from vacuum to the binding site, yielding the free energy of complexation, $\Delta G_{\mathrm{complex}}$. The free energy of binding is then obtained as the difference between the complexation free energy and the solvation free energy, $\Delta G_{\mathrm{bind}} = \Delta G_{\mathrm{complex}} - \Delta G_{\mathrm{solv}}$. This is illustrated for the first binding event in Figure 8.1, where a Cl^- ion has been added to keep the system electroneutral. In the first process, we want the potassium and the chloride ion to only sample the solution bulk. This poses an obstacle, since we need

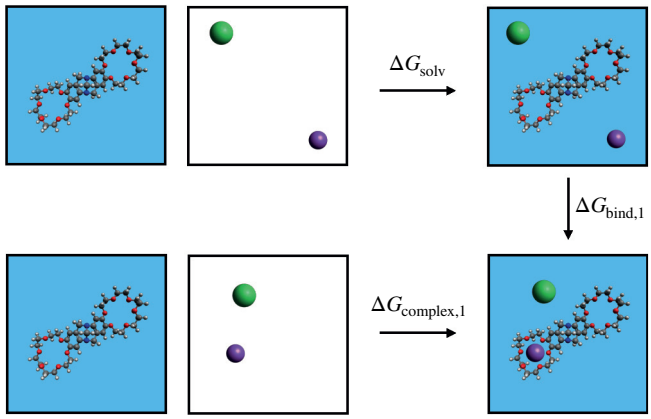

Figure 8.1: Schematic illustration of how the first binding event is divided up into two separate processes. The purple and green spheres represent the potassium and chloride ions, respectively.

to prevent the ions from moving into the binding sites. A common way to circumvent this problem is to remove the receptor, in this case the BCETB. Thus we end up with the scheme illustrated in Figure 8.2 for the first binding event. To remove the BCETB is an approximation, but since the ions are supposed to be transferred to the bulk solution, the presence of the BCETB is not expected to influence the free energy of solvation of the ions. In this way, we ensure that the ions are always in the bulk solution.

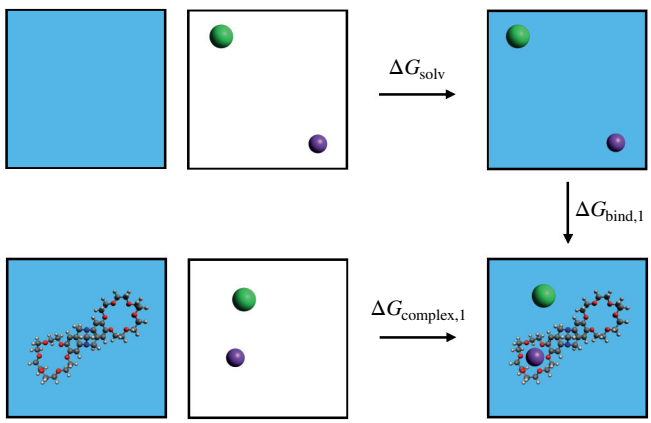

Figure 8.2: Schematic illustration of the first binding event, with the approximation that the BCETB is removed when the free energy of solvation is calculated (top row).

For the second binding event, the scheme is similar, with the difference that when the second potassium ion binds, the first one is already bound. This scenario is illustrated in Figure 8.3.

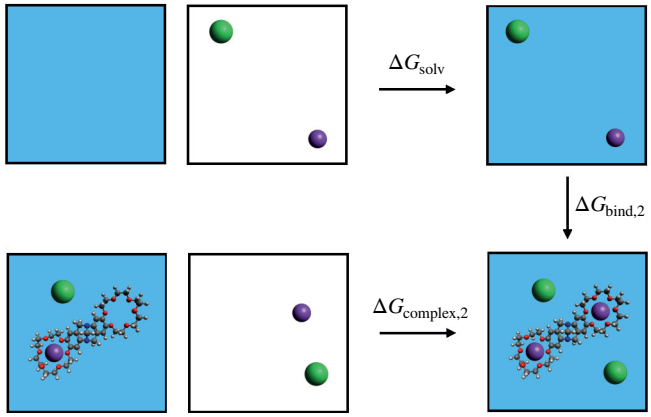

Figure 8.3: Schematic illustration of second binding event, with the approximation that the BCETB is removed when the free energy of solvation is calculated (top row).

In the current study, we are investigating the free energies and enthalpies associated with the two binding events using two different approaches. Of special interest is the difference between the free energies and enthalpies of binding of the first and second potassium ion, *i.e.* $\Delta\Delta G = \Delta G_{\text{bind,2}} - \Delta G_{\text{bind,1}}$ and $\Delta\Delta H = \Delta H_{\text{bind,2}} - \Delta H_{\text{bind,1}}$, respectively. In the first approach, MD simulations are performed and the BAR method is applied in order to calculate free energies and enthalpies. In the second approach, we use MC simulations together with the WL algorithm.

1 MD simulations with the BAR method

In the first approach, the MD simulation software GROMACS v. 2019 [48,49,50] is used, applying the BAR method to calculate free energies. Both the LJ parameters and the charges on the potassium and chloride ions are altered by a parameter λ that takes values in the range $[0, 1]$ according to $\lambda\epsilon$, $\lambda\sigma$, and λq. When $\lambda = 0$, the potassium and chloride ions are hence non-interacting, whereas when $\lambda = 1$, the ions are fully interacting. Three sets of simulations are performed. The first set of simulations correspond to the transfer of a potassium and a chloride ion from vacuum to the bulk solution (top row of Figure 8.2 and Figure 8.3). The second set corresponds to the transfer of a potassium ion from vacuum to one of the binding sites and the chloride ion from vacuum to the bulk solution (bottom row of Figure 8.2). The third set represent the process of transferring a second potassium ion from vacuum to the second binding site, and a second chloride ion from vacuum to the bulk solution (bottom row of Figure 8.3). Each set of simulations is consisting of n separate simulations, where each simulation is assigned a unique value of λ in the range $[\lambda_0,$

λ_1, λ_2,...,λ_n]. Here, $\lambda_0 = 0$ and hence corresponds to the simulation where the ions are not interacting at all, whereas $\lambda_n = 1$ and corresponds to the simulation where the ions are fully interacting. The λ's in between take intermediate values and correspond to intermediate states. We are in principle only interested in the end-states, *i.e.* λ_0 and λ_n, since we want to calculate the free energy difference between the states when the ions are not present at all and when they are fully interacting. However, the intermediate λ's are important to bridge the end-states so that all neighboring λ-states have sufficient configurational overlap. The free energy difference between each neighboring state is obtained by finding the constant C (see Chapter 6, Section 1), and the free energy between the end-states is then obtained as the sum over all free energies between neighboring states. To keep the potassium ions in their binding sites, a harmonic potential is used for each binding site. The equilibrium position for each harmonic potential is defined as the center of mass of the oxygen atoms in each respective crownether, keeping the potassium ions in the centers of the rings.

2 MC simulations with the WL algorithm

In the second approach, we construct a somewhat more simple model to see if it is good enough to reproduce the experimental values of $\Delta\Delta G$ and $\Delta\Delta H$. When calculating $\Delta\Delta G$:

$$\Delta\Delta G = \Delta G_{\text{bind},2} - \Delta G_{\text{bind},1} = (\Delta G_{\text{complex},2} - \Delta G_{\text{solv}}) - (\Delta G_{\text{complex},1} - \Delta G_{\text{solv}}), \tag{8.1}$$

we see that the free energy of solvation terms cancel, and hence the simulation corresponding to that part of the problem can be omitted. Thus, we only have to calculate $\Delta G_{\text{complex},2}$ and $\Delta G_{\text{complex},1}$ in order to get $\Delta\Delta G$. Furthermore, we choose to not include the chloride counterion since the contribution from that is expected to cancel when taking the difference $\Delta\Delta G = \Delta G_{\text{complex},2} - \Delta G_{\text{complex},1}$. This is an approximation, but it is advantageous in the sense that by including couterions, we need to allow them to explore the whole simulation box in order to get converged results for the binding free energies. Hence, by omitting the counterions, we reduce the problem, hopefully obtaining converged results quicker. Omitting the chloride counterion results in a system that is not electroneutral. The net charge of the system is compensated for by using the q-potential (developed in Paper II) for handling long-range electrostatics, cancelling local moments at the cutoff. The potential artefacts introduced by compensating net charges of the system like has been studied using Ewald summation,[51] and the error was found to depend on the homogeneity of the system. Compensating the charge was found to introduce much stronger artefacts in heterogeneous systems involving for example proteins or membranes, whereas the approximation was more accurate for systems with more uniform charge distribution.

In our case, the system is expected to be sufficiently homogeneous for applying the charge compensation without significant errors, but the potential artefacts are still a matter under investigation.

Further, it is assumed that the experimentally observed values of $\Delta\Delta G$ and $\Delta\Delta H$ result mainly from electrostatic interactions. As we have seen, the LJ interaction decay relatively fast with r. Hence, compared to the Coulomb interaction, it is assumed to contribute little to $\Delta\Delta G$ since the LJ interaction between the potassium ions in the binding sites ($r \sim 15$ Å) is small compared to the Coulomb interaction. Therefore, we designed the model so that it describes the first complexation event by keeping the LJ parameters for the first potassium ion constant while increasing its charge from 0 to $+1$. The second complexation event is similarly described by keeping the LJ parameters for the second potassium ion constant while increasing its charge from 0 to $+1$, while the first ion is all the time fully interacting. The idea is to find out whether $\Delta\Delta G$ can be reproduced by only altering the electrostatic interactions. For

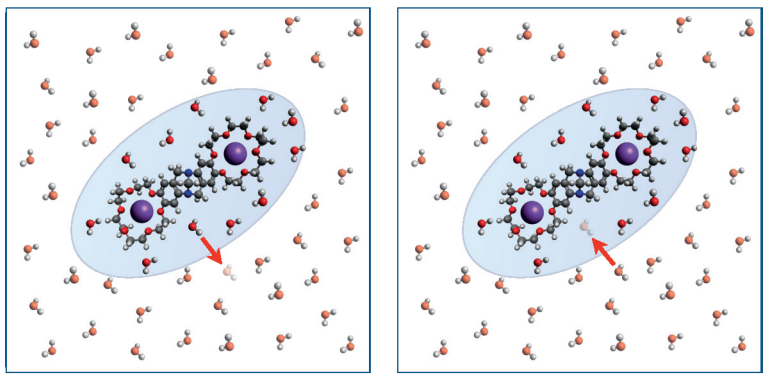

Figure 8.4: Schematic illustration of a 2D slice of the system used for sampling the binding of potassium ions to the BCETB. The scheme shows attempts to move water molecules across the boundary of R_{sol} which is illustrated as the blue ellips. The purple spheres in the crownethers are bound potassium ions, and the water molecules around the BCETB are assigned a darker or brighter color depending on if they are "in" or "out" molecules.

the purpose of this model, the MC simulation software Faunus is used. The charges on the potassium ions are altered by conducting charge trial moves in the range $q = [0, +1]$, where the Metropolis criteria decides whether or not to accept the trial move. Similarly to the example presented in Chapter 5, Subsection 2.1, the WL algorithm is then applied to estimate the density of states along the two reaction coordinates. In addition, an implementation of the preferential sampling method presented in Chapter 5, Subsection 1.1 is used. The region R_{sol} is defined as an ellipsoid enclosing the BCETB (see Figure 8.4), to preferentially move water molecules close to the two binding sites. A final approximation in the second approach is that the BCETB and the potassium ions are kept fixed, while the surrounding water molecules are allowed to move.

In both approaches, the difference in enthalpy associated a binding event is calculated using the linear form of the van't Hoff equation [52]

$$\frac{\Delta G^{\ominus}}{k_B T} = -\log K_{\text{assoc}} = -\frac{\Delta H^{\ominus}}{k_B T} + \frac{\Delta S^{\ominus}}{k_B}, \tag{8.2}$$

where $\Delta G^{\ominus}$ is the standard free energy of binding, K_{assoc} is the association constant, $\Delta H^{\ominus}$ is the standard enthalpy of binding, and $\Delta S^{\ominus}$ is the standard entropy of binding. This linear form of the equation is thus only valid if the changes in enthalpy and entropy are constant over the temperature range on which the linear regression is performed. Hence, it is an approximation since in reality, ΔH and ΔS do change with temperature. By choosing the temperature range sufficiently narrow it can however be assumed that the changes are negligible. By plotting the free energy difference as a function of $1/T$, the enthalpy of binding is thus obtained as the negative slope times the Boltzmann constant, k_B.

3 Preliminary results

Since the free energy of binding is needed in order to calculate the enthalpy of binding according to Equation 8.2, the first step towards developing an accurate model has been to try to reproduce the experimentally measured $\Delta\Delta G$ ($\Delta\Delta G^{\text{exp}} = 1.9$ $k_B T$). With the two approaches, we have achieved simulation results ranging from $\Delta\Delta G^{\text{sim}} = 0.1$ $k_B T$ to $\Delta\Delta G^{\text{sim}} = 1.9$ $k_B T$, depending on the choice of model parameters. Even though some sets of parameters using the first approach have yielded values of $\Delta\Delta G$ close to the experimental result (some actually gave $\Delta\Delta G^{\text{sim}} = 1.9$ $k_B T$), the corresponding single free energies of binding, $\Delta G^{\text{sim}}_{\text{bind},1}$ and $\Delta G^{\text{sim}}_{\text{bind},2}$, have deviated from the experimentally measured values, $\Delta G^{\text{exp}}_{\text{bind},1}$ and $\Delta G^{\text{exp}}_{\text{bind},2}$. Simulations of the first binding process show results ranging from $\Delta G^{\text{sim}}_{\text{bind},1} = -22.8$ $k_B T$ to $\Delta G^{\text{sim}}_{\text{bind},1} = 2.6$ $k_B T$ whereas for the second binding process, we have obtained values in the range $\Delta G^{\text{sim}}_{\text{bind},2} = -20.8$ $k_B T$ to $\Delta G^{\text{sim}}_{\text{bind},2} = 2.3$ $k_B T$. Some of these clearly deviates significantly from the experimental results, $\Delta G^{\text{exp}}_{\text{bind},1} = -4.7$ $k_B T$ and $\Delta G^{\text{exp}}_{\text{bind},2} = -2.8$ $k_B T$, and the current goal is thus to reproduce simultaneously $\Delta G^{\text{exp}}_{\text{bind},1}$, $\Delta G^{\text{exp}}_{\text{bind},2}$, and $\Delta\Delta G^{\text{exp}}$.

Chapter 9

Research and Outlook

1 Paper 1

The weakly hydrated thiocyanate anion, SCN^-, is located at the extreme of the salting-in end in the Hofmeister series.[9] This is apparent in two aspects: SCN^- shows a strong attraction to the air-water interface, making it one of the most surface active small inorganic ions, and it increases protein solubility by binding the the amide bond in the protein backbone,[53] making it relevant in biological contexts. The lack of a thermodynamically consistent force field of SCN^- that could reproduce properties in both bulk solution and the air-water interface motivated this work, resulting in the development of a classical, all-atom force field for the sodium and potassium salts of SCN^-, NaSCN and KSCN. As a consistency check, to validate the parameters on Na^+ and K^+, we decided to extend the force field to also include the salts NaCl, NaI, KCl and KI.

As a first step to derive the force field, the partial charges on S, C and N in the solvated thiocyanate ion were determined using quantum chemical calculations. Secondly, to achieve accurate bulk solution properties, the Lennard-Jones parameters on all ions were adjusted to reproduce experimental activity coefficients using molecular dynamics simulations and Kirkwood-Buff theory. Ion cluster formation, reorientational dynamics of SCN^- and solution density were also studied to further validate the bulk solution properties of the model. To validate the air-water interface behavior of the model, simulation data on excess surface tension and orientation of SCN^- at the air-water interface were compared with experiments.

Accurate activity derivatives were obtained for all salts except KI. The extent of ion

cluster formation in solutions of NaSCN and KSCN showed good agreement with experiments, and the slower reorientational dynamics observed for NaSCN compared to KSCN was captured. The excess surface tension was qualitatively reproduced for all salts, with quantitative agreement for most salts and concentrations, and the orientation of SCN^- at the air-water interface was in agreement with previous studies. Overall, the results suggest that our newly developed model is suitable for simulations both in bulk solution and the air-water interface. Although solution densities where reproduced for the Na^+ salts, they were systematically underestimated for all K^+ salts, which poses a limitation of our model, indicating the need for further reparametrisation of the K^+ ion.

Our model shows that K^+ preferentially coordinates the S atom in SCN^-, whereas Na^+ strongly attracts the N atom. The relative strength between K^+–S and Na^+–N interactions has consequences for bulk solution structure and dynamics. The weaker K^+–S interaction causes the K^+ ion to form a diffuse first coordination shell around the S atom, resulting in a stronger tendency to form ion pairs compared to NaSCN. The relatively strong Na^+–N attraction results in a more distinct first coordination shell, with the consequence of larger and more closely packed clusters being formed, particularly at high salt concentration. The larger clusters in solutions of NaSCN result in slower reorientational dynamics of SCN^- compared to KSCN. The effect of the cation is however not noticeably affecting the orientation of SCN^- at the air-water interface, which shows the same average angle, 44°, between the N–S bond and the normal to the air-water interface irrespective of cation, with the S atom pointing towards the vapor phase.

2 Paper II

The accuracy and efficiency of the scheme used to calculate long-ranged electrostatic interactions is important in the field of molecular simulation since it greatly affects the space and time feasible to simulate. The idea behind this study was to use a short-ranged potential to describe long-ranged electrostatics, potentially enhancing the efficiency of the calculation scheme while still producing accurate results. The theory behind the scheme is formally exact and has physical foundation in that it cancels an arbitrary number of electrostatic moments, P, at the cutoff.

The accuracy of the developed scheme was tested against several electrostatic models, with emphasis on the comparison with the commonly used Ewald and particle mesh Ewald (PME) methods. Using molecular dynamics simulations, we calculated a range of thermodynamic properties for water and salt solutions including radial distribution functions, activity derivatives obtained with Kirkwood-Buff theory, angular

correlation of water, dielectrics, diffusion coefficients and solution density.

The agreement with Ewald and PME was shown to depend strongly on how many electrostatic moments that are cancelled at the cutoff. For $R_c = 1.28$ nm, optimal values for the number of cancelled electrostatic moments turned out to be $P \in \{4, 5, 6\}$ for the systems studied, which resulted in good agreement for all calculated properties. The complexity of our newly developed scheme is $\mathcal{O}(N)$ and thus it scales linearly with the number of particles. The complexity of Ewald is, with optimal parameters, $\mathcal{O}(N^{3/2})$, whereas PME scales with $N\log(N)$. This suggests that our newly developed model is less costly than both Ewald and PME, thus providing an efficient alternative to describe long-range electrostatics.

3 Outlook

The overall aim of the work in this thesis has been to create a toolbox for describing the thermodynamics of systems containing water and ions. For this, simulation methods have been used in order to explain experimentally measured properties and get further insight into the underlying molecular mechanisms. For the binding of potassium ions to the BCETB, I will continue to refine the models so that they reproduce the experimentally measured values on the free energies and enthalpies of binding. Then, I plan to perform further analyses as to explain the observed difference in the enthalpy of binding to the first and second binding site from a molecular point of view.

Next, I would like to use KB theory in order to develop an atomistic model of seawater. Apart from the sodium, chloride, and potassium ions which were all parametrised in Paper I, the model will include other ions present in seawater such as the divalent magnesium and calcium ions. Subsequently, the aim is to apply the seawater model on problems of environmental relevance, for example chemistry occuring at the interface between mineral surfaces and seawater, and at the air/water interface of marine aerosols. For this, I expect to find the remainder of the methods presented in this thesis useful.

References

[1] Elizabeth A. Pillar, Marcelo I. Guzman, and Jose M. Rodriguez. Conversion of iodide to hypoiodous acid and iodine in aqueous microdroplets exposed to ozone. *Environmental Science & Technology*, 47(19):10971–10979, September 2013.

[2] Frank J. Millero, Rainer Feistel, Daniel G. Wright, and Trevor J. McDougall. The composition of standard seawater and the definition of the reference-composition salinity scale. *Deep Sea Research Part I: Oceanographic Research Papers*, 55(1):50–72, January 2008.

[3] Frank J. Millero. *Chemical oceanography*. CRC Press, fourth edition, 2013.

[4] Daan Frenkel and Berend Smit. *Understanding Molecular Simulation: From Algorithms to Applications*. Academic Press, second edition, 2002.

[5] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, June 1953.

[6] B. J. Alder and T. E. Wainwright. Phase transition for a hard sphere system. *The Journal of Chemical Physics*, 27(5):1208–1209, November 1957.

[7] B. J. Alder and T. E. Wainwright. Studies in molecular dynamics. i. general method. *The Journal of Chemical Physics*, 31(2):459–466, August 1959.

[8] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1):97–109, April 1970.

[9] Franz Hofmeister. Zur lehre von der wirkung der salze. *Archiv für Experimentelle Pathologie und Pharmakologie*, 25(1):1–30, August 1888.

[10] Terrell L. Hill. *An Introduction to Statistical Thermodynamics*. Dover Publications Inc, 1986.

[11] David Chandler and Jerome K. Percus. Introduction to modern statistical mechanics. *Physics Today*, 41(12):114–118, December 1988.

[12] W. Pauli. Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren. *Zeitschrift für Physik*, 31(1):765–783, February 1925.

[13] John D. Weeks, David Chandler, and Hans C. Andersen. Role of repulsive forces in determining the equilibrium structure of simple liquids. *The Journal of Chemical Physics*, 54(12):5237–5247, June 1971.

[14] Jacob N. Israelachvili. *Intermolecular and Surface Forces*. Elsevier, 2011.

[15] John E. Jones. On the determination of molecular fields. —II. from the equation of state of a gas. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, 106(738):463–477, October 1924.

[16] P. P. Ewald. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik*, 369(3):253–287, 1921.

[17] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: AnN·log(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, June 1993.

[18] D. Wolf, P. Keblinski, S. R. Phillpot, and J. Eggebrecht. Exact method for the simulation of coulombic systems by spherically truncated, pairwise r-1 summation. *The Journal of Chemical Physics*, 110(17):8254–8282, May 1999.

[19] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, January 1982.

[20] T. Schneider and E. Stoll. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B*, 17(3):1302–1322, February 1978.

[21] Shūichi Nosé. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2):255–268, June 1984.

[22] William G. Hoover. Canonical dynamics: Equilibrium phase-space distributions. *Physical Review A*, 31(3):1695–1697, March 1985.

[23] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, October 1984.

[24] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, December 1981.

[25] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. Oxford University Press, second edition, November 2017.

[26] J.C. Owicki and H.A. Scheraga. Preferential sampling near solutes in monte carlo calculations on dilute solutions. *Chemical Physics Letters*, 47(3):600–602, May 1977.

[27] Fugao Wang and D. P. Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical Review Letters*, 86(10):2050–2053, March 2001.

[28] Fugao Wang and D. P. Landau. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Physical Review E*, 64(5), October 2001.

[29] John E. Hunter and William P. Reinhardt. Finite-size scaling behavior of the free energy barrier between coexisting phases: Determination of the critical temperature and interfacial tension of the lennard-jones fluid. *The Journal of Chemical Physics*, 103(19):8627–8637, November 1995.

[30] Ola Engkvist and Gunnar Karlström. A method to calculate the probability distribution for systems with large energy barriers. *Chemical Physics*, 213(1-3):63–76, December 1996.

[31] Tushar S. Jain and Juan J. de Pablo. A biased monte carlo technique for calculation of the density of states of polymer films. *The Journal of Chemical Physics*, 116(16):7238–7243, April 2002.

[32] Qiliang Yan, Roland Faller, and Juan J. de Pablo. Density-of-states monte carlo method for simulation of fluids. *The Journal of Chemical Physics*, 116(20):8745–8749, May 2002.

[33] Nitin Rathore and Juan J. de Pablo. Monte carlo simulation of proteins through a random walk in energy space. *The Journal of Chemical Physics*, 116(16):7225–7230, April 2002.

[34] Björn Stenqvist, Axel Thuresson, Anıl Kurut, Robert Vácha, and Mikael Lund. Faunus– a flexible framework for monte carlo simulation. *Molecular Simulation*, 39(14-15):1233–1239, December 2013.

[35] Charles H Bennett. Efficient estimation of free energy differences from monte carlo data. *Journal of Computational Physics*, 22(2):245–268, October 1976.

[36] John G. Kirkwood and Frank P. Buff. The statistical mechanical theory of solutions. i. *The Journal of Chemical Physics*, 19(6):774–777, June 1951.

[37] Arieh Ben-Naim. *Water and Aqueous Solutions*. Springer US, 1974.

[38] Peter Krüger, Sondre K. Schnell, Dick Bedeaux, Signe Kjelstrup, Thijs J. H. Vlugt, and Jean-Marc Simon. Kirkwood–buff integrals for finite volumes. *The Journal of Physical Chemistry Letters*, 4(2):235–238, December 2012.

[39] Pritam Ganguly and Nico F. A. van der Vegt. Convergence of sampling kirkwood–buff integrals of aqueous solutions with molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 9(3):1347–1355, February 2013.

[40] Jasmin Milzetti, Divya Nayar, and Nico F. A. van der Vegt. Convergence of kirkwood–buff integrals of ideal and nonideal aqueous solutions using molecular dynamics simulations. *The Journal of Physical Chemistry B*, 122(21):5515–5526, January 2018.

[41] Kenneth E. Newman. Kirkwood–buff solution theory: derivation and applications. *Chem. Soc. Rev.*, 23(1):31–40, 1994.

[42] Charles J. Pedersen. Cyclic polyethers and their complexes with metal salts. *Journal of the American Chemical Society*, 89(26):7017–7036, December 1967.

[43] Alla Manjula and Madhavarao Nagarajan. New supramolecular hosts: Synthesis and cation binding studies of novel tröger's base-crown ether composites. *Tetrahedron*, 53(34):11859–11868, August 1997.

[44] Liem X. Dang and Peter A. Kollman. Free energy of association of the K^+:18-crown-6 complex in water: A new molecular dynamics study. *The Journal of Physical Chemistry*, 99(1):55–58, January 1995.

[45] Mark A. Thompson, Eric D. Glendening, and David Feller. The nature of K^+/crown ether interactions: A hybrid quantum mechanical-molecular mechanical study. *The Journal of Physical Chemistry*, 98(41):10465–10476, October 1994.

[46] Matthew C. Zwier, Joseph W. Kaus, and Lillian T. Chong. Efficient explicit-solvent molecular dynamics simulations of molecular association kinetics: Methane/methane, K^+/Cl^-, methane/benzene, and K^+/18-crown-6 ether. *Journal of Chemical Theory and Computation*, 7(4):1189–1197, February 2011.

[47] Carlos S. Arribas. *Tröger's base as scaffold for synthetic receptors*. PhD thesis, Lund University, 2007.

[48] S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7):845–854, February 2013.

[49] Szilárd Páll, Mark James Abraham, Carsten Kutzner, Berk Hess, and Erik Lindahl. Tackling exascale software challenges in molecular dynamics simulations with GROMACS. In *Lecture Notes in Computer Science*, pages 3–27. Springer International Publishing, 2015.

[50] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1-2:19–25, September 2015.

[51] Jochen S. Hub, Bert L. de Groot, Helmut Grubmüller, and Gerrit Groenhof. Quantifying artifacts in ewald simulations of inhomogeneous systems with a net charge. *Journal of Chemical Theory and Computation*, 10(1):381–390, January 2014.

[52] M. J. H. van't Hoff. Etudes de dynamique chimique. *Recueil des Travaux Chimiques des Pays-Bas*, 3(10):333–336, September 2010.

[53] Halil I. Okur, Jana Hladílková, Kelvin B. Rembert, Younhee Cho, Jan Heyda, Joachim Dzubiella, Paul S. Cremer, and Pavel Jungwirth. Beyond the hofmeister series: Ion-specific effects on proteins and their biological functions. *J. Phys. Chem. B*, 121(9):1997–2014, 2017.

Scientific publications

Author contributions

Paper I: Specific Cation Effects on SCN^- in Bulk Solution and at the Air-Water Interface

I performed part of the simulations and analysis, actively participated in discussions, and contributed to the writing of the paper.

Paper II: Generalized Moment Cancellation for Long-Range Electrostatics

I performed the simulations of the aqueous salt solutions and the subsequent analysis, actively participated in discussions, and contributed to the writing of the paper.

Models are essential in order to describe phenomena that we cannot directly see or measure. However, for a model to be useful it needs to be accurate yet simple enough so that it can be handled by whatever means that are available. The work in this thesis aims to describe intermolecular interactions in aqueous salt solutions, using atomistic simulation models. Interactions between molecules in these systems are of great importance in environmental chemistry, especially in chemistry involving seawater. While the models are designed to reproduce what has been observed from experimental measurements, they add further insight into the molecular mechanisms underlying these observations.

The thesis introduces the reader to the field of statistical thermodynamics, continues to describe how intermolecular interactions are modelled and how the theory is implemented in simulation, and ends with some applications.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

ISBN: 978-91-7422-743-7

Theoretical Chemistry
Department of Chemistry
Faculty of Science
Lund University