ALMA MATER STUDIORUM

Class LM-71 - Industrial Chemistry

Second Cycle Degree/Two Year Master in Industrial Chemistry

Department of Industrial Chemistry "Toso Montanari"

"In silico" seawater

Candidate:

Edoardo Errani

Supervisor:

Prof. Roberto Berardi

Cosupervisor:

Prof. Jose Luis Fernandez Abascal

Academic Year 2017-2018

Contents

List of Figures

List of Tables

Abstract

With "In silico" seawater we mean the calculation of some chemical-physics properties of seawater through computer simulations. Seawater is essentially an aqueous electrolyte solution characterized by a very complex composition, indeed an agreed Reference Composition required over a hundred years of study.

Since it has a such complex composition, it is easy to realize that the number of molecules needed to reproduce the Reference Composition Seawater is exceedingly large to be implemented in molecular simulations.

It seems thus interesting to design a simplified system with a composition close to the Reference Composition Seawater but more amenable to computer simulation. The first assumption we have made is to ignore minor constituents, i.e. ions with a mole fraction lower than a given value. In particular, since the mole fraction, X, of HCO_3^- is already very small ($X_{HCO_3^-} \approx 0.0015$) we may neglect it. Thus, our model for seawater would only deal in principle with Na^+, Cl^-, Mg^{2+}, SO_4^{2-}, Ca^{2+} e K^+.

One of the aims of this work is to investigate the contribution of the different ions to the properties of seawater. Following the same line of reasoning of the previous paragraph it seems interesting to analyse whether the replacement of Ca^{2+} and K^+ by Mg^{2+} and Na^+, respectively.

We may take the argument to the extreme: it makes sense to replace seawater by an equivalent NaCl solution? Despite the difficulty of defining an equivalence between seawater and NaCl solutions, the study of the latter system may be of interest. In this work, we have carried out a parallel study of the NaCl solution at "similar salinities" as those investigated for seawater.

Two different forcefields have been used in this work. For the first set we have chosen the Joung-Cheatham (JC) potential while the second one use the OPLS (optimized potentials for liquid simulations) forcefield. We have shown that the predictions for the JC and OPLS potentials are in good agreement with the experimental measurements for the thermodynamics and dynamics properties.

For the viscosity and density, the calculated data are compared with experimental. While for the structure and the self diffusion coefficient, for which there is no experimental data, the data calculated with the different models are compared to each other.

A first satisfactory model of seawater has been developed, but considerable efforts are still required to achieve a more accurate Force Field.

Abstract

Con "In silico" seawater si intende il calcolo di alcune proprietà chimico fisiche dell'acqua marina attraverso simulazioni al calcolatore. L'acqua marina è essenzialmente una soluzione elettrolitica caratterizzata da una composizione molto complessa, si pensi solo che per definire una composizione di riferimento ci sono voluti più di 100 anni di studio.

Data l'enorme complessità di tale soluzione è facile rendersi conto che il numero di molecole necessarie per la riproduzione della composizione dell'acqua marina è estremamente grande per essere implementato in simulazioni molecolari.

In questo lavoro viene descritto un modello semplificato di acqua marina adottato per una simulazione numerica. In prima approssimazione, sono stati trascurati gli ioni a bassa concentrazione, come, ad esempio, HCO_3^- ($X_{HCO_3^-} \approx 0.0015$). Gli ioni considerati sono Na^+, Cl^-, Mg^{2+}, SO_4^{2-}, Ca^{2+} e K^+.

Uno degli scopi principali di questo lavoro è quello di indagare il contributo dei diversi ioni sulle proprietà dell'acqua marina. Seguendo questa linea di ragionamento è stato interessante analizzare la sostituzione degli ioni Ca^{2+} e K^+ con Mg^{2+} e Na^+, rispettivamente.

Le proprietà dell'acqua marina sono quindi state confrontate con quelle di una soluzione di NaCl, che rappresenta il soluto maggioritario.

In questo lavoro di tesi sono stati usati due diversi *Force Field* (FF): il Joung-Cheatham (JC) e l'OPLS (Optimized Potentials for Liquid Simulations). Abbiamo dimostrato che entrambi i potenziali danno un risultato soddisfacente nel riprodurre le proprietà termodinamiche e dinamiche dell'acqua marina.

Nel caso della viscosità e della densità, i dati calcolati sono comparati con dati sperimentali. Mentre per quanto riguarda la struttura e i coefficienti di diffusione, per i quali non vi è alcun dato sperimentale per l'acqua marina, sono comparati tra di loro i dati calcolati con i diversi modelli.

È stato quindi sviluppato un primo modello soddisfacente di acqua marina, ma per realizzare un Force Field più accurato sono ancora necessari notevoli sforzi.

Chapter 1

Introduction

1.1 Overview of significant seawater properties

Seawater is water from a sea or ocean. Since the word sea is often used interchangeably with "ocean", in this work we will refer to the saline water contained in seas and oceans as "seawater". The seas and oceans contain 97% of Earth's water. Seawater is integral to life, forms part of the carbon cycle, and influences climate and weather patterns. Seawater is essentially an aqueous electrolyte solution. Although the ionic composition is quite complex, most of the water properties are essentially dependent (apart of the thermodynamic conditions) on the salinity, i.e. the amount of dissolved salts.

The salinity of the surface seawater is greatly influenced in coastal regions by the fresh water flow from rivers, and in polar regions by the processes of freezing and thawing of ice. Also, the salinity of water in the surface layer of oceans is dependent on evaporation and precipitation. Leaving aside these surface or coastal effects, on average, seawater has a salinity of about 3.5%. This means that every kilogram (roughly one litre by volume) of seawater has approximately 35 grams of dissolved salts. However, the salinity changes from one ocean to another and, for normal open ocean, ranges between 33 and 37% (see Fig. 1.1). Several factors influence the saline content of the oceans. Evaporation accounts for the larger salinity of equatorial regions while the delivery of fresh water from the melting of ice seems to be the reason of the low levels of salinity in polar areas. In addition to these factors it is worth noting the existence of ocean currents. Below, we will comment more thoroughly the characteristics and importance of the ocean currents. Salinity is a fundamental property of seawater and basic to understanding biological and physical processes in oceans. Many processes that have significant socio-economic impacts depend critically on salinity. These are not limited to fishing productivity but also include biological effects on the preservation of marine ecosystems (coral reefs, estuary and coastal wetlands, mangroves), as well as the development of harmful algal
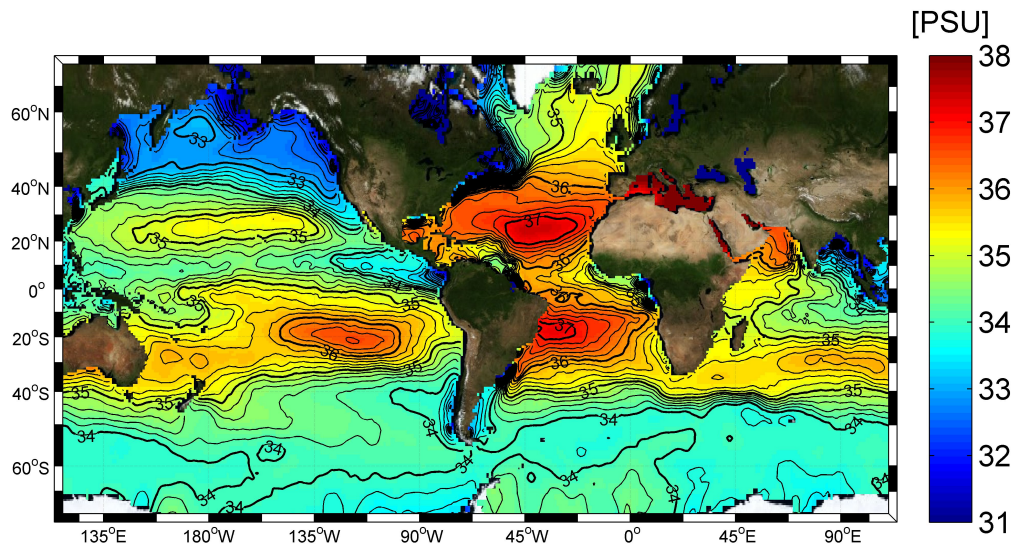
Figure 1.1: **Horizontal distribution of salinity in Earth's oceans.**

blooms or jellyfish invasions, survival of invasive species, etc.

The average temperature of the ocean surface waters is about 17 degrees Celsius and the temperature varies mainly with latitude. The polar seas (high latitude) can be as cold as -2 degrees Celsius while the Persian Gulf (low latitude) can be as warm as 36°C. Then, strictly speaking, the relevant range of seawater only spans about 40 degrees. However, in desalination processes the temperature of the brine can be higher than 100°C. This extends considerably the range of possible relevant temperatures for saline water.

Besides the horizontal changes of temperature and salinity there is a vertical profile for both magnitudes. Bodies of seawater are made up of layers, determined by temperature and salinity. Most of the heat energy of sunlight is absorbed in the first few centimeters at the ocean's surface, which heats during the day and cools at night as heat energy is lost to space by radiation. The effect of the wind creates turbulence, thus mixing water to create a layer of almost constant temperature, salinity and density called the wind-mixed layer. Below this mixed layer, the temperature, salinity and density remain relatively stable over day/night cycles and change gradually as the depth increases. The profile of the thermocline depends on the latitude. It is shallow to nonexistent in the polar regions (the water column is cold from the surface to the bottom) and variable in temperate regions. Fig. 1.2 shows a tropical ocean thermocline. Note the rapid change between 100 and 1000 meters. It is also interesting that the temperature is nearly constant after 1000 meters depth. The pattern of a typical halocline (Fig. 1.3) is quite similar to that of the thermocline represented in Fig. 1.2. Average density of seawater at the surface is 1.025 kg/L. Thus, seawater is denser than pure water (density 1.0 kg/L at 4C). The density of seawater depends essentially on salinity and temperature so it changes with latitude
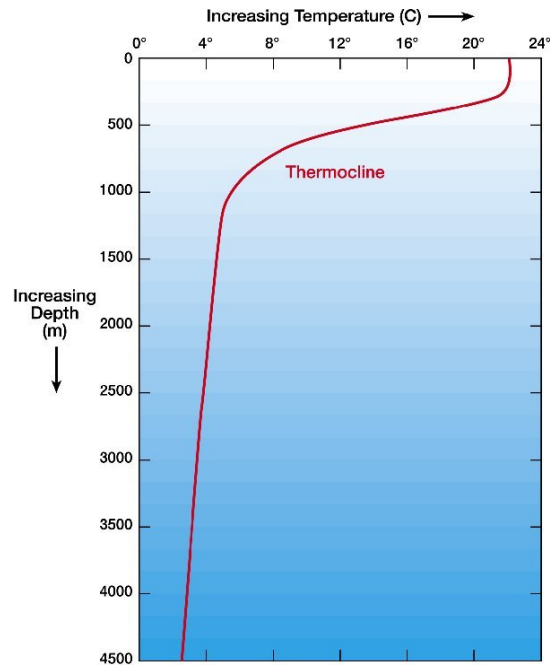
Figure 1.2: **Typical thermocline at tropical latitudes.**

and depth accordingly with the variations of these properties. Thus, the vertical profile of the density is almost constant at the polar regions but changes with depth at other latitudes: it increases in the first 1000 meters but changes very little at higher depths (see Fig. 1.3).

1.2 Thermohaline circulation and its effects on global climate

How are ocean currents caused, and how do they affect climate? These questions were hotly debated in the nineteenth century. Some argued that water is simply pushed along by the wind; others postulated convection currents caused either by heating and cooling or by evaporation and precipitation. Even today, the driving forces and climatic effects of ocean currents are still not completely understood. In 1908, Johan Sandström[4] laid the foundations of our modern understanding of ocean currents and elucidated the properties of wind-driven and thermal circulation. The latter term was amended by the 1920s to **thermohaline circulation**[4, 5], because water density in the ocean is determined by both temperature and salinity.

Sandström found that thermal gradients can give rise to a steady circulation only if heating occurs at a greater depth than cooling, a fact that is familiar to oceanography students as Sandstorm's theorem. But fluxes of heat and freshwater occur mostly at the ocean's surface. So what is the deep heat source that drives the ocean's observed thermohaline circulation? Sandström recognized that it is the downward penetration of heat at low latitudes, due to
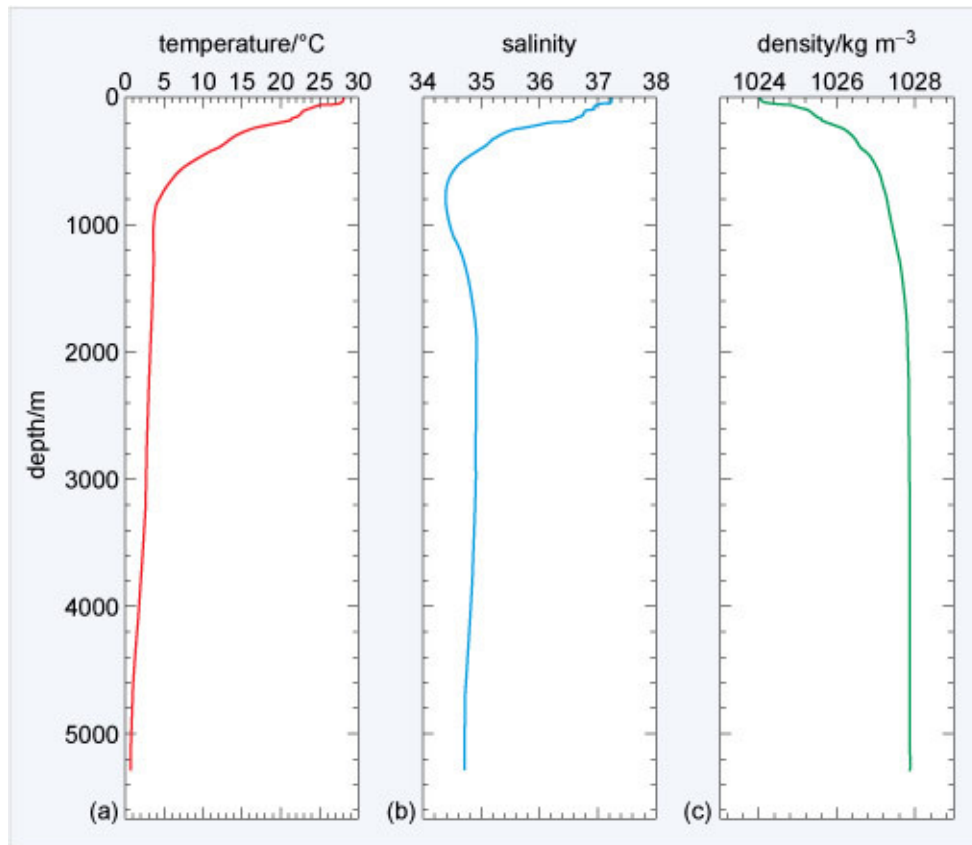
Figure 1.3: **Vertical profiles of temperature, salinity and density at tropical latitudes.**

turbulent mixing, that provides this thermal engine. Turbulent mixing is powered by winds and tides. Thermohaline circulation is thus caused by the joint effect of thermohaline forcing and turbulent mixing. It can be defined as currents driven by fluxes of heat and freshwater across the sea surface and subsequent interior mixing of heat and salt.

Thermohaline and wind-driven currents cannot therefore be separated by oceanographic measurements. There are thus two distinct forcing mechanisms, but not two separate circulations. Change the wind stress, and the thermohaline circulation will change; alter thermohaline forcing, and the wind-driven currents will also change. However, although thermohaline circulation is not measurable, the concept is still a useful one, and modern models of oceans can be used to carry out computer experiments to study the properties of these currents.

As we can see in Figure 1.4 at high latitudes the water sinks, both for the low temperature and for the high salinity caused by the formation of the ice floe. Moving towards the equator, the bottom water decreases its density by interacting with the other waters and tends to rise, particularly in the south of the Indian Ocean. The ascent of deep water (upwelling) promotes biological productivity as it causes the ascent of mineral nutrients. The water masses involved in this circulation carry both energy (in the form of heat) and materials (dissolved substances, gases and particles) with the consequence of significantly affecting both the terrestrial climate
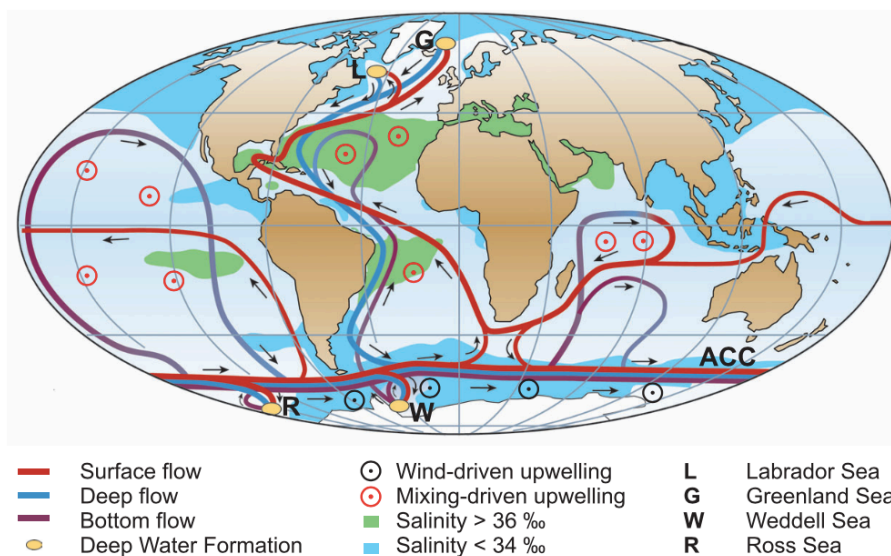
Figure 1.4: **Schematic representation of the global thermohaline circulation. Surface currents are shown in red, deep waters in light blue and bottom waters in dark blue. The main deep water formation sites are shown in orange. (from Kuhlbrodt et al.[1] modified after[2]).**

and marine biology. Simulations in which the ocean's heat transport is switched off consistently show a large winter cooling over the northern Atlantic and adjacent land areas, reaching several degrees in inland Europe, up to 10°C over Greenland and even exceeding 20°C over the Nordic seas. This heat transport warms the climate on both sides of the Atlantic, and is therefore not the main reason that Europe is warmer than America. This phenomenon is mainly due to the prevailing winds in the two regions. Nevertheless, ocean currents do make the northern Atlantic much warmer than at comparable latitudes in the northern Pacific. Changes in these currents are our best explanation for the abrupt and marked climate swings that occurred over the north Atlantic many times during the most recent glacial period, as shown by Greenland's ice cores and by deep-sea sediments. Circulation changes might again be triggered by global warming.

1.3 'In silico' seawater

Molecular computer simulation[6, 7] is a powerful technique that allows to calculate the properties of a molecular system from first principles or from the knowledge of the interactions (the forcefield) between the molecules. Computer simulation provides information on properties not easily accessible by experiment. For instance, due to the complex composition, the detailed ionic structure of seawater cannot be resolved from diffraction experiments. Also, the relative

contribution of each ion to the final properties of the solution can not be obtained experimentally but it is trivially evaluated in computer simulation calculations. The goal of this work is to investigate for the first time the possibility of accounting for the physico-chemical properties of seawater from scratch using the state of the art computer simulation techniques and knowledge of water and ionic forcefields. We have mentioned that the thermohaline circulation depends essentially on the gradients of density which, in turn, depend on the gradients of temperature and salinity.

Temperature is a well defined thermodynamic variable but, what is salinity? Salinity[8] is the amount of dissolved compounds (mostly ions such as chlorine, sodium, magnesium, sulfate, calcium, potassium and others) within seawater. As commented above, the salinity changes from a point of an ocean to another. But how about the composition? An essential property of seawater is the Principle of Constant Proportions: the ratio between the concentrations of the different components of seawater is essentially constant and independent of the total amount of ions. This simplifies considerably the problem

The aim of the present thesis is then the development of a model that can reproduce and predict, through molecular dynamics (MD) simulations carried out in a computer, the properties of seawater at different temperatures and salinities. More specifically, the aims of this work are:

1. The development of a *model* for seawater solutions based on the successful TIP4P/2005 water model[9].

2. To determine if two widely employed forcefields, namely, OPLS[10] and Joung-Cheatham (JC)[11], are able to reproduce thermodynamic, dynamic and structural properties of seawater.

3. To analyse the relative contributions of the different ions. This will allow to check to what extent the model can be simplified by reducing the number of ions of the system while preserving the predictions for the seawater properties.

The structure of the thesis is as follows. Chapter 2 annotates the composition of seawater. It also describes two important standards defined by the scientific community: the Reference Seawater Composition and the Absolute Salinity. Chapter 3 provides the theoretical framework of the molecular dynamics technique, the simulation tool used in this work. In the fourth chapter we will comment the methodology. There, we will discuss the specific features of the implemented seawater composition and give the technical details of the simulations. The presentation of the results is given in Chapter 5. In particular, we will analyse our calculations for

the density, viscosity and radial distribution functions at different temperatures and salinities. A summary of the main conclusions (Chapter 6) will close this thesis.

8

Chapter 2

Seawater Composition and Salinity

2.1 Seawater composition

2.1.1 Principle of Constant Proportions

The first investigations of the major inorganic components of seawater were made by A. Marcet in 1819[12]. He wrote "... all the specimens of sea water which I have examined, however different in their strength, contain the same ingredients all over the world, these bearing very nearly the same proportions to each other; so that they differ only as to the total amount of their saline contents". In 1862 Forchhammer reported the determination of the concentrations of Cl^-, SO_4^{2-}, Mg^{2+}, Ca^{2+} and K^+ directly and Na^+ by difference[13]. He made these measurements on several hundred surface water samples from all parts of the world. Based on these analyses, he concluded that the ratio of major salts in samples of seawater from various locations was almost constant, consistent with earlier speculation by Marcet. This constant ratio is known as Forchhammer's Principle, or the Principle of Constant Proportions. In addition to this principle, Forchhammer is credited with defining the term salinity to mean the concentration of major salts in seawater. Some years later, Dittmar[14] made careful determinations on 77 water samples representative of all oceans collected on the voyage around the world of the H.M.S. *Challenger*. Besides the major components, he measured the concentrations of Br^- and CO_3^-. The results of Dittmar's work showed that there were no significant regional differences in the relative composition of seawater.

Although the ion proportions are virtually constant, the constancy is not absolute. This lead to the proposal of a standard representing the composition of seawater. Martin Knudsen, at the international conference for the exploration of the sea held in Stockholm in 1899[15], proposed to define arbitrarily as "Standard Seawater" certain samples taken from the North Atlantic surface water in a specific region. Modern studies of the major components of Standard

Seawater were carried out by Cox, Culkin and Riley[16–19] as part of an international study of the salinity of seawater. Measurements of Carpenter and Manella[20] for Mg^{2+} showed that previous determinations were incorrect. Later work has been done to determine the amount of minor components such as F^-[21] and boron[22].

2.1.2 Reference Composition and Reference Seawater

Along the years it became evident the need of a benchmark for the composition of Standard Seawater. In 2008 a *Reference Composition* — consisting of the proportions of major components of Atlantic surface seawater (referred to as Standard Seawater) — was *defined* using earlier analytical measurements. The stoichiometry of sea salt introduced there was thus based on the most accurate prior determinations of the composition, adjusted to achieve charge balance and making use of the 2005 atomic weights[3]. The Reference Composition defines the mole fractions (see Table 2.1) so that each mole fraction, X_i, when multiplied by 10^7, is an integer.

The Reference Composition gives only the relative abundance of the solutes in seawater but it does not give any indication of the water content. *Reference Seawater* is defined as seawater with Reference Composition solute dissolved in pure water as the solvent and adjusted to its thermodynamic equilibrium state.

2.2 Salinity

2.2.1 Chlorinity and Practical Salinity

Earlier determinations of salinity were made with time-consuming gravimetric analysis. Besides, the procedure is subject to the chemical difficulty in drying the salts in seawater. Determination of the salinity using chemical analysis is difficult for several reasons: (1) some of the dissolved substances (namely, Cl^- and Na^+) are present in high concentrations while others are present in such minute quantities that they have not even been detected in seawater although they have been found in marine organisms or salt deposits; (2) two of the major constituents (Na^+ and K^+) are difficult to determine accurately (in fact, Na^+ is usually obtained indirectly); (3) it is extremely difficult to separate chemically related substances (Ca^{2+}/Sr^{2+}). However, once the Principle of Constant Proportions was well established, new methods of evaluating the salinity were at hand.

At the end of the nineteenth century, Dittmar showed that the evaluation of salinity could be done as well, if not better, by determining the weight of chlorine in a sample. The calculation

Table 2.1: The ionic stoichiometry used in the definition of the Reference Composition, expressed in mole fraction of ions X_i. Third column gives the mass fractions, W_i, using the 2005 atomic weights[3]. Fourth column presents the mass ratios respect to chlorinity r_i. Using the defined X_i and the current accepted values for the molecular masses, the average atomic weight of sea salt is 31.4038218 g/mol. Last columns give the properties of KCl-normalized Reference Seawater: mass per kg of solution w_i, and molality m_i. The sum of the fifth column is the special Reference Salinity corresponding to $S = 35$, S_R^{35}, which is exactly 35.16504 g/kg.

Solute i	Reference Composition			KCl-normalized Reference Seawater	
	$10^7 X_i$ (Definition)	W_i	r_i	w_i(g/kg)	m_i(mol/kg)
Na^+	4 188 071	0.3065958	0.556492	10.78145	0.4860597
Mg^{2+}	471 678	0.0365055	0.066260	1.28372	0.0547421
Ca^{2+}	91 823	0.0117186	0.021270	0.41208	0.0106568
K^+	91 159	0.0113495	0.020600	0.39910	0.0105797
Sr^{2+}	810	0.0002260	0.000410	0.00795	0.0000940
Cl^-	4 874 839	0.5503396	0.998904	19.35271	0.5657647
SO_4^{2-}	252 152	0.0771319	0.140000	2.71235	0.0292643
HCO_3^-	15 340	0.0029805	0.005410	0.10481	0.0017803
Br^-	7520	0.0019134	0.003473	0.06728	0.0008728
CO_3^{2-}	2134	0.0004078	0.000740	0.01434	0.0002477
$B(OH)_4^-$	900	0.0002259	0.000410	0.00795	0.0001045
F^-	610	0.0000369	0.000067	0.00130	0.0000708
OH^-	71	0.0000038	0.000007	0.00014	0.0000082
$B(OH)_3$	2807	0.0005527	0.001003	0.01944	0.0003258
CO_2	86	0.0000121	0.000022	0.00042	0.0000100
Sum	10 000 000	1.0000000	1.815069	35.16504	1.1605813
H_2O				964.83496	55.5084720
Sum				1000.00000	56.6690534

of the chlorine content can be made with a rapid titration of the seawater against silver nitrate solution. Notice that the bromide and iodide anions also precipitate with silver. Accordingly, chlorinity Cl was firstly defined as the weight of chlorine in grams per kilogram of seawater after the bromides and iodides had been replaced by chlorides. Multiplying the chlorinity by a constant factor gives the salinity.

In the 1960s and 1970s, various properties of seawater samples were studied quantitatively as a function of chlorinity. In fact, the composition of Standard Seawater was usually represented by the measured values of r, the mass ratio of the major components of seawater relative to the chlorinity Cl. Properties derived experimentally from Standard Seawater samples allowed the mathematical construction of quantitative relations between the properties such as the equation of state of seawater $\rho = \rho(Cl, T, p)$.

Chlorinity is not the only quantity that has been used as a 'standard' measure of salinity; conductivity has also served this purpose. The two approaches are essentially equivalent for Standard Seawater (Millero et al., 1977), but the resulting values for derived quantities such as densities are not precisely consistent when composition anomalies are present, as for example in the Baltic Sea (Millero and Kremling, 1976). Thus, prior to 1978, two incompatible definitions of salinity were in use, one based on the chlorinity, the other one on the conductivity of seawater. This fact complicated the history of salinity measurements and caused severe problems in the comparability of their results.

In 1978 the Joint Panel on Oceanographic Tables and Standards recommended the Practical Salinity Scale (PSS-78) to all oceanographic organizations as the scale in which to report future salinity data. In the formulation of PSS-78[23], the Practical Salinity, S, of a sample of seawater is defined in terms of the ratio K_{15} of the electrical conductivity of the seawater sample at the temperature of 15°C and the pressure of one standard atmosphere, to that of a potassium chloride (KCl) solution, in which the mass fraction is 32.4356×10^{-3}, at the same temperature and pressure. The K_{15} value exactly equal to 1 corresponds, by definition, to a practical salinity exactly equal to 35[23, 24].

2.2.2 Absolute and Reference Salinity

Absolute Salinity S_A is defined as the mass fraction of dissolved material in seawater. A precise direct experimental determination of S_A is practically impossible. One of the purposes of the introduction of the Reference Composition was just to determine a "best estimate" of the Absolute Salinity for Standard Seawater. This new scale is the Reference-Composition salinity S_R (or Reference salinity for short).

Seawater with a Practical Salinity of 35 has provided a benchmark in past discussions of salinity. Let us refer to it as KCl-normalized Reference Seawater. By definition the Reference Salinity of the normalized Reference Seawater is fixed to our current best estimate for the the Absolute Salinity of this solution, namely 35.165031 g/kg[25]. The complete definition of S_R is

(a) The *Reference-Composition Salinity of pure water* is defined to be zero.

(b) The *Reference-Composition Salinity of normalized Reference Seawater* is defined to be exactly 35.16504 g kg^{-1}.

(c) The *Reference-Composition Salinity* is defined to be conservative during mixing. If a seawater sample of mass m_1 and Reference-Composition Salinity S_{R1} is mixed with another seawater sample of mass m_2 and Reference-Composition Salinity S_{R2}, the final Reference-Composition Salinity S_{R12} of this sample is

$$S_{R12} = \frac{m_1 S_{R1} + m_2 S_{R2}}{m_1 + m_2} \tag{2.1}$$

Negative values of m_1 and m_2 corresponding to the removal of seawater with the appropriate salinity are permitted.

(d) The *Reference-Composition Salinity of a seawater sample at a given temperature and pressure* is equal to the Reference-Composition Salinity of the same sample at another temperature and pressure when the transition process is conducted without exchange of matter, in particular, without evaporation, precipitation or degassing of substance from the solution.

Table 2.1 presents the weight fractions and molalities of the constituents of the KCl-normalized Reference Seawater. Notice that the Reference Salinity of this solution (the sum of the masses of the ionic components in the fifth column) gives 35.16504 g/kg in accordance with the above definitions.

The definition of the Reference Salinity is particularly suited for theoretical and simulation calculations. On the other hand, the possibility of determining a correction for the Absolute Salinity for actual samples whose composition differ from the Reference Composition it is also possible[26]. The largest such corrections would be based on measurements of changes in the carbonate system and the addition of $CaCO_3$, CO_2 and nutrients from the oxidation of plant material. Since this work is devoted to the properties of Reference Seawater there is no need for such corrections.

Chapter 3

Molecular Dynamics

Following the earlier success of Monte Carlo simulations, Alder and Wainwright[27] developed the method of molecular dynamics (MD) to simulate perfectly elastic collisions between hard spheres in an IBM 704 computer. Loup Verlet[28] made the first MD simulation for a system interacting via the Lennard-Jones potential[29]. Rahman and Stillinger[30] were able to show that a more realistic (and relatively complex) system like liquid water could be described using a potential model proposed by Fermi.

Molecular Dynamics (MD) is a computer simulation technique for studying the physical motion of atoms and molecules and to predict the macroscopic properties of a system from the molecular interactions between particles[6, 7]. The atoms and molecules are allowed to interact for a fixed period of time, giving a view of the dynamic evolution of the system. In the most common version, the trajectories of atoms and molecules are determined by numerically solving the Newton's equations of motion for a system of interacting particles. The forces between particles and their potential energies are often calculated using interatomic potentials or molecular mechanics forcefields. For this reason it can be defined as a deterministic method. This approach is successful within the limits in which it is not necessary to take into account the typical nuclear quantum effects of, for example, very small atoms or light molecules. A large number of trajectories are derived by integrating over time the Newton's equations of motion. From the ensemble of trajectories, classical statistical mechanics is employed to evaluate the averages leading to the macroscopic properties of the system. Thermodynamic, dynamical or structural properties are accessible using the MD method.

3.1 Force Field

The classical MD method rely on the knowledge on the interaction potential between the molecules. These are usually referred to as the forcefield[31]. The most commonly used force-

fields represent the potential energy of the system as the sum of two contributions, namely intramolecular and intermolecular, corresponding to interactions between bounded and unbounded atoms

$$V(r) = V_{bond}(r) + V_{unbond}(r) \tag{3.1}$$

The interaction between bonded atoms, may include stretching, bending and torsion terms and takes a form of the type

$$V_{unbond}(r) = \sum_{i=1}^{N_{stretch}} k_i^{str}(r_i - r_{0i})^2 + \sum_{i=1}^{N_{bend}} k_i^{ben}(\theta_i - \theta_{0i})^2 + \sum_{i=1}^{N_{tors}} k_i^{tor}[1 + \cos(n_i\phi_i - \gamma_i)], \tag{3.2}$$

where the constants r_{0i} and θ_{0i} represent bond distances and bond angles with corresponding k_i^{str} and k_i^{ben} strength constants, ϕ_i is the torsional angle for the i-torsion while k_i^{tor}, n_i e γ_i are parameters whose physical meaning depends on the atoms involved. The sums are extended respectively to all stretching, rotations and torsions present in the molecules under consideration. In this work the intramolecular potential has been neglected because we consider the water molecules as rigid.

The intermolecular potential can be defined as the difference of energy when the molecules are in a specific configuration (given by the coordinates $\vec{r}_1, \omega_1, ..., \vec{r}_N, \omega_N$), and the energy of these molecules when they are at infinite separation ($V = 0$). In principle, the determination of V requires quantum calculation. In a liquid, for example, the number of molecules is of the order of the number of Avogadro, so that a quantum calculation of its energy is impossible (quantum calculations can currently address systems of the order of up to ≈ 100 molecules). So we work with an empirical expression of V which is known by the name **Force Field**. The most common forcefields are based on pair-additive potentials. In this approximation the total energy of the system is a sum of the interactions between pairs of molecules (n, m), as seen in the following equation:

$$V_{unbond} = \sum_{n} \sum_{m>n} V_{nm} \tag{3.3}$$

Often, the molecules consist of a number of interaction sites (which may or may not be coincident with the atoms of the molecule) and the potential energy between a pair of molecules is in turn a summation over all possible site-site interactions

$$V_{nm} = \sum_{i} \sum_{j>i} V_{ij}(\mathbf{r_{ij}}). \tag{3.4}$$

In most cases, the interaction between pairs of sites of different molecules, V_{ij}, contains terms. The first correspond to exchange-dispersion forces (usually represented by a Lennard-Jones potential, and a coulombic interaction coming from the partial charges q_i, q_j placed respectively

at sites i and j of molecules n and m. Thus,

$$V_{ij}(\mathbf{r}_{ij}) = 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{1}{4\pi\varepsilon_0} \frac{q_i q_j}{r_{ij}}. \tag{3.5}$$

where ϵ_{ij} represents the depth of the potential well, and σ_{ij} is the distance at which $V_{ij}(\mathbf{r}_{ij}) = 0$. The minimum of the potential occurs at $r_e = 2^{1/6}\sigma_{ij}$ (see Figure 3.1). Notice that the first term of the Lennard-Jones potential, representing the repulsive forces, is short-ranged (it is always positive and inversely proportional to the twelfth power of r. The attractive term (negative and inversely proportional to the sixth power of r) is also short ranged. On the contrary, the coulombic interactions may be attractive (between unlike-charged sites) or repulsive (between like-charged sites) but are long ranged since they depend on r^{-1}. In the above expression, ε_0 is the dielectric constant in vacuum.
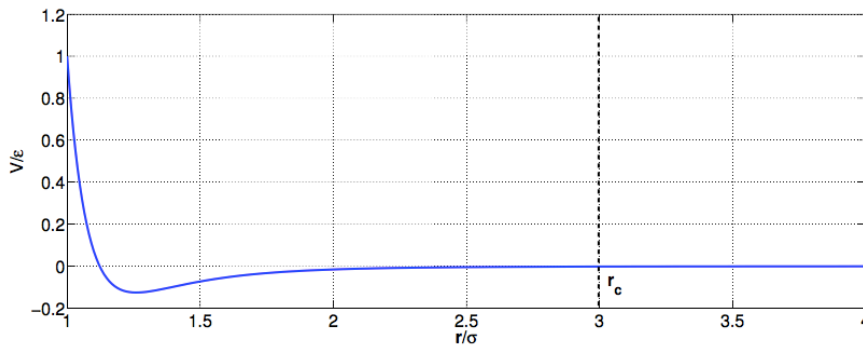
Figure 3.1: **Lennard-Jones potential. The dotted line shows the typical cutoff distance, r_c**

3.2 Classical Molecular Dynamics

Let us consider a system of classical particles and a forcefield describing their interactions. Classical Molecular Dynamics (MD) is the technique allowing to generate a number of trajectories in a given thermodynamic ensemble. In the microcanonic ensemble (also termed as NVE), the trajectories are generated by solving the Newton's law which relates the total force acting on the element i to the variation produced in its velocity:

$$\vec{F}_i = m_i \frac{\partial \vec{v}_i}{\partial t} = -\frac{\partial V\left(\vec{r}^N\right)}{\partial \vec{r}_i}. \tag{3.6}$$

As seen in equation (3.6), the force acting on i may be simply evaluated as the (minus) gradient of the potential energy V at the position $\vec{r}_i$. Assuming that we know $\vec{r}_i(0)$ y $\vec{v}_i(0)$ for a starting configuration, equation (3.6) allows the calculation of the new positions of the particles at a time δt, thus generating a new configuration in the NVE ensemble. Analytical solution of the Newton equation is not possible for a molecular system. Thus, one must resort to numerical algorithms for differential equations to solve Eq. 3.6.

3.3   Verlet's algorithm

The intent of a MD simulation is to derive the trajectories of the particles by solving Newton's equations over time. In order to implement this calculation, the simulation time must be discretized so the integration is done for finite intervals. The time intervals take the name of timestep (Δt) and the algorithms based on this strategy are called finite difference methods. One of the widely used algorithms in MD is the *Verlet's algorithm*[28]. His basic idea is to write two Taylor series approximations truncated at the third term (where the accelerations appear)

$$\vec{r_i}(t + \delta t) = \vec{r_i}(t) + \vec{v_i}(t)\delta t + \frac{1}{2}\vec{a_i}(t)(\delta t)^2, \tag{3.7}$$

$$\vec{r_i}(t - \delta t) = \vec{r_i}(t) - \vec{v_i}(t)\delta t + \frac{1}{2}\vec{a_i}(t)(\delta t)^2, \tag{3.8}$$

where $\vec{v_i}$ is the velocity and $\vec{r_i}$ is the position of the particle i at time t. We want to know the velocity and the positions in a time $t + \delta t$. Adding both equations we obtain

$$\vec{r_i}(t + \delta t) = 2\vec{r_i}(t) - \vec{r_i}(t - \delta t) + \vec{a_i}(t)(\delta t)^2. \tag{3.9}$$

Let us analyze the previous expression in more detail. The first term is the position of a particle at time t, the second is the position of the same particle at time $-\delta t$ (i.e., at the previous timestep) and the last one depends on the acceleration. To get the new positions, the only unknown variable is the acceleration of the particle, which can be determined from equation (3.6)

$$\vec{a_i} = \frac{\partial \vec{v_i}}{\partial t} = -\frac{\nabla_i V(\vec{r_1}.....\vec{r_N})}{m_i}. \tag{3.10}$$

Once the accelerations are known, the velocities can be evaluated by finite differences as follows

$$v_i = \frac{\partial \vec{r_i}}{\partial t} \approx \frac{\Delta r_i}{\Delta t} = \frac{\vec{r_i}(t + \delta t) - \vec{r}(t - \delta t)}{2\delta t}. \tag{3.11}$$

In summary, the steps required to perform a *Molecular Dynamics* simulation using the *Verlet's algorithm* are:

1. Setup the initial configuration with the positions and velocities of the particles.

2. Evaluate the forces on the particles in that configuration.

3. Calculate the acceleration with the Newton equation.

4. Determine the positions and velocities for the new configuration.

5. Repeat the procedure from step 2 several million times.

For each configuration, a number of properties may be evaluated (either on the fly or after the simulation if the configurations are saved). The potential energy is readily accessible when evaluating any new configuration. Also, the kinetic energy (E_{kin}), is easily obtained $E_{kin} = \frac{1}{2}mv^2$. Notice that the total energy $E_{tot} = E_{kin} + E_{pot}$ must be conserved along the whole simulation. This limits the timestep to values around 1-2 fs. In fact, the stability of the solution of any differential equation is very sensitive to the details of the algorithm.

An alternative integration method is the so-called *Velocity Verlet* algorithm whose main advantage is that it provides directly the velocities. It proceeds according to the steps described below:

$$\vec{v_i}(t + \frac{1}{2}\delta t) = \vec{v_i}(t) + \frac{1}{2}\vec{a_i}(t)(\delta t) \tag{3.12}$$

$$\vec{r_i}(t + \delta t) = \vec{r_i}(t) + \vec{v_i}(t + \frac{1}{2}\delta t)\delta t \tag{3.13}$$

$$\vec{a_i}(t + \delta t) = -\frac{1}{m}\nabla \vec{V_i}(r(t + \delta t)) \tag{3.14}$$

$$\vec{r_i}(t + \delta t) = \vec{v_i}(t + \frac{1}{2}\delta t)\delta t + \frac{1}{2}\vec{a_i}(t + \delta t)\delta t \tag{3.15}$$

$$\vec{r_i}(t - \delta t) = \vec{r_i}(t) - \vec{v_i}(t)\delta t + \frac{1}{2}\vec{a_i}(t)(\delta t)^2. \tag{3.16}$$

3.4 Periodic boundary conditions

The number of molecules in a liquid is of the order of the number of Avogadro. However the maximum number of molecules that can be dealt with in a MD simulation with the current multicore computers is of the order of one thousand per core. A system made of a few thousand molecules show important surface effects and the properties of the system do not correspond to those of the bulk. A useful tool to avoid edge effects is the use of *periodic boundary conditions*[6, 7]. In this procedure the system of N molecules is replicated indefinitely along the three dimensions (ot two in a 2D system). The simulation box of size L (a cubic box is the more general choice) is considered the primitive cell of an infinite and periodic network of identical cells, as can be seen Figure 3.2. The molecules of the replicated cells are called *images* of the corresponding molecule in the central box. If in the course of the simulation a molecule of the central box moves in one direction, its replicas also move in the same way. Thus, when a molecule leaves the central box, one of the images will enter the opposite side, thus preserving the number of molecules per box.

In Figure 3.2 a two-dimensional version of a periodic system is shown. Replicated boxes are arbitrarily labeled A, B, C, etc. When particle 7 moves along a border, its images 7A, 7B, etc (where the subscript specifies the replicated box corresponding to each image) move across their contiguous frontiers. In this way, the density of the central box (and hence, that of the whole system) is preserved. During the simulation it is not necessary to store the coordinates of all the images (since it is an infinite number) but simply those of the central box.
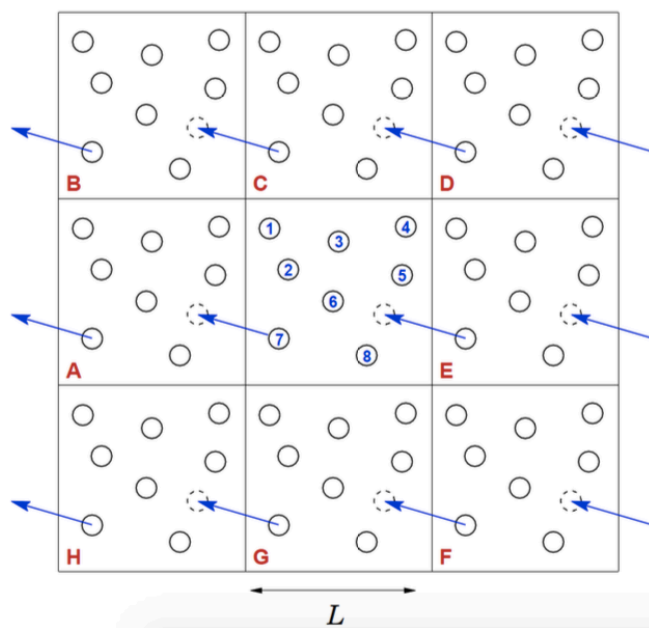
Figure 3.2: **Periodic boundary conditions for a two-dimensional system.**

3.5 Cut-off

The intermolecular energy is zero when all the molecules of the collective are infinitely separated. In any other case, the potential energy is calculated by adding the intermolecular potentials between all the molecules. If periodic boundary conditions are used, the energy of a molecule with the other $N-1$ particles in M replicas is obtained as

$$V\left(1\right) = V\left(1,2\right) + V\left(1,3\right) + ... + V\left(1,N\right) + V\left(1,1_A\right) + V\left(1,1_B\right) + ... + V\left(1,1_M\right) +$$
$$+V\left(1,2_A\right)+V\left(1,2_B\right)+...+V\left(1,2_M\right)+V\left(1,3_A\right)+V\left(1,3_B\right)+...+V\left(1,3_M\right)+...+V\left(1,N_M\right).$$
(3.17)

If infinite replicas would be taken into account, the sum would contain infinite terms. In practice, the evaluation of the interatomic potential requires the use of approximations. As we observe in Figure 3.1, the Lennard-Jones potential is almost negligible beyond $\approx 3\sigma$. In general, for any short ranged potential, it is customary to neglect the interactions between particles separated by a distance larger than a given value, r_c, called the *cutoff* radius.

When using periodic boundary conditions, it is important that any molecule can not interact with any of its replicas nor with more than one replica of another molecule (the opposite could lead to spurious errors). Only the closest neighbours are thus considered for the evaluation of the intermolecular interaction. This procedure is called *minimum image criterion*. Notice (see Figure 3.3) that the nearest image of a given particle needs not to be in the same cell. Therefore, the interactions between pairs of molecules separated by distances greater than the cutoff radius are not evaluated, which defines a truncated potential $V_{trunc}(\mathbf{r_{ij}})$:

$$V_{trunc}(\mathbf{r}_{ij}) = \begin{cases} V(\mathbf{r_{ij}}) & r_{ij} \leq r_c \\ 0 & r_{ij} \geq r_c \end{cases} \tag{3.18}$$

The choice of a certain cut-off distance imposes a minimum value to the unit cell size. For consistency with the minimum image criterion, the cutoff radius r_c must be less than half the edge of the simulation box $L/2$. For instance, if the cutoff distance is 3σ and the size of a molecule is $\sigma \approx 3$ Å then $\mathbf{r}_c \approx 10$ Åand the length of the box must be $L \geq 20$ Å.
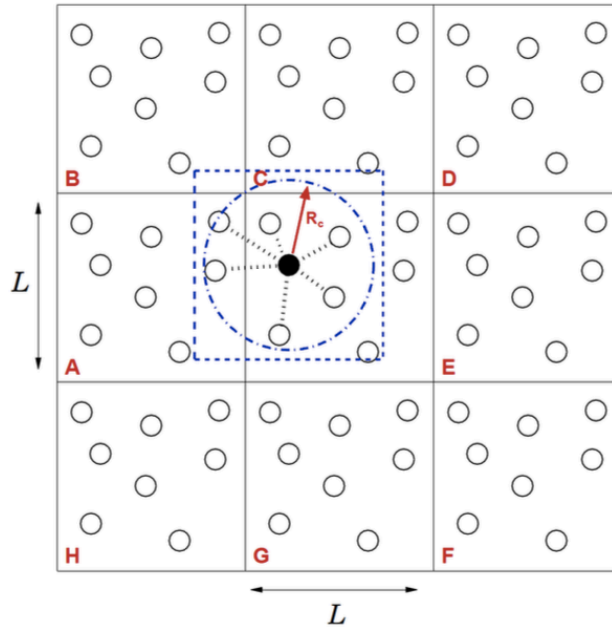
Figure 3.3: **Periodic boundary conditions for a two-dimensional system. Notice that the central particle (filled circle) interacts only with the nearest neighbours (using the minimum image convention) within the cutoff distance (dashed line).**

3.6 Ewald Sum and Particle-Mesh Ewald

The potential of Coulomb, described in the last term of the equation (3.5), is a long-range potential because it decays slowly with the distance between the particles ($V \propto r^{-1}$). Therefore, the truncation of the coulombic forces is not possible. In a system with charged particles, each

charge is usually surrounded by successive layers carrying alternating charge signs. Although there is some cancellation between these layers, the overall series is slowly convergent. The **Ewald Sum**[32] is a technique devised to speed-up the convergence. Here, we give a brief description of its physical basis and concentrate on its implementation[6, 7].

The Ewald sum method manages to transform the Coulomb sum, which converges slowly and conditionally, into two rapidly convergent sums plus a constant term. In this method, each punctual charge q_i, is surrounded by a charge distribution of equal magnitude and opposite sign that extends radially from it. For convenience, this charge distribution is taken as Gaussian:

$$\rho_i\left(\mathbf{r}\right) = \frac{q_i \alpha^3}{\pi^{3/2}} e^{-\alpha^2 r^2}, \tag{3.19}$$

where α is an arbitrary parameter that determines the width of the Gaussian distribution and r is the distance to the particle i where the distribution is centered. This extra charge distribution acts as an ionic cloud that surrounds the punctual charge and shields the interaction between neighbouring charges (see Figure 3.4). The resulting screened interactions are now short-ranged and converge more or less quickly depending on the choice of α. This parameter is then usually chosen so that the interactions beyond the cutoff distance r_c are negligible.

Obviously, each of these shielding distributions have to be compensated by another charge distribution with sign opposite to the previous one, so that the sum of these two distributions recovers the original charges of the system. The intermolecular potential associated to the latter distribution is also a rapidly convergent function when evaluated in the reciprocal (Fourier) space.
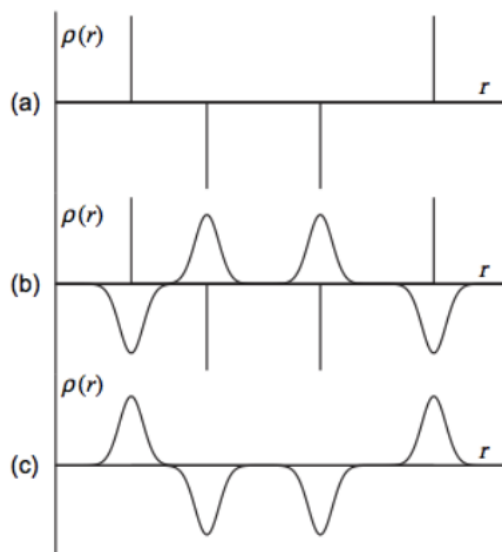
Figure 3.4: **Charge distributions in *E*wald Sum. (a) Original point charges. (b) Punctual charges and shielding distributions. (c) Compensating charge distributions.**

On the other hand, notice that when introducing the Gaussian distributions we have in-

cluded terms of autointeraction. The interaction of each Gaussian with its corresponding punctual charge must be subtracted from the total interaction. In summary, the equation of the *Ewald Sum* is

$$V_C = V_{real} + V_{reciprocal} + V_{self}, \tag{3.20}$$

where

$$V_{real} = \sum_i \sum_{\substack{j \\ j>i}} \frac{q_i q_j}{r_{ij}} \, \text{erfc} \left(\alpha \, r_{ij} \right), \tag{3.21}$$

$$V_{reciprocal} = \sum_{\vec{k}} \frac{4\pi}{L^3} \frac{1}{\left| \vec{k} \right|^2} \exp \left[-\frac{k^2}{4\alpha^2} \right] \left| \rho \left(\vec{k} \right) \right|^2, \tag{3.22}$$

$$V_{self} = -\frac{\alpha}{\sqrt{\pi}} \sum_{i=1}^{N_q} (q_i)^2, \tag{3.23}$$

$erfc(x)$ being the complementary error function

$$\text{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty e^{-t^2} dt, \tag{3.24}$$

which tends to zero for large x. The first term V_{real}, due to the potential of the screened system, is calculated in real space. It is in fact the Coulomb's law damped with the *erfc* function. A typical value of α is 1/3 Å^{-1}. The original Coulomb's law decays slowly but, multiplied by erfc(x), tends to zero quickly. Therefore the resulting interaction may be truncated at $\approx$ 10 Å. The $V_{reciprocal}$ term corresponds to the interactions between the Gaussian charge distributions. In Eq. 3.22, $\vec{k}$ is a vector of the reciprocal space lattice (usually, the sum extends over about 200-1000 reciprocal space vectors), and $\rho \left(\vec{k} \right)$ is the structure factor,

$$\rho \left(\vec{k} \right) = \sum_{i=1}^{N_q} q_i \exp \left[i \vec{k} \vec{r}_i \right], \tag{3.25}$$

where $\vec{r}_i$ defines the location of the i charge in the real space. Finally, the third term V_{self} is the above mentioned self-interaction term which does not depend on the coordinates of the particles of the system. It is a constant that may be evaluated once, at the beginning of the simulation.

The Particle-Mesh Ewald (PME) methods[33] approximate the reciprocal-space term of the standard Ewald summation by a discrete convolution on an interpolating grid, using the discrete Fast-Fourier transforms (FFT). The PME algorithm scales as $N \log(N)$, and is substantially faster than ordinary Ewald summation on medium to large systems. On very small systems it might still be better to use Ewald to avoid the overhead in setting up grids and transforms.

3.7 Long-range corrections

If the intermolecular potential is not rigorously null for greater distances than r_c, truncated interactions will result in a systematic error in V. If intermolecular interactions decay rapidly, energy can be corrected by adding a contribution to V. Let us start from the expression for the average potential energy, $\langle V \rangle$, of the system[7]

$$\langle V \rangle = \frac{N}{2} \int_0^\infty \underbrace{\underbrace{dr\, 4\pi r^2}_{dV} \underbrace{\rho\, g\left(r\right)}_{\rho(r)}}_{\substack{N(r) \\ V\text{of molecules} \\ \text{in r+dr}}} v\left(r\right). \tag{3.26}$$
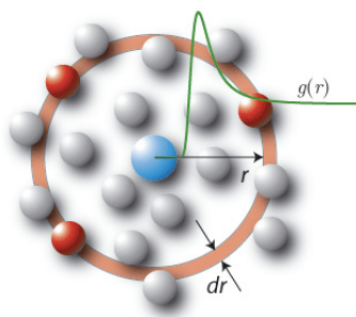
Figure 3.5: **Instantaneous configuration of a set of N identical particles viewed from a reference particle**

Now, by tweaking the potential, we reorganize the integral of the equation (3.26) as follows:

$$\langle V \rangle = \frac{N}{2} \int_0^{r_c} 4\pi r^2 \rho\, g\left(r\right) v\left(r\right) dr + \frac{N}{2} \int_{r_c}^\infty 4\pi r^2 \rho\, g\left(r\right) v\left(r\right) dr. \tag{3.27}$$

The first term is the potential energy calculated by Molecular Dynamics for the truncated potential. The second term is a long-range correction (it is the missed energy from r_c until ∞). The truncation of potential affects very little to the calculation of properties with respect to the real system, excepting the potential energy. To avoid this problem, a long-range correction is usually added assuming that beyond the cutoff the distribution of particles is coincident with the density, $g\left(r\right) = 1$.

3.8 NVE, NpT and NVT ensembles

A molecular dynamics simulation within the NVEensemble, or *microcanonical ensemble*, is defined by a system of N molecules in a constant volume V and a constant total energy E. This type of simulation corresponds to an adiabatic process where no heat exchange is produced.

There, the potential and the kinetic energy vary according to the particles trajectory along the simulation. However the sum of the kinetic and potential energies, that is, the total energy, remains constant.

For the calculation of properties of a system and to be able to compare these properties with the experimental ones, it is sometime convenient to work at a fixed pressure, temperature, or volume. When we impose conditions of constant pressure and temperature the configurations of the system are those of the isothermal-isobaric ensemble (NpT). If we fix the temperature and volume we are working in the isothermal-isochoric ensemble or *canonical ensemble* (NVT). In the canonical ensemble the system can exchange energy with the environment, experimentally assimilated to a sample immersed in a thermal bath. The question is: how do we fix the temperature and/or pressure in our system? In the following we describe two of the most commonly used algorithms to reach our goals: the Nosé-Hoover thermostat[34, 35] and the Parrinello-Rahman barostat[36].

3.8.1 Nosé-Hoover thermostat

It is well known that the temperature is linked to the kinetic energy of the particles (K) as

$$\frac{3}{2}Nk_BT = \langle K \rangle. \tag{3.28}$$

Thus, it is trivial to calculate the temperature of the system from the mean velocities of the molecules in a large number of configurations:

$$\frac{3}{2}k_BT = \frac{1}{2}m_i \langle v^2 \rangle \implies T = \frac{m_i \langle v^2 \rangle}{3k_B}. \tag{3.29}$$

To fix the temperature we need to modify Newton's motion equations, adding a parameter ξ as follows

$$m_i\frac{d^2\vec{r}_i}{dt^2} = -\nabla_i V\left(1, 2, ..., N\right) - \xi m_i\vec{v}_i, \tag{3.30}$$

$$\frac{d\xi}{dt} = \left[\sum_{i=1}^{N}\frac{1}{2}m\vec{v}_i^{\,2} - \frac{3}{2}Nk_BT_f\right]\frac{1}{\tau}, \tag{3.31}$$

where $\xi m_i\vec{v}_i$ is a viscous force, and ξ is the coefficient of this viscous force which can be negative or positive depending on the needs of the system. If $\xi > 0$, then $-\xi\left(m_i\vec{v}_i\right) < 0$, and the added force slows down the particles. On other hand if $\xi < 0$, $-\xi\left(m_i\vec{v}_i\right) > 0$, so the added force accelerates the particles. Thus, the thermostat works differently (accelerating or retarding the particles) depending on the sign of ξ so the temperature fluctuates around the desired temperature T_f. The parameter τ has time units and is called *thermostat relaxation time*. Typically τ is set to around 1 ps.

The extended-ensemble approach was first proposed by Nosé[34] and later modified by Hoover[35]. In the latter formulation, the system Hamiltonian is extended by introducing a thermal reservoir and a friction term ξ (also called heat bath variable). The equations of motion are replaced by

$$\frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i} - \xi\frac{d\vec{r}_i}{dt},$$ (3.32)

The friction parameter is a fully dynamic quantity with its own equation of motion

$$\frac{d\xi}{dt} = \frac{1}{Q}\left(T - T_0\right).$$ (3.33)

The reference temperature is denoted T_0, while T is the instantaneous temperature of the system. The strength of the coupling is determined by the constant Q (usually called the 'mass parameter' of the reservoir) in combination with the reference temperature. The mass parameter is a somewhat awkward way of describing coupling strength, especially due to its dependence on the reference temperature. To maintain the coupling strength, one would have to change Q in proportion to the change in reference temperature. **GROMACS** works instead with the period τ_T of the oscillations of kinetic energy between the system and the reservoir. It is directly related to Q and T_0 via

$$Q = \frac{\tau_T^2 T_0}{4\pi^2}.$$ (3.34)

This provides a much more intuitive way of selecting the Nosé-Hoover coupling strength. In addition, τ_T is independent of system size and reference temperature.

3.8.2 Parrinello-Rahman barostat

Analogous to the thermostat of Nosé-Hoover, in the Parrinello-Rahman barostat[36], a parameter is used to adjust the pressure. To do this we consider that the vectors of the simulation box are represented by a matrix $\mathbf{b}$ that meets the motion equations

$$\frac{d\mathbf{b}^2}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}\left(\mathbf{p} - \mathbf{p}_{ref}\right),$$ (3.35)

where V is the volume of the box, and $\mathbf{W}$ is a matrix of parameters that determines the strength of the coupling. $\mathbf{p}$ and $\mathbf{p}_{ref}$ are the current and reference pressure matrices, respectively. The equations of Newton are modified by adding a new term

$$\frac{d^2\vec{r}_i}{dt^2} = \frac{\vec{F}_i}{m_i} - \mathbf{M}\frac{d\vec{r}_i}{dt},$$ (3.36)

with

$$\mathbf{M} = \mathbf{b}^{-1}\left[\mathbf{b}\frac{d\mathbf{b}'}{dt} + \frac{d\mathbf{b}}{dt}\mathbf{b}'\right]\mathbf{b}'^{-1}.$$ (3.37)

The inverse matrix of the mass parameter $\mathbf{W}^{-1}$ determines the strength of the coupling and how the box can be deformed. Finally, the pressure of the of the i configuration is given by

$$p_i = \frac{Nk_BT}{V} + \frac{1}{3V}\left[\sum_{i=1}^{N} \vec{r}_i \cdot \vec{F}_i\right]. \tag{3.38}$$

The first term corresponds to a system without intermolecular forces. The second term adds the intermolecular forces.

3.9 Constraint algorithms

The use of constraints in MD stems from the need to eliminate the vibrations of high-frequency bonds that require short integration steps. The factor that determines the timestep in the integration of the Newton equation is the highest frequency that occurs in the system. For the O-H bond the stretching frequency is of the order of 10^{14} Hz, corresponding to an average period of the order of 10 fs. Accepting that, in order to correctly reproduce the progress of a periodic function, it must be sampled at least twenty times per period, limits the timestep is limited to less than 0.5 fs. If the molecules do not have torsional degrees of freedom, the stretching motion can be averaged and the forcefield may be calculated assuming that the molecules are *rigid*. The integration of the motion of a rigid molecule may be done by decoupling the motion of the center of mass and the rotation. Such a procedure has an important drawback: it is very difficult to implement a general code to solve the equations for the motion of any arbitrary molecule.

A more general procedure is to assume that the molecules are made of interaction sites. Each site has its own equations of motion so the problem reduces to the motion of a number of punctual sites. In this approach the angular motion is the result of the different velocities of the sites forming the molecule. An obvious problem is that this does not ensure that the shape of the molecule is preserved after the integration. The implementation of constraints solve this question. By introducing a method that constraint the distances between the sites of a rigid molecule, the timestep of the integration may increased up to 2 fs. Since the vibrations of bonds are practically decoupled from the other vibrations of the system, the constraints do not alter the remainder of the dynamics of the system.

The first method proposed for performing a constrained trajectory is the so-called SHAKE, that was introduced in 1977 [37]. The SHAKE algorithm changes a set of unconstrained coordinates $\vec{r}'$ to a set of coordinates $\vec{r}''$ that fulfill a list of distance constraints, using a set $\vec{r}$ SHAKE($\vec{r}' \to \vec{r}''; \vec{r}$). This action is implemented by solving a set of Lagrange multipliers in the constrained equations of motion. Assume that the equations of motion must fulfill K

holonomic constraints, expressed as

$$\sigma_k(\vec{r}_1 \ldots \vec{r}_N) = 0; \quad k = 1 \ldots K, \tag{3.39}$$

(e.g. $(\vec{r}_1 - \vec{r}_2)^2 - b^2 = 0$). Then the forces are defined as

$$-\frac{\partial}{\partial \vec{r}_i} \left(V + \sum_{k=1}^{K} \lambda_k \sigma_k \right), \tag{3.40}$$

where λ_k are Lagrange multipliers which must be solved to fulfill the constraint equations. The second part of this sum determines the *constraint forces* $\vec{G}_i$, defined by

$$\vec{G}_i = -\sum_{k=1}^{K} \lambda_k \frac{\partial \sigma_k}{\partial \vec{r}_i} \tag{3.41}$$

The displacement due to the constraint forces in the leapfrog or Verlet algorithm is equal to $(\vec{G}_i/m_i)(\Delta t)^2$. Solving the Lagrange multipliers (and hence the displacements) requires the solution of a set of coupled equations of second degree. These are solved iteratively by SHAKE until all the constraints are satisfied within a *relative tolerance* `TOL`.

3.10   Gromacs

GROMACS (GROningen MAchine for Chemical Simulations)[38] is an engine to perform molecular dynamics simulations and energy minimization, developed in 1991 at Department of Biophysical Chemistry, University of Groningen, Netherlands. It is a versatile package to perform Molecular Dynamics, i.e. simulate the Newtonian equations of motion for systems with hundreds to millions of particles.

GROMACS is primarily designed for biochemical molecules like proteins, lipids and nucleic acids that have a lot of complicated bonded interactions, but since it is extremely fast at calculating the nonbonded interactions (that usually dominate simulations) many groups are also using it for research on non-biological systems, e.g. polymers and, like us, water solutions.

GROMACS provides extremely high performance compared to all other programs. A lot of algorithmic optimizations have been introduced in the code. In **GROMACS** 4.6 and up, on almost all common computing platforms, the innermost loops are written in C using intrinsic functions that the compiler transforms to SIMD machine instructions, to utilize the available instruction-level parallelism. This program is Free Software, available under the GNU Lesser General Public License (LGPL), there is no scripting language - all programs use a simple interface with command line options for input and output files.

This program requires three input for the simulation of a model files, which are as follows:

1. the *topol.top* file where there are the parameters of the *Force Field*.

2. the configuration of the system *conf.g96* where it can be found the initial velocities $\vec{v}_i(0)$, positions $\vec{r}_i(0)$ of the atoms, and size of the simulation box.

3. the *grompp.mdp* file where it can be found the time and the condition (pressure and temperature) of the system.

When the simulation ends, the results can be analyzed using a **GROMACS** command called *g_energy*. With this command you can get the average values of many properties of the system, including density, kinetic and potential energies, etc.

Chapter 4

Methodology

4.1   Simplified seawater composition for numerical simulations

Having a look at Table 2.1 it is easy to realize that the number of molecules needed to reproduce the Reference Composition Seawater is exceedingly large to be implemented in molecular simulations. The total number of ions in the definition of the Reference Composition amounts to 10^7. The number of water molecules for a Reference Seawater of salinity around 35 g/kg would then be of the order of 500 millions. The huge size of such system would pose a number of problems to perform molecular simulations. Not only as a result of the computing cost of the simulations but also because of the difficulties of implementing a working starting configuration. More importantly perhaps is the fact that the calculated properties are not sensitive to extremely fine composition details. It is not necessary to reproduce exactly the Reference Composition Seawater to arrive at significant conclusions about the ability of the current force-fields to account for the seawater properties. It seems thus interesting to design a simplified system with a composition close to the Reference Composition Seawater but more amenable to computer simulation.

The first assumption we have made is to ignore *minor constituents*, i.e. ions with a mole fraction lower than a given value. In particular, since the mole fraction, x, of HCO_3^- is already very small ($x = 0.0015$) we may neglect it as well as the rest of components with lower concentration. Thus, our model for seawater would only deal in principle with Cl^-, Na^+, Mg^{2+}, SO_4^{2-}, Ca^{2+}, K^+. However, most of the neglected chemical compounds are ions carrying a negative charge so it is not possible to ignore them completely in order to fulfill the electroneutrality condition. It is thus required to include a monovalent ion representing the ensemble of these minor constituents. A system containing just one ion representing the minor constituents would consist of about 300-350 ions and $15000 - 20000$ water molecules. With these premises in mind, it would seem trivial the design of an acceptable model for seawater. However, consid-

erable tweaking is required because the number of ions must be integer numbers. The relative abundance of all the ions can not be easily reproduced with small integer numbers. Besides, the rounding of numbers leads quite often to deviations of electroneutrality. After a cumbersome trial and error we have finally arrived at a composition fulfilling the requirements.

The resulting composition for our "in silico" seawater of salinity $S_R = 35.16$ g/kg, is shown in Table 4.1. It matches the Reference Composition Seawater to a great accuracy: the mole fractions of our model are correct approximately up to the fourth decimal figure. In fact, such good agreement is somewhat unexpected given the relatively small number of ions in our sample. Notice that, since a single ion represents all the minor constituents, the properties of this single ion are not well defined.

Let us recall that we are interested in properties which derive from the interaction between particles. Since the principal contribution to the potential energy in a ionic solution is the coulombic term, the main feature of the single anion representing the minor constituents is that it must ensure the charge neutrality. Small variations in the dispersion interactions of this anion with the rest of particles have a negligible impact in the properties of the system. In fact, the most significant effect of the dispersion forces in a ionic solution is the excluded volume determined by the van der Waals radius. Any reasonable choice for the van der Waals radius of the single anion would suffice for our purposes. For simplicity, we have chosen to use for this anion the same potential parameters as those of Cl^-. In summary, the resulting composition for our "in silico" seawater of salinity $S_R = 35.16$ g/kg is shown in the fifth column of Table 4.1 and contains 15210 water molecules and 318 ions (Cl^-, Na^+, Mg^{2+}, SO_4^{2-}, Ca^{2+}, K^+). From the composition of this sample, intended to simulate the KCl-normalized Reference Seawater, it is straightforward to obtain the Reference Composition seawater at any desired salinity by adding or removing water molecules.

One of the aims of this work is to investigate the contribution of the different ions to the properties of seawater. More specifically, we intend to shed light on the effect of a further simplification of the sample. Following the same line of reasoning of the previous paragraph it seems interesting to analyse whether the replacement of Ca^{2+} and K^+ by Mg^{2+} and Na^+, respectively, has a noticeable effect on the solution properties. We may take the argument to the extreme: it makes sense to replace seawater by an equivalent NaCl solution? Because of the large set of experimental data available for both systems, there is no need to carry out molecular simulations to answer the question: there is no such equivalence for a wide set of properties. It is easy to realize that a rescaling of salinities by a factor is able to provide an excellent match of the densities of both systems. However the factor is slightly dependent on temperature. But a rescaling trying to match the corresponding experimental values for a dynamical property

such as the viscosity does not work at all (even using a different factor than that employed for the density).

Despite the difficulty of defining an equivalence between seawater and NaCl solutions, the study of the latter system may be of interest. Although other ions make significant contributions to the properties of seawater, the largest ionic contributions come from the Na^+ and Cl^- interactions. In this work, we have carried out a parallel study of the NaCl solution at "similar salinities" as those investigated for seawater. Since we will compare each solution with its corresponding experimental data, there is no room for ambiguity.

4.2 Force Field

4.2.1 Water model

To represent the water potential we have chosen a rigid non-polarizable simple model because it provides an excellent compromise between performance and simplicity. Among these type of water models, **TIP4P/2005**[9] has shown to give the best performance for a large number of properties[39, 40]. Figure 4.1 displays the essential features of TIP4P/2005.
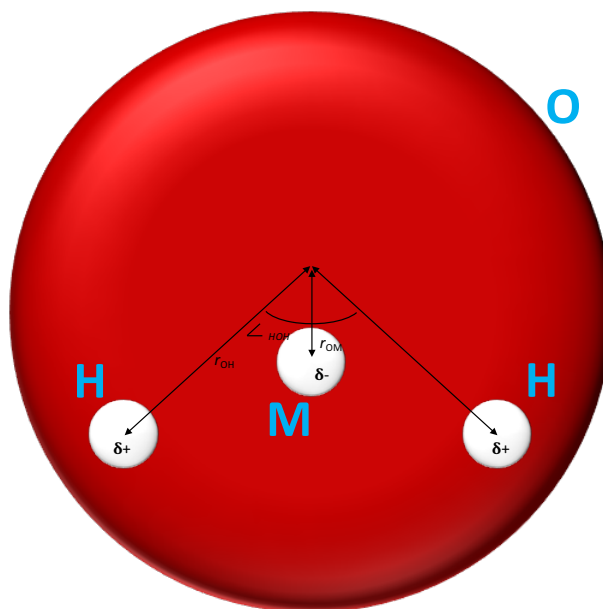
Figure 4.1: **TIP4P/2005 Water model**

It has a Lennard-Jones site centered at the oxygen atom, and three charges located in the following way: two positive charges on each hydrogen, q_H, and a negative one located at the point M, $-2q_H = q_M$ (M is placed along the bisector of the H-O-H angle). The following table gives the parameters of the model.

Table 4.1: "In silico" models for the molecular simulation of seawater. The second column shows the values of the Reference Composition (for the components with $x > 0.002$). The third column gives the amount of substances for the KCl-normalized Salinity $S_R = 35.16504\,g/kg$. Fourth column represents the number of water molecules and ions for a KCl-normalized solution closely matching the Reference Composition using a single anion to represent the minor constituents. Sixth column shows our proposal for a "In silico" seawater (ISsw) model (it is identical to the third column where the single anion is represented by a chloride). The simplified "In silico" seawater (sISsw) model is obtained by replacing Ca^{2+} and K$^+$ ions by Mg^{2+} and Na$^+$, respectively.

Component	mole fraction	KCl norm	N molec	mole fraction	N molec (ISsw)	N molec (sISsw)
Water (H$_2$O)	—	964.8350	15210	—	15210	15210
Chloride (Cl$^-$)	0.48748	19.35271	155	0.487	156	156
Sodium (Na$^+$)	0.41881	10.78145	133	0.418	133	136
Sulfate (SO$_4^{2-}$)	0.02522	2.71235	8	0.025	8	8
Magnesium (Mg^{2+})	0.04717	1.28372	15	0.047	15	18
Calcium (Ca^{2+})	0.00918	0.41208	3	0.009	3	—
Potassium (K$^+$)	0.00912	0.39910	3	0.009	3	—
Minor constituents	0.00302	0.22363	1	0.003	—	—
Sum Ion			318		318	318
Salinity		35.16504	35.16			

	r_{OH} [nm]	H-O-H [°]	σ [nm]	ϵ[kJ/mol]	q_H [e]	r_{OM} [nm]
TIP4P/2005	0.09572	104.52	0.31589	0.7749	0.5564	0.01546

Table 4.2: Parameters of the TIP4P/2005 water model. The distance between the oxygen and hydrogen is denoted as r_{OH}. The angle formed by hydrogen, oxygen and the second hydrogen is denoted as H-O-H. The LJ centre is located at the oxygen with parameters σ and ϵ/k_B [K]. The assigned charge of the hydrogen atom is q_H. r_{OM} it is the distance from the oxygen to the site M placed along the bisector H-O-H.

4.2.2 Ionic Interactions

Two different forcefields have been used in this work. Both sets differ in the parameters of Na^+ and Cl^-. For the first set we have chosen the Joung-Cheatham (JC)[11] potential while the second one use the OPLS (optimized potentials for liquid simulations)[10] forcefield to represent the sodium-chloride interactions. The same parameters are used in both sets for the rest of ions. The OPLS forcefield also provides parameters for the divalent cations and these are the obvious choice for the second forcefield which will be denoted as OPLS. But these will also be used in the other set (it will be referred to as JC despite being actually a merge between JC and OPLS). Finally, for the sulfate anion, the parameters given by Cannon et al.[41] will be incorporated to both the JC and the OPLS sets. The following tables show the parameters of the main ions.

Table 4.3:

Name	Mass	Charge	JC		OPLS	
			σ[nm]	ϵ [kJ/mol]	σ[nm]	ϵ [kJ/mol]
Na^+	22.98977	1	2.1595E-01	1.4764	3.33045E-01	1.15980E-02
Cl^-	35.453	-1	4.8305E-01	0.535E-01	4.41724E-01	4.92833E-01
Mg^{2+}	24.305	2	1.64447E-01	3.66118E+00	1.64447E-01	3.66118E+00
O_{SO_4}	24.01565	-1	3.15000E-01	0.83740E+00	3.15000E-01	0.83740E+00
S_{SO_4}	0	2	3.55000E-01	1.04670E+00	3.55000E-01	1.04670E+00

4.2.3 Cross Interactions

Cross interactions refer to interactions between different atoms such as ion-ion, ion-water or water-water. The cross interactions have been calculated through the well-known Lorentz-

Berthelot combining rules

$$\epsilon_{ij} = (\epsilon_i \cdot \epsilon_j)^{\frac{1}{2}} \tag{4.1}$$

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \tag{4.2}$$

4.2.4 Sulfate Anion

Unlike the other ions, the sulfate, is a polyatomic molecule and, as for water, it may be considered a rigid molecule. Since **GROMACS** is based on *site-site* interactions a constraint method (we use SHAKE, see chapter 3.9) is required to preserve the molecular geometry. But the constraint algorithm can not work for a molecule with tetrahedral symmetry because the motion of the outer atoms determine that of the central ion. However the site-site interactions can not ensure this requirement and, thus, the constraint algorithm fails.

This failure forced us to use virtual atoms; these are interacting atoms, but without their own mass. The force exerted on a virtual atom is projected to the rest of sites. After evaluating the new positions of the massive sites, the position of the virtual sites can be obtained. Two problems arise from the use of virtual sites: the preservation of a) the total mass of the molecule, and b) the moments of inertia. In order to get rid of the first problem the mass of the sulphur was distributed among the oxygen atoms (see the table 4.3. However this modify the moments of inertia of the molecule. In order to keep them constant, the massive interaction sites should be moved towards the center of the molecule. In this way we would have five virtual interacting centers (at the positions of the four oxygens and the sulfur), and four mass centers for a total of nine interaction centers. This would complicate the calculations very much and, probably, would not affect the results of the simulations as the sulfate is the minority component of the solution. Thus, we have kept the positions of the O and S atoms at their original position.

4.3 Evaluation of properties

4.3.1 Temperature

The absolute temperature T is related to the kinetic energy of the N-particle system

$$E_{kin} = \frac{1}{2} \sum_{i=1}^{N} m_i v_i^2, \tag{4.3}$$

through

$$\frac{1}{2} N_{df} kT = E_{kin}, \tag{4.4}$$

where k is Boltzmann's constant and N_{df} is the number of degrees of freedom which can be computed from:

$$N_{df} = N - N_c - N_{com}. \tag{4.5}$$

Here N_c is the number of constraints imposed on the system. When performing molecular dynamics $N_{com} = 3$ additional degrees of freedom must be removed, because the three center-of-mass velocities are constants of the motion, which are usually set to zero. When simulating in vacuo, the rotation around the center of mass can also be removed, in this case $N_{com} = 6$. When more than one temperature-coupling group is used, the number of degrees of freedom for group i is

$$N_{df}^i = (3N^i - N_c^i)\frac{3N - N_c - N_{com}}{3N - N_c} \tag{4.6}$$

The kinetic energy can also be written as a tensor, which is necessary for pressure calculation in a triclinic system, or systems where shear forces are imposed

$$\mathbf{E}_{kin} = \frac{1}{2}\sum_{i=1}^{N} m_i \vec{v_i} \otimes \vec{v_i} \tag{4.7}$$

4.3.2 Pressure and virial

The stress (or pressure) tensor $\mathbf{P}$ is calculated as the difference between kinetic energy E_{kin} and the virial Ξ

$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{kin} - \Xi), \tag{4.8}$$

The virial Ξ tensor is defined as

$$\Xi = -\frac{1}{2}\sum_{i<j} \vec{r_{ij}} \otimes \vec{F_{ij}}, \tag{4.9}$$

where V is the volume of the computational box. The scalar pressure p, which can be used for pressure coupling in the case of isotropic systems, is related to the trace of the stress tensor

$$P = trace(\mathbf{P})/3 \tag{4.10}$$

4.3.3 Density

The simulations for the calculation of the density were carried out in the isothermal-isobaric ensemble NpT. In this ensemble, the volume of the box is not constant, but fluctuates around an average value, so the density is simply calculated as:

$$\rho = \frac{M}{\langle V \rangle} \tag{4.11}$$

where $\langle V \rangle$ is the ensemble averaged box volume and M is the total mass of the sample.

4.3.4 Self diffusion coefficient

The self diffusion coefficients (D) are obtained through the Stokes-Einstein equation[42]

$$D = \frac{1}{6} \frac{\left\langle [\vec{r}_i(t) - \vec{r}_i(0)]^2 \right\rangle}{t} = \frac{MSD}{6t}, \tag{4.12}$$

where $\vec{r}_i(0)$ is the particle's position i in a certain origin of time t, and $\vec{r}_i(t)$ is the position of this particle at time t. $[\vec{r}_i(t) - \vec{r}_i(0)]^2$ is just the square of the distance that the particle has moved. Making an average over all the particles of the system and time origins we get $\left\langle [\vec{r}_i(t) - \vec{r}_i(0)]^2 \right\rangle$, usually called *Mean Square Displacement, MSD*. Large MSD indicate large self diffusion coefficients.

From equation 4.12 we get $MSD = 6Dt$, so if we represent the mean square displacement as a function of t, the slope of the curve provides the self diffusion coefficient: $D = \frac{1}{6} \cdot m$.

4.3.5 Shear viscosity

The shear viscosity has been calculated using the Green-Kubo equation[43]

$$\eta = \frac{V}{k_b T} \int_0^\infty \left\langle P_{\alpha\beta}(t_0) P_{\alpha\beta}(t_0 + t) \right\rangle dt, \tag{4.13}$$

where $\left\langle P_{\alpha\beta}(t_0) P_{\alpha\beta}(t_0 + t) \right\rangle$ it is the autocorrelation function of the off-diagonal components of the stress tensor. In isotropic systems, the stress tensor is symmetric and the components out of the diagonal, P_{xy}, P_{xz}, and P_{zy}, are equivalent. Besides, due to the rotational invariance of the molecule, the terms $(P_{xx} - P_{yy})/2$ y $(P_{yy} - P_{zz})/2$ are also equivalent[44, 45]. Thus, in our numerical integration calculations, the resulting autocorrelation function was calculated as an average of five equivalent components of the stress tensor.

To evaluate the shear viscosity[45, 46] we need to perform simulations in the NVT ensemble. Thus, we must know the volume of the system and keep it constant in the simulation. The calculation of the volume requires a previous simulation in the NpT ensemble from which we get the volume of the system at the desired temperature and pressure.

4.3.6 Radial distribution functions

The *radial distribution function* (rdf) or pair correlation function $g_{AB}(r)$ between particles of type A and B is defined in the following way

$$g_{AB}(r) = \frac{\langle \rho_B(r) \rangle}{\rho_B} = \frac{1}{\rho_B} \frac{1}{N_A} \sum_{i \in A}^{N_B} \sum_{j \in B}^{N_B} \frac{\delta(r_{ij} - r)}{4\pi r^2}, \tag{4.14}$$

$\langle \rho_B(r) \rangle$ being the particle density of type B at a distance r from particles A, and ρ_B the average particle density of type B (see Figure 4.2b). Essentially, the radial distribution function

represents the probability of finding a pair of particles i and j at the interparticle distance r_{ij} in the case of an isotropic fluid.

In practice, the analysis program `gmx rdf` divides system into spherical slices (from r to $r + dr$, see Figure 4.2 (a)) and replaces the δ-function by an histogram.
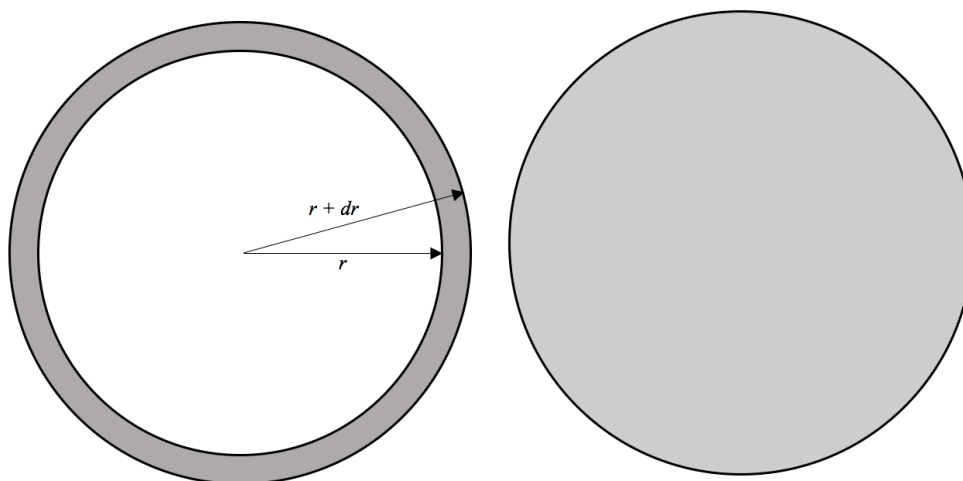
Figure 4.2: **(a) $g_{AB}(r)$. The slices are colored gray. (b) Normalization $\langle \rho_B \rangle_{local}$. Normalization volumes are colored gray.**

An example of a rdf is given in Fig. 4.3 Note that $g(r)$ presents two clear local maximum, the first corresponds to $r \approx 0.25$ nm which is the first coordination sphere. At larger distances, the function oscillates around 1. The rdf must become essentially flat at distances smaller than half the box length indicating that there is no correlation between a particle and any of its replicas.
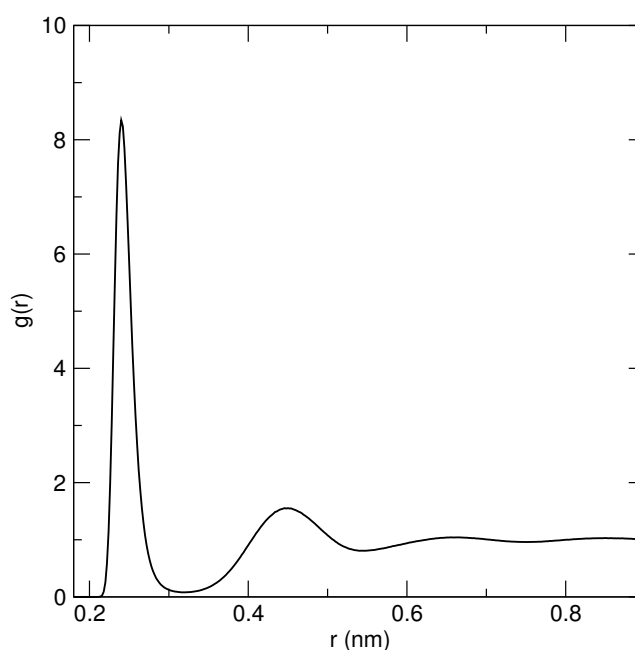
Figure 4.3: **Typical shape of a radial distribution function.**

4.3.7   Ions pairs and hydration numbers

The rdf allows the calculation of the number of ion pairs and hydration (see Figure 4.4).
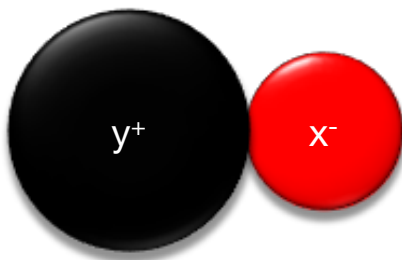
Figure 4.4: **Ion pairs formed by an ion x^+ and the other y^-**

If we consider an instant configuration of our system of particles and select one of them randomly, for example an ion x^+, the number of anions y^- included among r and $r + dr$ around x^+ is given by

$$N(r) = \rho_y \; 4 \; \pi \; r^2 \; g_{x^+-y^-}(r) \; dr. \tag{4.15}$$

If we integrate the above equation between 0 and r_{min} (the position of the first minimum in $g_{x^+-y^-}(r)$)

$$n\left(x^+\text{-}y^-\right) = \int_0^{r_{min}} N(r)dr = \rho_y \cdot 4\pi \int_0^{r_{min}} r^2 \, g_{x^+-y^-}(r) \; dr, \tag{4.16}$$

we get $n\left(x^+\text{-}y^-\right)$, the number of ion pairs. ρ_y is the number density of the ion y^-

$$\rho_y = \left(\frac{N_y}{V}\right) \tag{4.17}$$

The hydration numbers, i.e., the number of water molecules surrounding a given ion can be readily obtained in the same way from the ion-oxygen and/or ion-hydrogen rdf's.

4.4   Errors

A fundamental part of a scientific work is the calculation of the uncertainties or, more precisely, their estimation. **GROMACS** includes an error analysis package using block averages. The simulation is divided in blocks and, for each block, the average value for a given property is calculated. The error of the total average is calculated from the variance between the averages of the m blocks B_i in the following way:

$$\epsilon_B^2 = \frac{\sum_i \left(B_i - \langle B \rangle\right)^2}{m \cdot (m-1)} \tag{4.18}$$

These errors are dependent on the block size. An analytical curve of the block's uncertainties may be obtained by fitting them to a sum of two exponentials

$$\epsilon^2(t_{bl}) = \frac{2\sigma^2}{t}\left(\alpha\tau_1\left(1 + \frac{\tau_1}{t_{bl}}\left(e^{-\frac{t_{bl}}{\tau_1}} - 1\right)\right) + (1-\alpha)\cdot\tau_2\left(1 + \frac{\tau_2}{t_{bl}}\left(e^{-\frac{t_{bl}}{\tau_2}} - 1\right)\right)\right), \tag{4.19}$$

where t is the total time of the simulation, σ is the standard deviation of the data set. α(usually is close to 1), τ_1 and τ_2 are obtained by a least squares fit. Observing the Figure 4.5 we can divide it into two areas. A first one correspond to short blocks where there is a significant correlation between them. The second region correspond to large blocks (and thus there are few points) with almost no correlation between them. Since the number of blocks is reduced, there is a substantial noise in this region. However we may see that the analytical fit is asymptotic and the plateau provides a representative estimation of the uncertainty.

$$\lim_{t_{bl} \to \infty} \epsilon^2(t_{bl}) = \sigma \sqrt{\frac{2 \left(\alpha \tau_1 + (1 - \alpha) \tau_2 \right)}{T}} \tag{4.20}$$
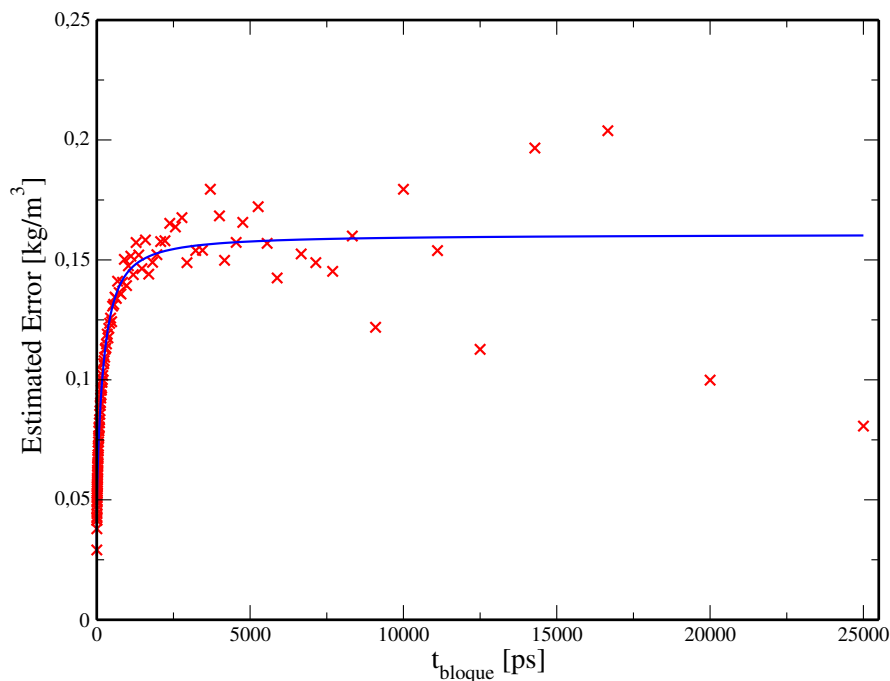
Figure 4.5: **Density errors estimation for a 5 m NaCl system at 1 bar and 298 K. The 'block size' time is on the abscissa axis. In the ordinates there is the estimated error according to the size of the block. Finally, the curve represents an adjustment of the block errors and corresponds to the equation (4.19).**

Chapter 5

Results

In this chapter we present the results of the simulations performed for the simplified seawater compositions discussed in 4.1. The simplest one contains water and the ions Na^+, Cl^-, Mg^{2+} and SO_4^{2-} while the more complex seawater sample used in this work is made of the above mentioned components plus the ions Ca^{2+} and K^+. We have investigated different properties as a function of the temperature and salinity. In particular we have evaluated the equation of state, dynamical properties (viscosity and self-diffusion coefficients), and the structure of the solution (ion-water and ion-ion distribution functions). Since the most abundant ions of seawater are Na^+ and Cl^- we have also carried out an study of the behaviour of NaCl solutions using the same interaction potentials as for seawater.

5.1   Density

As already mentioned, one of the aims of this work was to investigate how much we can simplify the composition of seawater including more or less abundant ions. Figure 5.1 shows the variation of the density of seawater with temperature at the Reference-Composition salinity (S=35.16 g/kg) using the JC and OPLS interaction potentials. Here, we have represented the results for two simplified compositions: the first one contains the ions Na^+, Cl^-, Mg^{2+} and SO_4^{2-} (see column labelled sISsw in Table 4.1) while the other one also includes Ca^{2+} and K^+ (labelled as ISsw). Notice that the effect of replacing Ca^{2+} and K^+ by Mg^{2+} and Cl^-, respectively, is almost negligible. In fact, only a few points were calculated for the ISsw system since their results were essentially coincident with those for the simplified composition at the same conditions. The numerical results, presented in Table 5.1 indicate that the differences are close to the uncertainty of the calculations (around 0.15 kg/m^3).
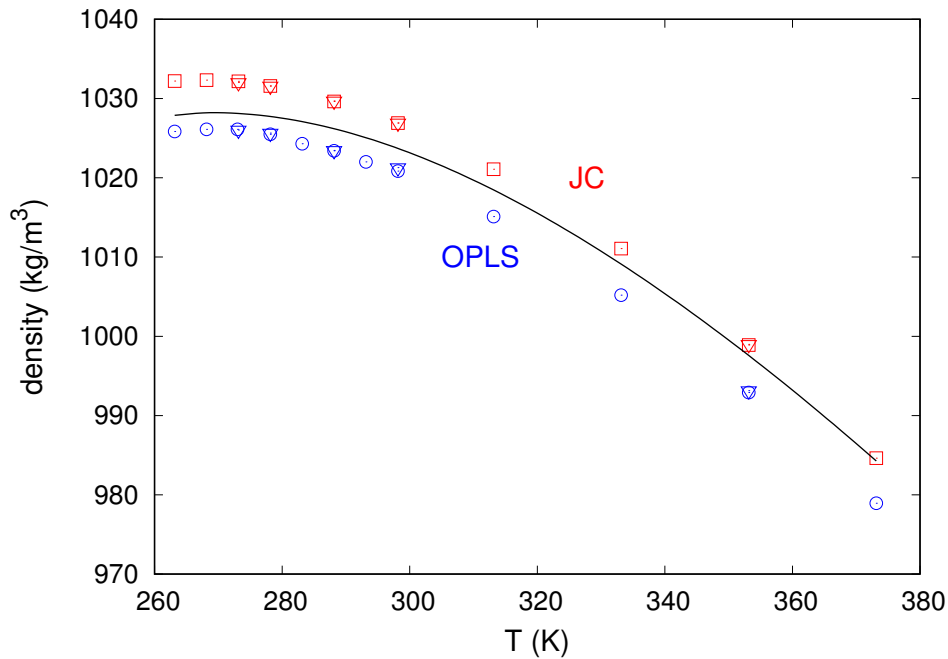
Figure 5.1: **Density at as a function of temperature compared to the experimental data (full line) at the Reference-Composition salinity (35.16 g/kg). The squares and circles represent the simulation values for the composition containing only Na^+, Cl^-, Mg^{2+} and SO_4^{2-} ions (see column labelled sISsw in Table 4.1); the triangles represent the results when the simulated seawater also contains Ca^{2+} and K^+ (ISsw).**

Table 5.1: Comparison of the numerical results for the densities of our simplified 'In silico' seawater compositions at 1 bar and $S = 35.16$ g/kg.

t(°C)	OPLS		JC	
	sISsw	ISsw	sISsw	ISsw
-10	1025.84	-	1032.20	-
-5	1026.11	-	1032.32	-
0	1026.07	1026.03	1032.17	1032.02
5	1025.50	1025.65	1031.59	1031.58
10	1024.29	-	-	-
15	1023.41	1023.45	1029.63	1029.70
20	1022.01	-	-	-
25	1020.85	1021.29	1026.91	1026.93
40	1015.10	-	1021.09	-
60	1005.19	-	1011.06	-
80	992.91	993.14	998.90	998.99
100	978.94	-	984.62	-

As expected, for both potentials, the density of the solution decreases as the temperature increases. The JC forcefield tends to overestimate the density while the OPLS to underestimate it. The OPLS model reflects better the experimental data at low temperatures compared to the JC, while the JC forcefield produces a better agreement with the experimental measurements at high temperatures.

Figure 5.2 shows the density as a function of salinity for our simplified 'In silico' seawater composition, sISsw, (see Table 4.1) at ambient pressure and the temperatures 298.15 K and 373.15 K. At low salinities the simulation results essentially match the experimental data. This could be expected taking into account the excellent performance of the TIP4P/2005 water model[9, 39, 40]. As the salinity increases, so does the departures between simulation and experiment albeit the differences are never important. As in our previous figure, the JC forcefield overestimates the density while the OPLS underestimate it, so the simulation results bracket the experimental data. The JC predictions are slightly better at 373.15 K while the OPLS results deteriorate a little bit at increasing temperatures. Overall, the JC forcefield provides a quite acceptable description of the density of seawater and its dependence with salinity along two isotherms. The performance of the OPLS forcefield is slightly worse.
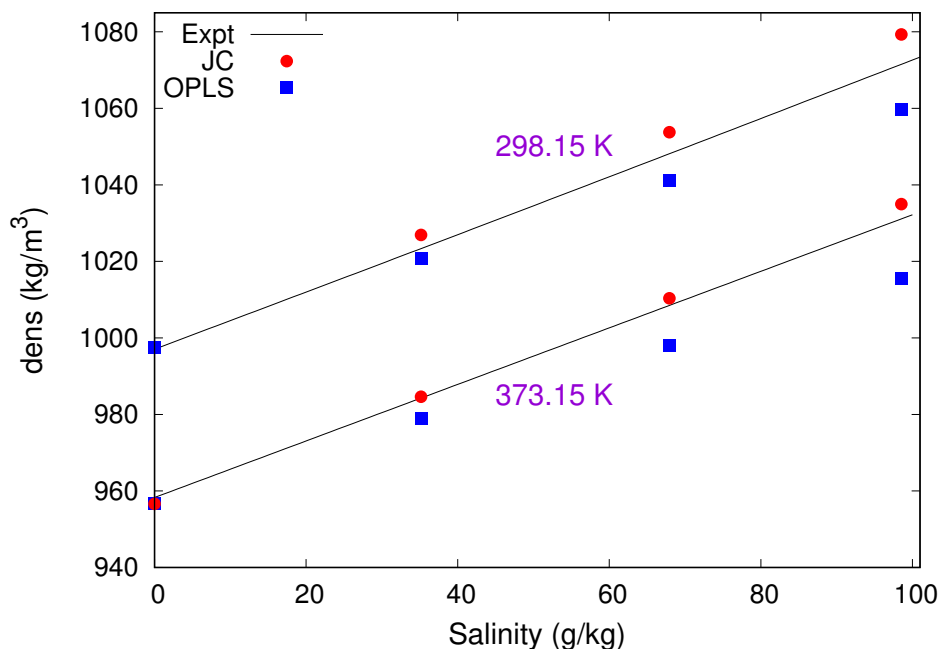
Figure 5.2: **Density as a function of salinity for two isotherms (as obtained using the JC and OPLS potentials) compared to the experimental values for seawater**

We have demonstrated that the effect of replacing Ca^{2+} and K^+ by Mg^{2+} and Cl^-, respectively, is almost negligible. We could try to take further this replacement scheme and propose an extremely simple NaCl solution representative of seawater at a given salinity. However,

the replacement of the divalent ions magnesium and sulfate by simple monovalent ions is not trivial (among other problems we must face the question of preserving the electroneutrality of the solution because the abundance of the sulfate anion is different from that of the magnesium cation). In fact, we have tried to map the *experimental* seawater properties into the *experimental* NaCl behaviour and we have found that there is not such simple mapping. Despite this failure, it seems interesting to assess the importance of correctly describing the interactions of Na^+ and Cl^-. To this end, we may examine the *simulated* properties of a NaCl solution (more or less close to seawater of a given salinity) and check it against the *experimental* properties of the NaCl solution. It remains the question of how to define the NaCl concentration with properties more or less close to those of seawater at a given salinity. We have observed that the experimental data for a NaCl solution are noticeably different from those of seawater at the same ionic strength. We have found that the more simple approach which produces similar results for both solutions consist in replacing every divalent ion by two monovalent ions. Notice that this approach preserves the electroneutrality condition. In summary, our interest is to check the simulation results against experimental data for NaCl solutions in a concentration range producing similar properties to those of seawater.

As for seawater, the density of the NaCl solutions has been calculated with the JC and OPLS forcefields. In Figure 5.3, we have depicted the simulation results for a 0.628 molal NaCl solution as a function of the temperature at constant pressure (1 bar) and compare them with the experimental data for NaCl. The similarity of this figure to Fig. 5.1 is evident which indicates that the NaCl 0.628 m solution may be considered as a simple proxy for the Reference-Composition standard seawater. Again, the simulation results for both models bracket the experimental ones. Also in concordance with previous results for seawater, the experimental data at low temperature for NaCl solutions are better represented by the potential OPLS while JC performs better at high temperatures. The similarities between Figs. 5.3 and 5.1 indicate that, as expected, most of the departures observed between simulation and experimental data in seawater were due to inaccuracies in the Na^+ and Cl^- potential interactions. This statement is corroborated by the results presented in Fig. 5.4 showing a remarkable parallelism with Fig. 5.2.
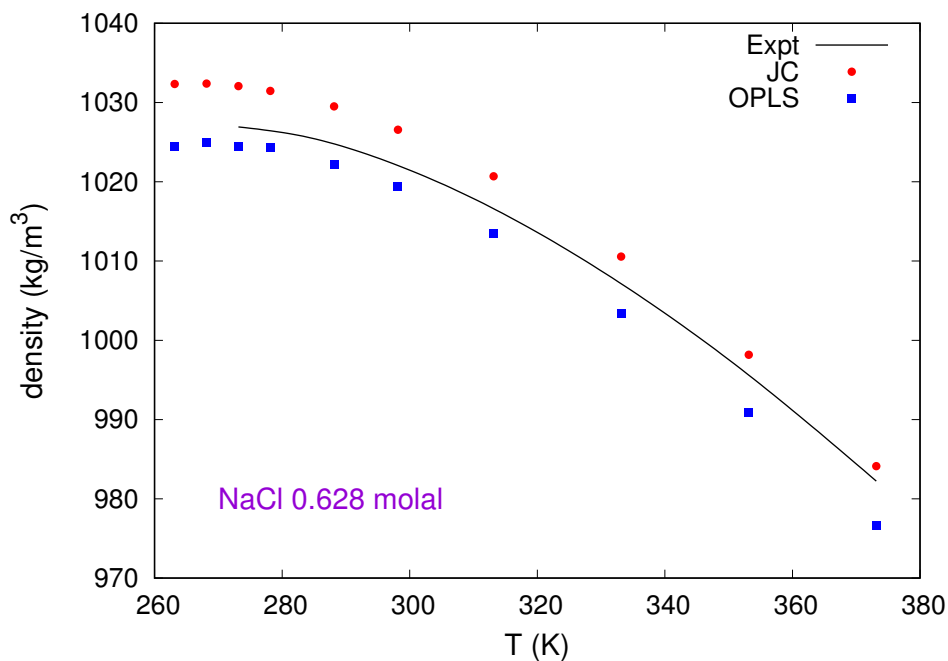
Figure 5.3: **Density values as a function of temperature for a 0.628 molal NaCl solution using the JC and OPLS forcefields compared to experimental data.**
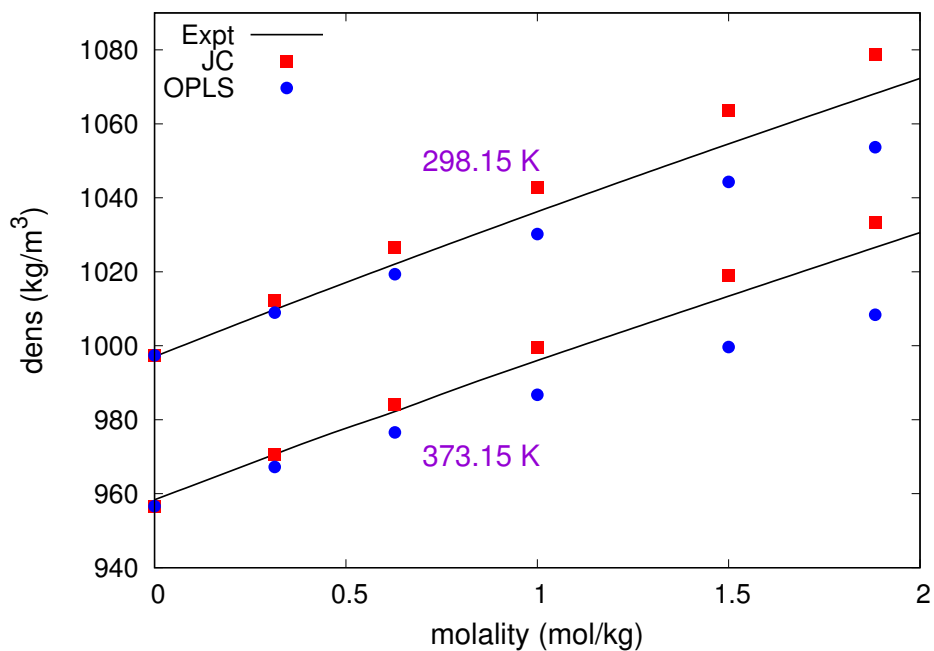
Figure 5.4: **Density as a function of concentration for NaCl aqueous solutions for the 298.15 K and 373.15 K isotherms at 1 bar.**

5.2 Dynamical properties

5.2.1 Viscosity

In this section we present the calculations for the shear viscosity obtained with the Green-Kubo formula[43] for the OPLS and JC forcefields. Figure 5.5 shows the shear viscosity of seawater as a function of temperature at the Reference-Composition salinity (S=35.16 g/kg) using the JC and OPLS interaction potentials. As in Fig. 5.1 we have represented the results for two simplified compositions: the first one contains the ions Na^+, Cl^-, Mg^{2+} and SO_4^{2-} (see column labelled sISsw in Table 4.1) while the other one also includes Ca^{2+} and K^+ (labelled as ISsw). The differences between the results for the ISsw and sISsw compositions are again quite small. The influence of Ca^{2+} and K^+ is visible only at the lowest temperature. In fact, the calculations for the this solution were carried out only up to of 298.15 K since their effect at this temperature seems already within the uncertainty of the results (around 5%).
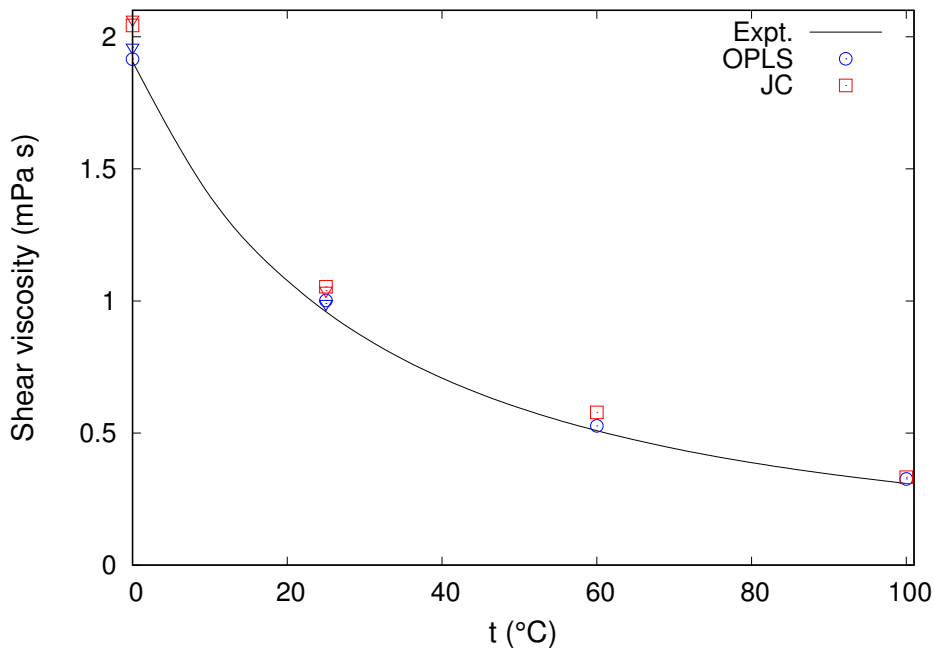
Figure 5.5: **Viscosity of seawater as a function of temperature compared to the experimental data (full line) at the Reference-Composition salinity (35.16 g/kg). The squares and circles represent the simulation values for the composition containing only Na^+, Cl^-, Mg^{2+} and SO_4^{2-} ions (see column labelled sISsw in Table 4.1); the triangles represent the results when the simulated seawater also contains Ca^{2+} and K^+ (ISsw).**

In Figure 5.7 we show the results at 1 bar and two temperatures, 298.15 K and 373.15 K. Both potential interactions account for the dependence on salinity of the shear viscosity of seawater. The agreement with experiment is almost quantitative in the low salinity and high temperature regime. However, especially at ambient temperature, the simulation data exagger-

ate the slope of the viscosity-salinity curves and the deviations become increasingly important at high salinities.

Despite that, in all cases, the OPLS forcefield seems to perform better than JC, the viscosity calculations at 298.15 K for the largest salinity (105.5 g/kg) evidenced that this state is beyond the solubility limit. The simulation for this point evolved in a rather bizarre way so we checked the trajectory and discovered the formation of solid NaCl. Figure 5.6 shows a snapshot of one of the final configurations of the system. It is important to note that, since the formation of solid nucleus is an activated process, the observation of spontaneous crystallization is only possible when the solubility limit is largely exceeded. This means that the results for the OPLS at medium to high salinities likely correspond to a metastable system. This fact rises serious doubts about the validity of the OPLS forcefield.
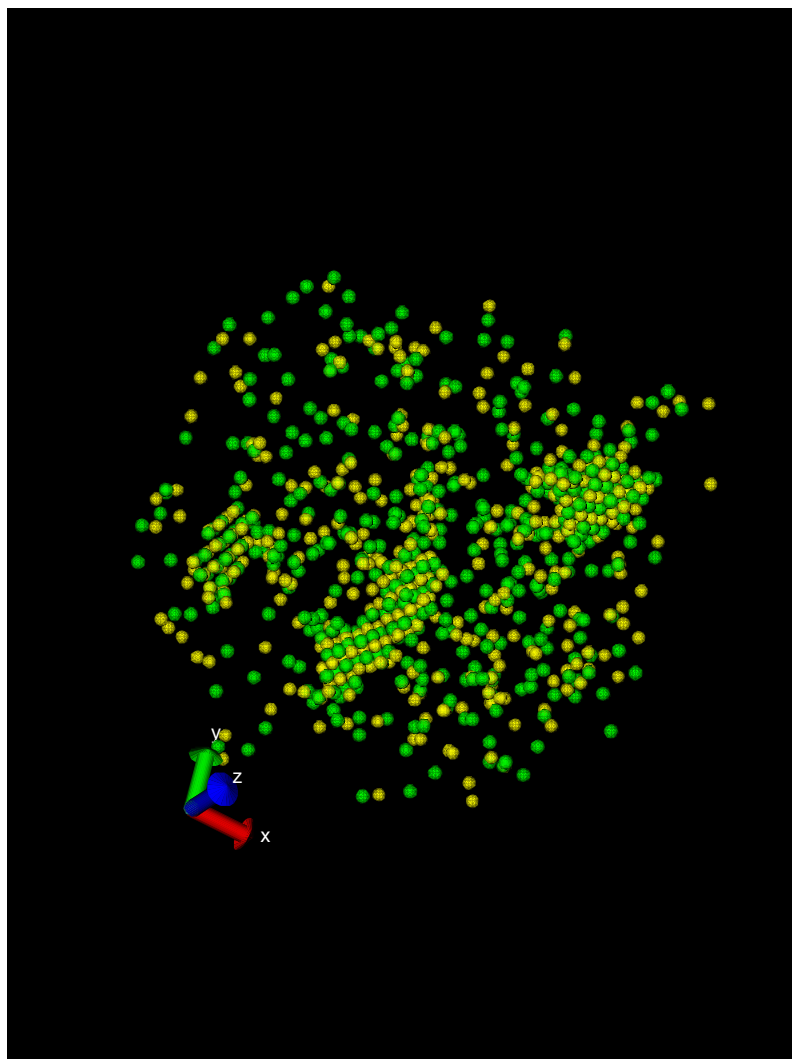
Figure 5.6: **A snapshot of one of the final configurations of a run for the OPLS forcefield at 298.15 K and S=105.5 g/kg showing the formation of NaCl crystallites (the water molecules have been removed for clarity).**
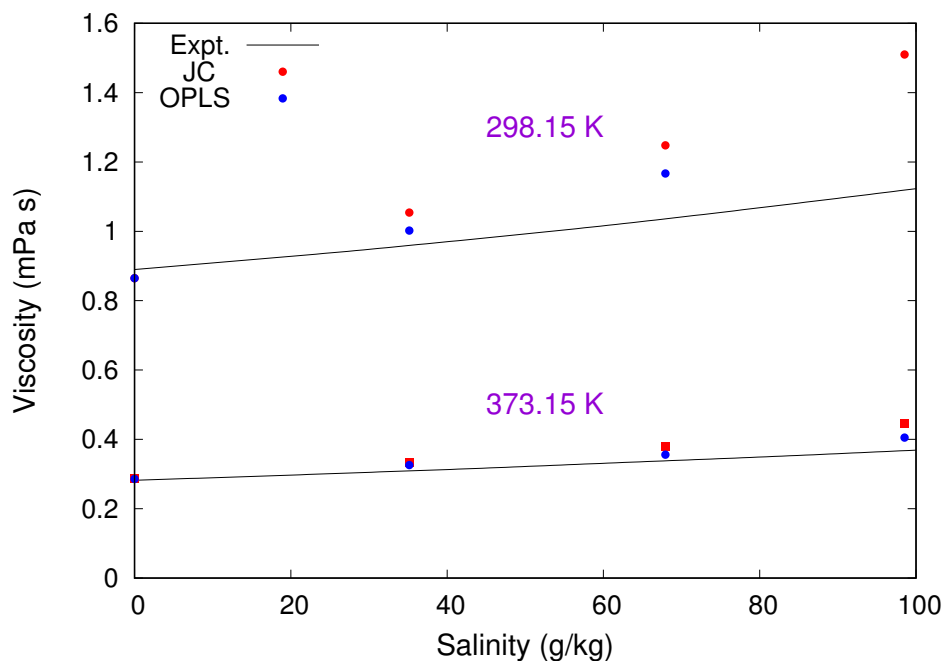
Figure 5.7: **Viscosity of seawater as a function of salinity for the 298.15 K and 373.15 K isotherms at 1 bar using the OPLS and JC interaction potentials.**

Following a similar line of reasoning as in the case of density, we are interested in assessing the role of the main ionic components of seawater and, thus, we have checked the performance of the OPLS and JC forcefields in aqueous NaCl solutions. Figure 5.8 represents the viscosities of a 0.628 molal NaCl solution calculated at several temperatures (at 1 bar) and Figure 5.9 depicts the viscosities as a function of the NaCl concentration for the 298.15 K and 373.15 K isotherms (the concentration range has been limited up to 2 molal to avoid the the solid NaCl nucleation). As in the previous case, the trends of the calculated viscosities for the NaCl solution are essentially the same as those for seawater. Identical conclusions to those obtained for the density can be drawn from the analysis of Figs. 5.8 and 5.9 (NaCl) which show similar patterns to Figs. 5.5 and 5.7 (seawater). In summary, the results for a dynamical property, as the shear viscosity, also leads to the conclusion that having a good model of NaCl interactions is a requisite for an accurate forcefield for seawater.
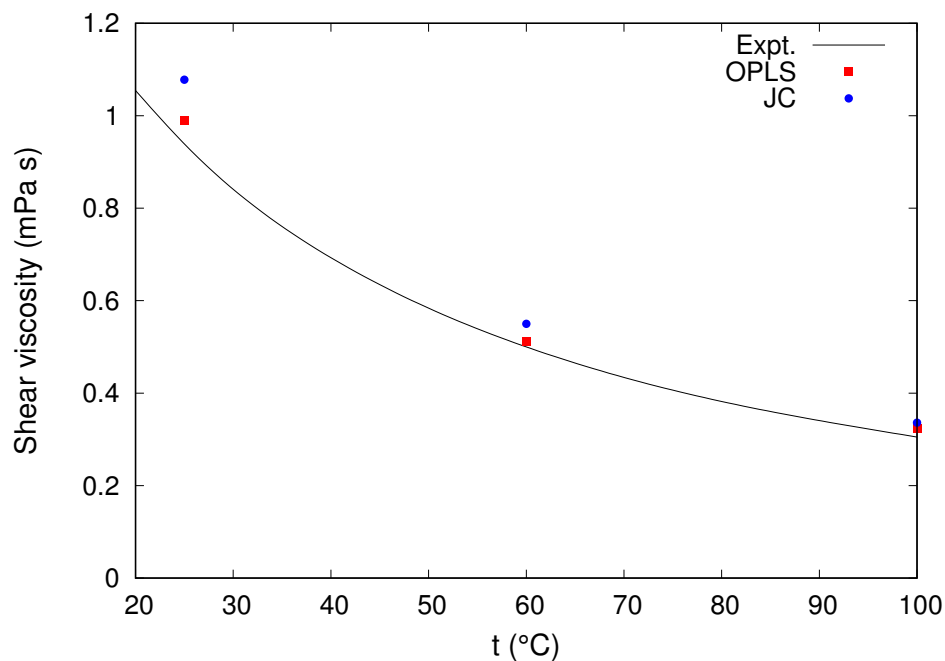
Figure 5.8: **Calculated shear viscosities for a 0.628 molal NaCl solution at several temperatures (using the OPLS and JC forcefields) compared to experimental values.**
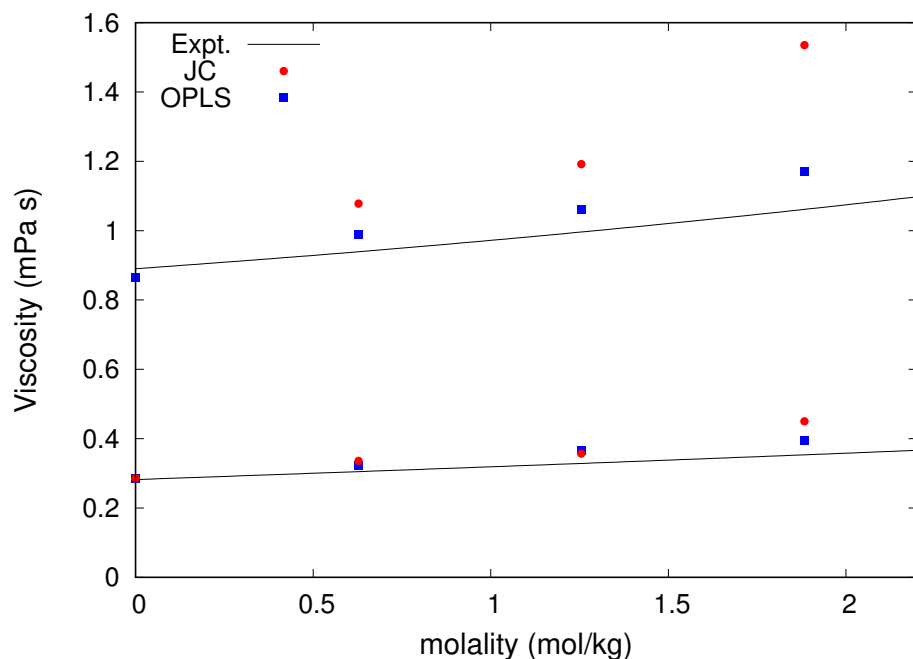
Figure 5.9: **Calculated shear viscosities of NaCl solutions (using the OPLS and JC forcefields) at different concentrations for the 298.15 K and 373.15 K isotherms at 1 bar compared to experimental data.**

5.2.2 Self diffusion coefficient

To the best of our knowledge no experimental data have been reported for the self-diffusion coefficients of the seawater components. However these results are easily at our reach in molec-

ular simulations. This fact shows the potentiality of the simulation in supplying information not accessible experimentally. The results presented in the precedent sections give us some confidence in the predictions of the OPLS and JC forcefields so we may get relevant information about the diffusivity of the seawater components.

Figure 5.10 shows the Mean Square Displacement as a function of time for seawater (the sample contains 15210 molecules) and Mg^{2+} (18 ions). The number of particles affects the uncertainty of the calculations: while the MSD of water exhibits an almost perfect linear dependence up to 2000 ns, the MSD of Mg^{2+} shows a considerable drift from ≈ 800 ps.
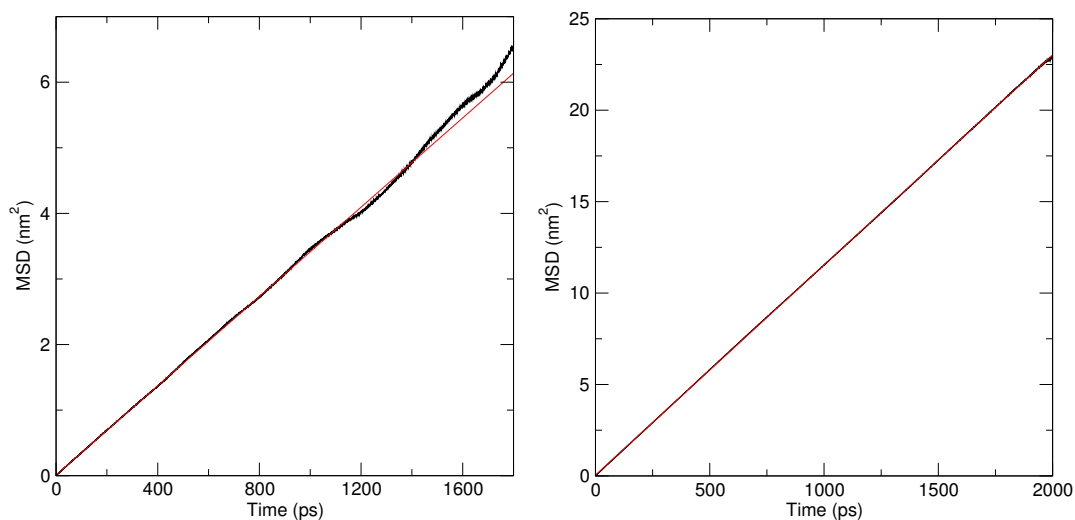
Figure 5.10: **MSD in function of t at 1 bar y 298 K. The MSD function is the black one and red one is the adjustment to a straight line. (a) Magnesium's MSD. (b) Water's MSD.**

Table 5.2 presents the self diffusion coefficients of water for the simplified ISsw and sISsw (see Table 4.1) seawater compositions evaluated with the OPLS and JC forcefields. All the coefficients were obtained at ambient conditions. The time used for the linear regression (Eq. 4.12) was 2000 ps. Although the self-diffusion coefficient of water changes depending on the involved potential interactions its value for a given forcefield is noticeably independent of the seawater composition: the results obtained when considering a more or less detailed seawater composition agree within the statistical uncertainty (around 3%). This means that, at this level of accuracy, the effect of replacing Ca^{2+} by Mg^{2+} and K^+ by Na^+ can not be detected.

The self-diffusion coefficients of the ionic components of the simplified model for the Reference-Composition (sISsw) seawater at ambient conditions are given in Table 5.3. Since the number of ions is much smaller than that of the water molecules the uncertainty is now much larger (although, probably, the distinct values obtained for the Na^+ ion with the OPLS and JC forcefields are a reflect of their differences in the Na^+-water interactions). The diffusivity of the cations is sensibly smaller than that of water and bear no relation with the atomic masses.

5. Results

Table 5.2: Self diffusion coefficient of water for the simplified ISsw and sISsw seawater samples (see Table 4.1) at 1 bar and 298.15 K .

| | $D_w/(10^{-5}\ \mathrm{cm^2/s})$ | | | |
| | OPLS | | JC | |
t(°C)	sISsw	ISsw	sISsw	ISsw
0	0.527	-	0.578	-
25	1.002	0.990	1.054	1.040
60	1.915	1.960	2.043	2.066

This suggest that the effective ionic mass includes the hydration shell In fact, it anticipates a strong hydration shell in the case of Mg^{2+}. D_{Cl^-} seems to be less affected by the surrounding water molecules (its atomic mass is almost twice that of water so it must have a reduced mobility). Finally, the high molecular mass of SO_4^{2-} explains in part a low diffusion coefficient though its polyatomic nature may also contribute its small diffusivity.

Table 5.3: Self-diffusion coefficients of ions in the simplified model for the Reference-Composition seawater (15210 water, 136 Na^+, 156 Cl^-, 8 SO_4^{2-}, 18 Mg^{2+}) at 298.15 K, 1 bar.

| | $D_{ion}/(10^{-5}\ \mathrm{cm^2/s})$ | |
	JC	OPLS
Na^+	0.73	0.88
Cl^-	1.14	1.11
Mg^{2+}	0.56	0.53
SO_4^{2-}	0.31	0.32

Table 5.4 presents the values of the self-diffusion coefficients of the Na^+ and Cl^- ions in a 0.628 molal NaCl solution. The results are coincident (within the statistical uncertainty) with those obtained for the Reference-Composition seawater. This indicates that the presence of Mg^{2+} and SO_4^{2-} ions does not affect the self-diffusion coefficient of Na^+ and Cl^- in seawater.

Table 5.4: Self-diffusion coefficients of ions in a 0.628 molal NaCl solution (15210 water, 172 Na$^+$, 172 Cl$^-$) at 298.15 K, 1 bar.

	$D_{ion}/(10^{-5}$ cm^2/s)	
	JC	OPLS
Na$^+$	0.77	0.87
Cl$^-$	1.17	1.07

5.3 Structure

5.3.1 Ion-Water Structure

The anion-water radial distribution functions (rdf) have been represented in Figures 5.11 and 5.12. The hydration numbers are also displayed in the figures and their numerical values are given in All the rdf's and the respective hydration numbers have been evaluated —using the JC and OPLS interaction potentials— for the simplified Reference-composition seawater (column sISsw in Table 4.1) and for a 0.628 m NaCl aqueous solution. Because of the similarities between the results for both models we only show the calculations for the JC potential. Also, as observed in the figures, the general aspect of the Na^+-water rdf's in seawater is coincident with that of the NaCl solution. In summary, the general features of the hydration of the anions are independent on the details of the forcefield or the composition. These characteristics are then essentially dependent on the nature of the ion, mostly the sign and the nominal value of the ionic charge.

Figure 5.11: **(left) Cl^--water radial distribution functions and the corresponding hydration numbers, calculated with the JC potential for our simplified Reference-Composition seawater, sISsw (15210 water, 136 Na^+, 156 Cl^-, 8 SO_4^{2-}, 18 Mg^{2+}) at $T = 298.15$ K, 1 bar. (right) Same as the left panel but for a 0.628 molal aqueous NaCl solution (15210 water, 172 Na^+, 172 Cl^-).**
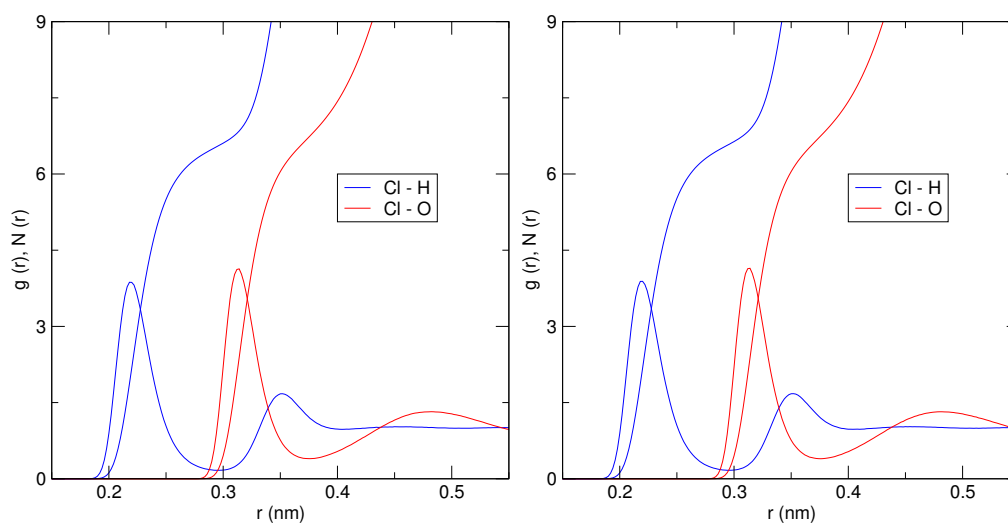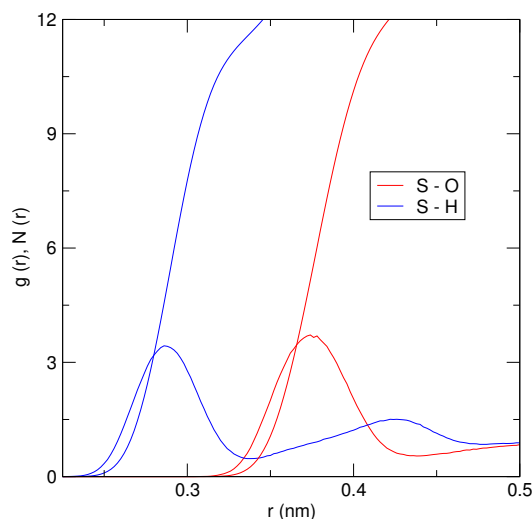
Figure 5.12: **(left) Sulfate-water radial distribution functions and the corresponding hydration numbers, calculated with the JC potential for our simplified Reference-Composition seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+})** at $T = 298.15$ **K, 1 bar. (right) Same as the left panel but for a 0.628 molal aqueous NaCl solution (15210 water, 172 Na$^+$, 172 Cl$^-$).**

In the case of anions, it may be noted that the first peaks for hydrogen and oxygen have almost identical amplitudes and heights, their only difference being that they appear shifted in the distance. Also their integrals seem to have similar trends. This means that the probabilities of finding an oxygen atom at a certain distance r from the center of the anion is essentially the same as that of finding an hydrogen atom at $r + \delta r$. Interestingly δr is just the OH bond distance. The numerical results of hydration numbers, presented Table 5.5, confirm these assertions.

The picture emerging from these results is presented in Figure 5.13(left). The water molecules in the first layer approach the anion with an hydrogen atom closer than the oxygen. The OH bond points towards the center of the anion so that the number of hydrogens in the first shell is coincident with that of oxygens. In addition the position of the first anion-H and anion-O peaks should appear (and, indeed, they do it) shifted by the OH length (roughly 0.1 nm). Moreover, since the water HOH angle is far from 180° the position of the second maximum of the anion-H rdf appears closer to the first anion-O peak than the first anion-H maximum.

Table 5.5: Ion-water hydration numbers (HN) at ambient conditions for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+}), and for a 0.626 molal NaCl solution (15210 water, 172 Na$^+$, 172 Cl$^-$). The HN values have been calculated as the average number of each ion at the distance (in nm) of the first minimum of the Ow-ion and Hw-ion rdf's.

sISsw seawater

| | Ow-ion | | | | Hw-ion | | | |
| | JC | | OPLS | | JC | | OPLS | |
	Distance	HN	Distance	HN	Distance	HN	Distance	HN
Cl$^-$	0.375	6.7	0.385	6.6	0.295	6.5	0.305	5.9
Na$^+$	0.32	5.9	0.325	5.2	0.38	14.9	0.38	13.9
Mg^{2+}	0.27-0.30	5.94	0.27-0.30	5.94	0.30	11.9	0.32	11.9
S	0.44	12.8	0.44	12.9	0.34	11.7	0.34	12.3

NaCl 0.628 molal

| | Ow-ion | | | | Hw-ion | | | |
| | JC | | OPLS | | JC | | OPLS | |
	Distance	HN	Distance	HN	Distance	HN	Distance	HN
Cl$^-$	0.375	6.7	0.385	6.6	0.295	6.6	0.305	5.9
Na$^+$	0.32	5.95	0.325	5.2	0.375	14.7	0.38	13.6

The hydration numbers of the Cl$^-$ ion are in all cases around 6. However, this quantity signals the only remarkable differences between the predictions of the JC and the OPLS force-fields, the latter providing slightly smaller Cl$^-$-H hydration numbers in seawater and the NaCl solution. As expected, the sulfate ion is surrounded by many more water molecules (≈ 12) than the chloride. This is a consequence of its double negative charge as well as its bigger size which allows to accommodate a larger number of water molecules around it. Unexpectedly, the number of hydrogens in first layer of the sulfates are now larger for the OPLS potential interaction (we recall that we have used the same potential function[41] for SO$_4^{2-}$ in both forcefields).
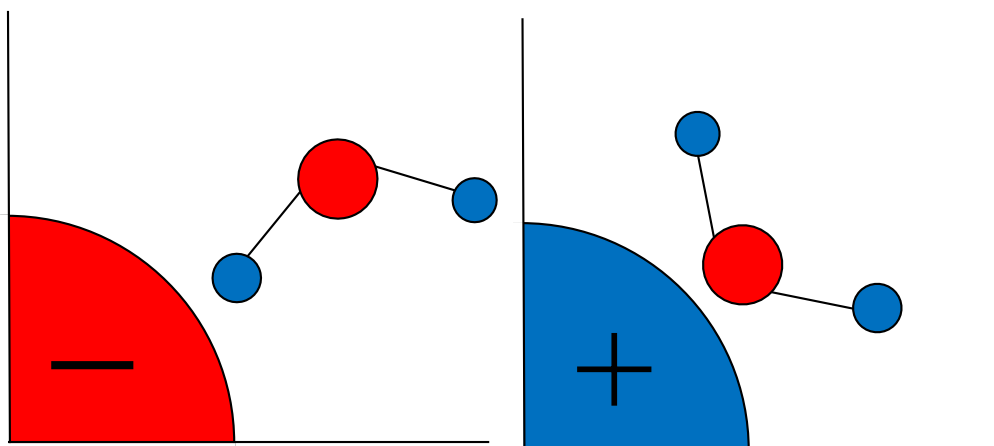
Figure 5.13: **(a) Configuration of the water molecule close to an anion. (b) Configuration of the water molecule close to a cation.**

Looking at the Figures for the cations (Figs. 5.15 and 5.13) we may notice that now the height of the first cation-O peaks are much larger than that of the first cation-H maxima. In fact, the hydration numbers (5.5) of the hydrogen atoms are roughly twice those of oxygens. These results results the configuration of the first water layer around the cations displayed in Fig. 5.13(right). The closer atom of a water molecule approaching a cation is the oxygen one with both hydrogens placed symmetrically and, thus, at almost the same distance from the ion. This picture is almost perfect for the magnesium cation where the Mg^{2+}-H hydration numbers exactly double those of Mg^{2+}-O. Notice also the large planar region of the magnesium rdf's corresponding to six strongly bounded water molecules in a very tight arrangement.

On the other hand, the hydration of the sodium cations is less rigid. Although the number of oxygens in the first layer is again around 6, the hydration numbers of the hydrogen atoms is more than twice this figure indicating that the hydrogens of some extra water molecules may enter into the first hydration shell. The HN numerical values for Na^+ also reveal a slightly different hydration properties of the JC and the OPLS forcefield. Finally, it is worth mentioning that the results for the hydration of Na^+ in seawater are completely coincident with those in a one-component NaCl solution indicating that the presence of other ions (even if they are divalent) does not disturb essentially the local Na^+-water structure.
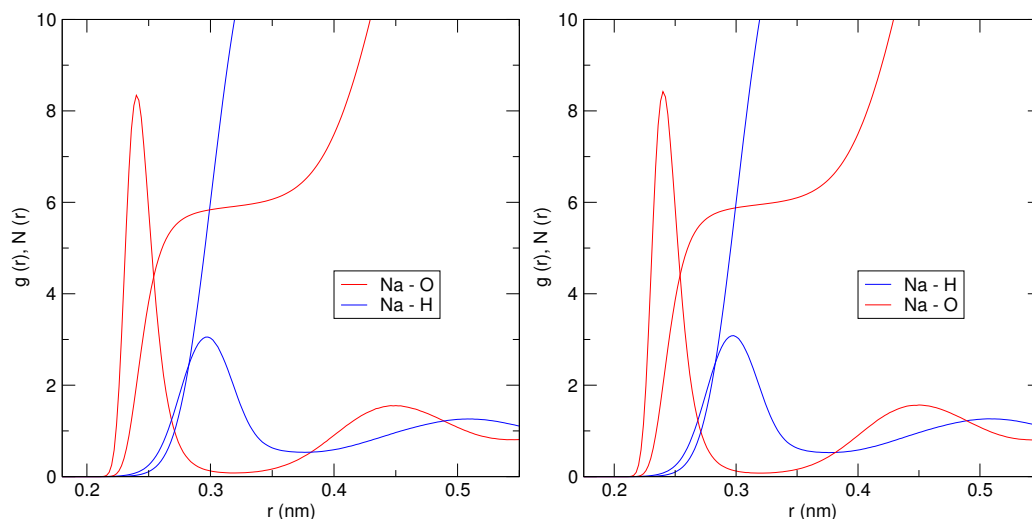
Figure 5.14: **(left) Na$^+$-water radial distribution functions and the corresponding hydration numbers, calculated with the JC potential for our simplified Reference-Composition seawater, sISsw (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+})** at $T = 298.15$ K, 1 bar. **Lower curves are the corresponding coordination numbers. (right) Same as the left panel but for a 0.628 molal aqueous NaCl solution (15210 water, 172 Na$^+$, 172 Cl$^-$).**
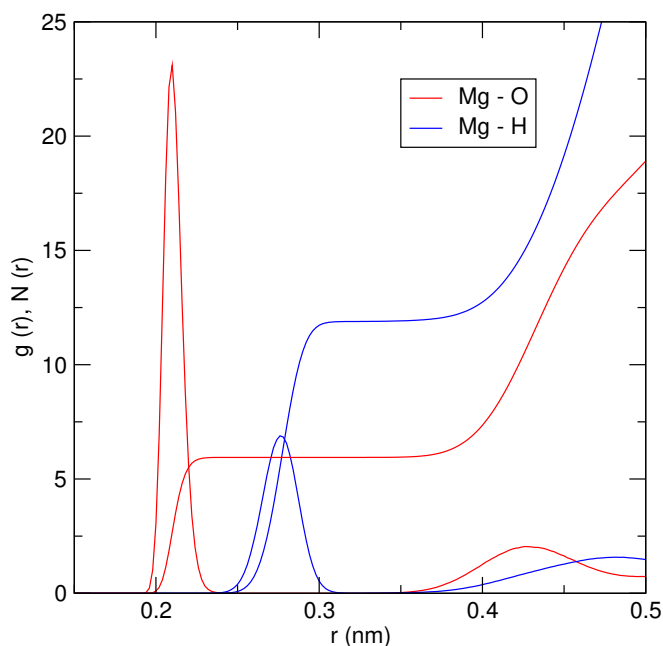
Figure 5.15: **Mg^{2+}-water radial distribution functions, calculated with the JC potential, for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+})** at $T = 298.15$ K, 1 bar.

5.3.2   Ion-Ion Structure

Figures 5.16, 5.17, 5.18, and 5.19 show the rdf's for the cation-anion pairs in sISsw seawater. In Fig. 5.16 the Na$^+$-Cl$^-$ radial distribution functions of a 0.628 molal NaCl solution are also displayed. At this point it cannot be a surprise that the rdf in seawater is completely coincident

with that in a one-component NaCl solution of similar concentration. This confirms, once again, that the presence of other ions (even if they are divalent) does not disturb essentially the local structure. Table 5.6, presents the numerical values of the more interesting features (position and heights of the first maximum, first minimum and second maximum) of selected distribution functions.
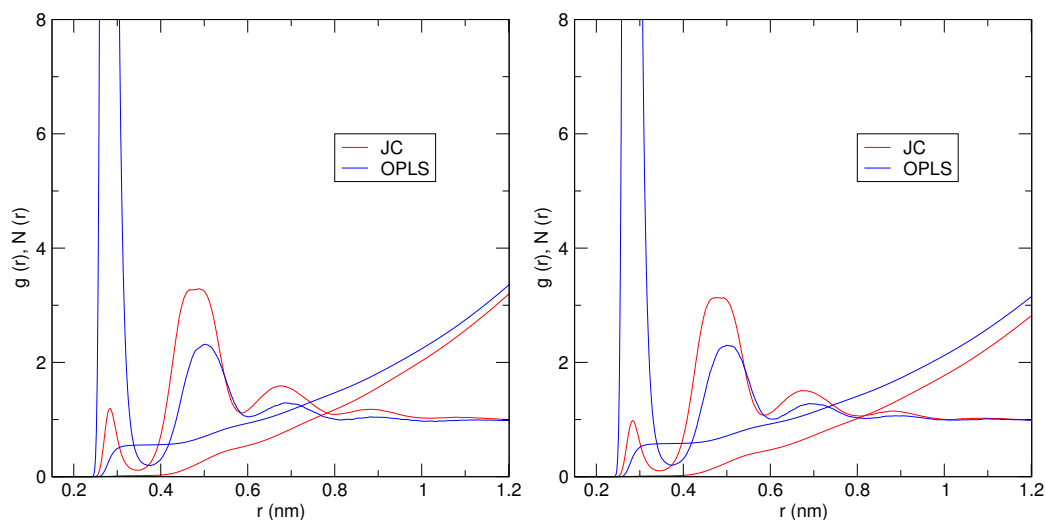

Figure 5.16: **(left) Na^+-Cl^- radial distribution functions and coordination numbers calculated with the JC and OPLS potentials for sISsw seawater (15210 water, 136 Na^+, 156 Cl^-, 8 SO_4^{2-}, 18 Mg^{2+}) at T = 298.15 K, 1 bar. (right) Same as the left panel but for a NaCl solution at similar salinity (15210 water, 172 Na^+, 172 Cl^-).**

Contrary to the trends exhibited by the ion-water distribution functions, the predictions for the rdf's between unlike charged ions are considerably different for the JC and OPLS forcefields. The Na^+-Cl^- rdf provides a paradigmatic example. In this case the OPLS system exhibits a huge first peak (in fact, we have been forced to cut it in the plot). This means that, in the OPLS system, every ion is surrounded by a significant proportion of ions of opposite charge at distances smaller than the hydration layer. It is usual to name them as contact ion pairs (CIP). The JC results lead to a quite different picture: a small first peak followed by second maximum of considerable height indicates that most of the unlike charged neigbours of an ion are placed beyond the hydration layer. These anion-cation pairs are usually referred to as solvent separated ion pairs (SSIP).

It has been suggested that the solubility of models of NaCl solutions at ambient conditions may be obtained as the concentration at which the number of ionic pairs is close to 0.075[47, 48]. Since the CIP reaches a value over 0.50 in the OPLS 0.628 NaCl solution, assuming the rule to be valid, it follows that the system would largely be beyond the solubility limit. Given the similarities between the behaviour of the sISsw seawater and the 0.628 NaCl solution, it seems clear that the OPLS fails completely in the prediction of the ion structure (more precisely the

Na$^+$−Cl$^-$ rdf). This fact explains the problems detected in the calculation of viscosity where we observed the formation NaCl crystallites in the high salinity systems.

The differences in the ionic structure produced by the OPLS and JC forcefields are not limited to the Na$^+$−Cl$^-$ pairs. Figure 5.17 shows that the Na$^+$-S distribution functions are somewhat different for both models. Now it is the JC forcefield that predicts a huge first peak, much higher than that of the OPLS model. The number of ionic pairs is significant for the JC system (an average of 0.75 sodium ions surround every sulfate) and considerably smaller for the OPLS potential (0.24). Notice that despite these relatively large number of ion pairs, the low concentration of sulfate anions disables the possibility of ionic aggregation.
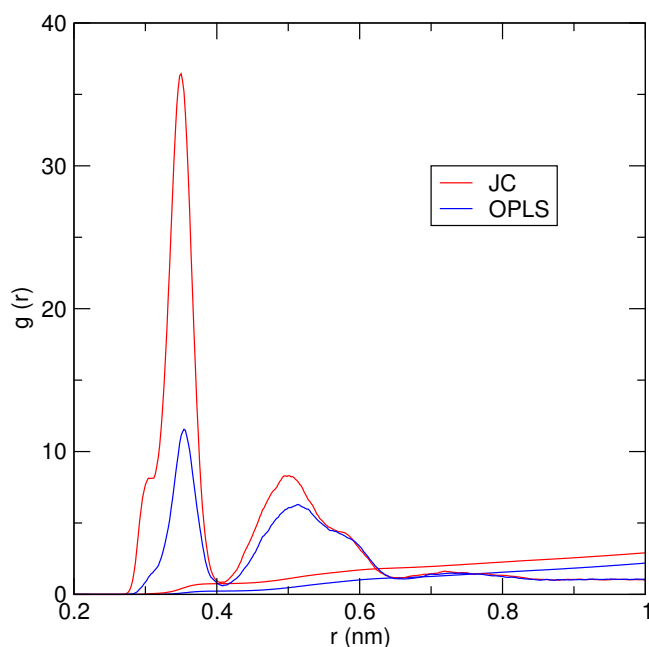
Figure 5.17: **Na$^+$−S distribution functions and coordination numbers for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+}) as obtained with the OPLS and JC forcefields at T=298.15 K, 1bar.**

Although the overall aspect of the Mg^{2+}-Cl distribution functions of OPLS and JC are quite similar (Figure 5.19), its most significant feature is that the positions of the extrema appear slightly shifted in both models. The first maximum occurs at quite small distances (0.25-0.26 nm). Because of this and also due to the low Mg^{2+} concentration, the height of the first maximum is not relevant: a peak height of around 10 only leads to a CIP around 0.05. The second maximum of this function appears considerably shifted, at slightly below 0.5 nm, and involves a significant number of ion pairs, 0.71 and 0.48 for the JC and OPLS systems, respectively.
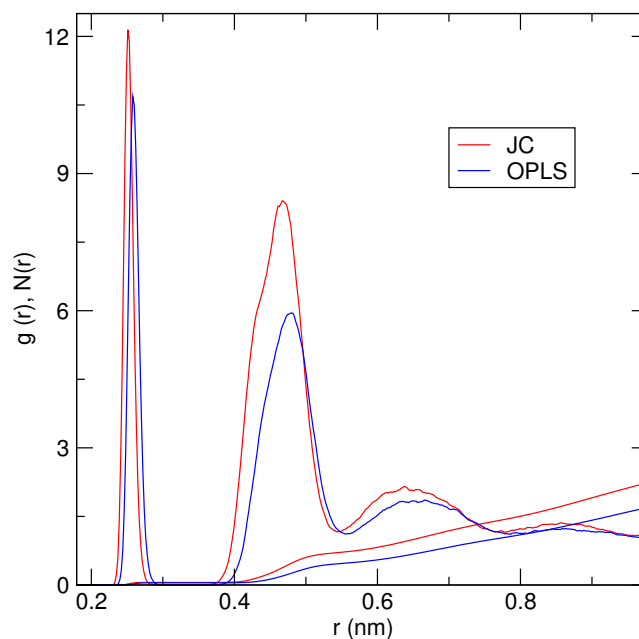
Figure 5.18: **Mg−Cl distribution functions and coordination numbers for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+}) as obtained with the OPLS and JC forcefields at T=298.15 K, 1bar.**

The Mg^{2+}-S distribution functions (Figure 5.19) show a quite peculiar behaviour: a large and wide first maximum centered at $r = 0.48$ followed by an extented and much less structured region. Consequently, the coordination numbers grow steadily up to the distance of the first minimum (at $r = 0.60$, the average number of sulfates surrounding a magnesium ion is already 0.21 and 0.07 in the OPLS and JC systems, respectively) and increases quite slowly afterwards ($N_c = 0.25$ and $N_c = 0.10$ at $r = 0.90$, respectively). These numbers also indicate that, even if the overall picture is the same for both forcefields, there are significant quantitative differences among them.

The features of the Mg^{2+}-S distribution functions are probably due to a combination of three factors. Firstly, these ions are divalent and, thus, the electrostatic interactions are very strong (four times larger than those between monovalent ions). In addition, as commented above, the magnesium ion is firmly hydrated which makes it act effectively as a large ion. This, together with size and shape of the sulfate, accounts for the large and wide first peak of the rdf. Finally, it should be expected that, apart of these strongly binded unlike charged ion pairs, most of the charge cancellation of a divalent ion would be due to presence of the most abundant monovalent ions. In other words despite the strong electrostatic interactions the scarce divalent ions cannot compete with the most abundant monovalent ones at medium to large distances.
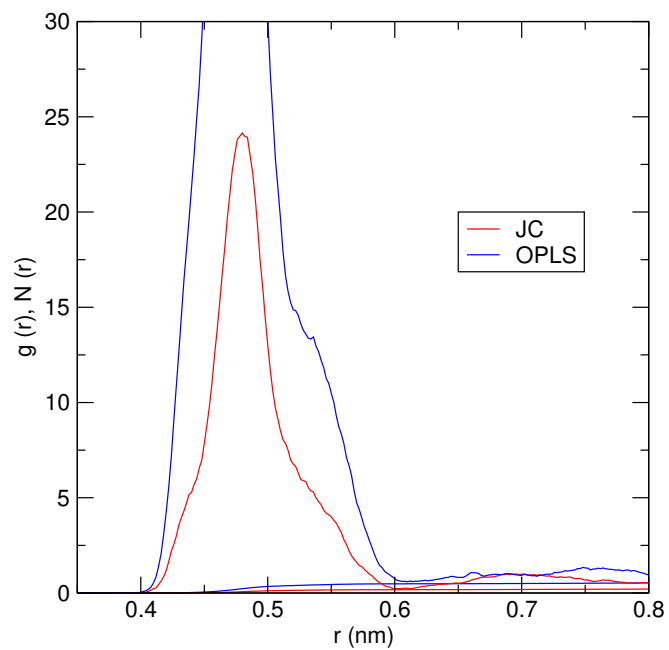
Figure 5.19: **Mg−S distribution functions and coordination numbers for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+}) as obtained with the OPLS and JC forcefields at T=298.15 K, 1bar.**

Table 5.6: Position and heights of the main peaks of the ion-ion radial distribution functions calculated with the JC and OPLS potentials for sISsw seawater (15210 water, 136 Na$^+$, 156 Cl$^-$, 8 SO$_4^{2-}$, 18 Mg^{2+}) at T=298.15 K, 1 bar. Data in brackets represent a range of $g(r) = 0$. Data separated by a semicolon represent a double peak.

| | First maximum | | | | First minimum | | | | Second maximum | | | |
| | JC | | OPLS | | JC | | OPLS | | JC | | OPLS | |
	r(nm)	height	r(nm)	height	r(nm)	height	r(nm)	height	r(nm)	height	r(nm)	height
Na - Cl	0.282	0.97	0.275	60	0.346	0.096	0.374	0.185	0.479	3.12	0.504	2.32
Na - Na	0.379;0.430	0.612;0.681	0.387	2.78	0.499	0.27	0.430	1.82	0.625	1.08	0.464	2.26
Cl - Cl	0.510	1.13	0.455	2.15	0.613	0.41	0.625	0.54	0.747	0.98	0.75	1.04
Mg - Cl	0.251	12.1	0.250	11.5	[0.29–0.34]	0	[0.30–0.36]	0	0.471	8.3	0.478	6.05
Na - S	0.349	37	0.352	11.9	0.407	0.88	0.413	0.60	0.500	8.4	0.515	6.4
Mg - S	0.480	23	0.472	71	0.607	0.23	-	-	-	-	-	-

Chapter 6

Conclusions

About 71% of the earth is covered with seawater, a very complex substance from the point of view of the composition. A lot of work has been done about its composition and its chemical-physical properties. These considerations give strong support for the development of a computational model able to simulate the seawater properties.

The first contribution of this work is just to show that the molecular simulation of seawater is at our reach. We have shown that we may discard very minor components of the complex seawater composition and substitute it by a simpler one which is amenable for computer simulation: the Reference Composition of Standard seawater may be quite well represented by a system consisting of 15210 water molecules and 318 ions (133 Na^+, 156 Cl^-, 8 SO_4^{2-}, 15 Mg^{2+}, 3 Ca^{2+}, and 3 K^+).

Through molecular dynamics simulations we have investigated the results for this seawater composition for two different forcefields based on the use of the TIP4P/2005 model to describe the water-water interactions. The only difference between them are the parameters representing the interactions between the most abundant ions in seawater, Na^+ and Cl^-. For these we have employed the Joung-Cheatham and the OPLS potentials. For the rest of the interionic potentials we use a common set of parameters (taken from the literature) in both forcefields. We have shown that the predictions for the JC and OPLS potentials are in good agreement with the experimental measurements for the density and acceptable for the viscosity.

One of the aims of this study was to investigate the importance of minor constituents in the seawater properties. To this effect we have carried out simulations for a system in which we replace the K^+ and Ca^{2+} ions by Na^+ and Mg^{2+}, respectively. A significant conclusion of this work is that the properties of seawater are almost the same (within the statistical uncertainties) in both systems.

Computer simulation is able to provide information on a number of properties not easily

accessible to the experiments. We have been able to yield some results for the water and ionic diffusivity in seawater as well as to give a description of the structure of this imposrtant fluid.

The comparison with experimental results (when possible) show that the predictions of the JC and OPLS forcefields are quite similar. However some of the simulations at high salinity using the OPLS potential lead to the formation of solid ionic nuclei indicating that this model has a solubility limit considerably lower than the experimental one. It is probably even more important that the structural properties of the OPLS system denoted a strong ionic association even at low salinities. This fact throws serious doubts about the validity of the OPLS forcefield despite their acceptable predictions for the density and the viscosity of seawater.

In order to assess the origin of the departures of the simulation results from the experimental data for seawater we have also performed some simulations for NaCl aqueous solutions. We have observed that the differences between simulation and experiment are exactly the same in seawater and NaCl solutions of equivalent salinity. This indicates that a more satisfactory description of the Na^+-water, Cl^--water and Na^+-Cl^- interations is a requisite for a better account of the seawater properties. A recently proposed forcefield for NaCl aqueous solutions could be an advance along these lines. However, since the model is based on partial ionic charges, it is not compatible with the currently proposed interactions for other ions. A considerable effort is yet to be done to provide a consistent forcefield for the simulation of seawater.

Bibliography

[1] T. Kuhlbrodt, A. Griesel, M. Montoya, A. Levermann, M. Hofmann, and S. Rahmstorf, "On the driving processes of the atlantic meridional overturning circulation," *Reviews of Geophysics*, vol. 45, 2007.

[2] S. Rahmstorf, "Ocean circulation and climate during the past 120,000 years," *Nature*, vol. 419, p. 207, 2002.

[3] M. E. Wieser, "Atomic weights of the elements 2005 (iupac technical report)," *Pure and Applied Chemistry*, vol. 78, pp. 2051–2066, 2006.

[4] J. W. Sandström, "Dynamische versuche mit meerwasser," *Ann. Hydrog. Mar. Meteorol.*, vol. 36, pp. 6–23, 1908.

[5] S. Rahmstorf, "Thermohaline circulation: The current climate," *Nature*, vol. 421, p. 699, 2003.

[6] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids*. Oxford: Oxford University Press, 1987.

[7] D. Frenkel and B. Smit, *Understanding Molecular Simulation*. London: Academic Press, 1996.

[8] R. Pawlowicz, "Key physical variables in the ocean: temperature, salinity, and density," *Nature Education Knowledge*, vol. 4, p. 13, 2013.

[9] J. L. F. Abascal and C. Vega, "A general purpose model for the condensed phases of water: Tip4p/2005," *J. Chem. Phys.*, vol. 123, p. 234505, 2005.

[10] W. L. Jorgensen and J. Tirado-Rives, "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin," *Journal of the American Chemical Society*, vol. 110, pp. 1657–1666, 1988.

[11] I. S. Joung and T. E. Cheatham, "Determination of alkali and halide monovalent ion parameters for use in explicit solvated biomolecular simulation," *J. Phys. Chem. B*, vol. 112, p. 9020, 2008.

[12] A. J. G. Marcet, "On the specific gravity, and temperature of sea waters, in different parts of the ocean, and in particular seas; with some account of their saline contents," *Philosophical Transactions of the Royal Society of London*, vol. 109, pp. 161–208, 1819.

[13] G. Forchhammer, "On the constitution of sea-water, at different depths, and in different latitudes," *Proc. R. Soc. London*, vol. 12, pp. 129–132, 1862-1863.

[14] W. Dittmar, "Report on researches into the composition of ocean water, collected by H.M.S. *Challenger*," *Challenger Repts., Phys. and Chem.*, vol. 1, pp. 1–251, 1884.

[15] M. Knudsen, "On the standard-water used in the hydrographical research until july 1903," *Conseil permanent international pour l'exploration de la mer. Publications de circonstance no. 2*, 1903.

[16] R. A. Cox, F. Culkin, and J. P. Riley, "The electrical conductivity/chlorinity relationship in natural sea water," *Deep Sea Research*, vol. 14, pp. 203–220, 1967.

[17] F. Culkin and R. A. Cox, "Sodium, potassium, magnesium, calcium and strontium in sea water," *Deep Sea Research*, vol. 13, pp. 789–804, 1966.

[18] A. W. Morris and J. P. Riley, "The bromide/chlorinity and sulphate/chlorinity ratio in sea water," *Deep Sea Research*, vol. 13, pp. 699–705, 1966.

[19] J. P. Riley and M. Tongudai, "The major cation/chlorinity ratios in sea water," *Chemical Geology*, vol. 2, pp. 263–269, 1967.

[20] J. H. Carpenter and M. E. Manella, "Magnesium to chlorinity ratios in seawater," *Journal of Geophysical Research*, vol. 78, pp. 3621–3626, 1973.

[21] T. B. Warner, "Normal fluoride content of seawater," *Deep Sea Research*, vol. 18, pp. 1255–1263, 1971.

[22] L. R. Uppström, "The boron/chlorinity ratio of deep-sea water from the pacific ocean," in *Deep Sea Research*, vol. 21, pp. 161–162, 1974.

[23] UNESCO, "The practical salinity scale 1978 and the international equation of state of seawater," *UNESCO Technical Papers in Marine Science*, vol. 36, 1981.

[24] UNESCO, "Background papers and supporting data on the practical salinity scale 1978," *UNESCO Technical Papers in Marine Science*, vol. 37, 1981.

[25] F. J. Millero, R. Feistel, D. G. Wright, and T. J. McDougall, "The composition of standard seawater and the definition of the reference-composition salinity scale," *Deep Sea Research I*, vol. 55, pp. 50–72, 2008.

[26] D. G. Wright, R. Pawlowicz, T. J. McDougall, R. Feistel, and G. M. Marion, "Absolute salinity,"density salinity"and the reference-composition salinity scale: present and future use in the seawater standard TEOS-10," *Ocean Science*, vol. 7, no. 1, pp. 1–26, 2011.

[27] B. J. Alder and T. E. Wainwright, "Phase transition for a hard sphere system," *J. Chem. Phys.*, vol. 27, pp. 1208–1209, 1957.

[28] L. Verlet, "Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules," *Phys. Rev.*, vol. 159, pp. 98–103, 1967.

[29] J. E. Jones, "On the determination of molecular fields. ii. from the equation of state of a gas," *Proceedings of the Royal Society of London A*, vol. 106, pp. 463–477, 1924.

[30] A. Rahman and F. H. Stillinger, "Molecular Dynamics study of liquid water," *J. Chem. Phys.*, vol. 55, pp. 3336–3359, 1971.

[31] A. R. Leach, *Molecular Modelling Principle and Applications*. Longmans, 1996.

[32] P. Ewald, "Die berechnung optischer und electrostatischer gitterpotentiale," *Annalen der Physik*, vol. 64, p. 253, 1921.

[33] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *J. Chem. Phys.*, vol. 103, pp. 8577–8593, 1995.

[34] S. Nosé and M. L. Klein, "Constant pressure molecular dynamics for molecular-systems," *Mol. Phys.*, vol. 50, pp. 1055–1076, 1983.

[35] W. G. Hoover, "Canonical dynamics: equilibrium phase-space distributions," *Phys. Rev. A*, vol. 31, pp. 1695–1697, 1985.

[36] M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new Molecular Dynamics method," *J. Appl. Phys.*, vol. 52, pp. 7182–7190, 1981.

[37] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical integration of the cartesan equations of motion of a system with constraints. molecular dynamics of n-alkanes.," *J. Comput. Phys.*, vol. 23, p. 327, 1977.

[38] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "Gromacs: Fast, flexible and free," *J. Comput. Chem.*, vol. 26, p. 1701, 2005.

[39] C. Vega, J. L. F. Abascal, M. M. Conde, and J. L. Aragones, "What ice can teach us about water interactions: a critical comparison of the performance of different water models," *Faraday Discuss.*, vol. 141, pp. 251–276, 2009.

[40] C. Vega and J. L. F. Abascal, "Simulating water with rigid non-polarizable models: a general perspective," *Phys. Chem. Chem. Phys*, vol. 13, pp. 19663–19688, 2011.

[41] W. R. Cannon, B. M. Pettitt, and J. A. McCammon, "Sulfate anion in water: Model structural, thermodynamic, and dynamic properties," *J. Phys. Chem.*, vol. 98, pp. 6225–6230, 1994.

[42] A. Einstein, "On the movement of small particles suspended in stationary liquids required by the molecular-kinetic theory of heat," *Annalen der Physik*, vol. 17, pp. 549–560, 1905.

[43] J. J. Erpenbeck, "Shear viscosity of the lennard-jones fluid near the triple point: Green-kubo results," *Phys. Rev. A*, vol. 38, pp. 6255–6266, 1988.

[44] D. Alfè and M. J. Gillan, "First-principles calculation of transport coefficients," *Phys. Rev. Lett.*, vol. 81, pp. 5161–5164, 1998.

[45] G.-J. Guo and Y.-G. Zhang, "Equilibrium molecular dynamics calculation of the bulk viscosity of liquid water," *Mol. Phys.*, vol. 99, pp. 283–289, 2001.

[46] M. A. Gonzalez and J. L. F. Abascal, "The shear viscosity of rigid water models," *J. Chem. Phys.*, vol. 132, p. 096101, 2010.

[47] A. L. Benavides, J. L. Aragones, and C. Vega, "Consensus on the solubility of nacl in water from computer simulations using the chemical potential route," *J. Chem. Phys.*, vol. 144, p. 124504, 2016.

[48] A. L. Benavides, M. A. Portillo, J. L. F. Abascal, and C. Vega, "Estimating the solubility of 1:1 electrolyte aqueous solutions: the chemical potential difference rule," *Molec. Phys.*, vol. 115, pp. 1301–1308, 2017.