

## UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF CHEMICAL & PROCESS ENGINEERING

M.ENG CHEMICAL & PROCESS ENGINEERING 18530

---

# Computer-aided Design of Bio-inspired Nanoporous Silica Materials

---

*Author:*

André CREScenzo

*Supervisors:*

Miguel JORGE

Alessia CENTI

Carlos F. RANGEL

September 10, 2015

## 1 Summary

...

## Contents

	Page
<b>1 Summary</b>	<b>i</b>
<b>2 Acknowledgements</b>	<b>iii</b>
<b>3 Introduction</b>	<b>1</b>
<b>4 Theoretical Basis</b>	<b>2</b>
4.1 Simulation Methods . . . . .	2
4.1.1 Molecular Dynamics simulation . . . . .	2
4.1.2 Monte Carlo simulation . . . . .	3
4.2 Software Description . . . . .	4
4.2.1 About GROMACS . . . . .	4
4.2.2 About MagiC . . . . .	5
<b>5 Methods Description</b>	<b>7</b>
5.1 Atomistic Molecular Model . . . . .	7
5.2 Atomistic Simulation . . . . .	8
5.3 Coarse-grained molecular Model . . . . .	10
<b>6 Results and Discussion</b>	<b>12</b>
6.1 Atomistic Simulations . . . . .	12
6.2 Coarse-Graining Process: Convergence Analysis . . . . .	14
6.2.1 Pure DMDDn and water . . . . .	15
6.2.2 DMDDs and Bromide . . . . .	16
6.2.3 Pure DMDDs and Bromide (Integrated Potentials) . . . . .	17
6.2.4 DMDDn and Silica . . . . .	18
6.3 Charged Systems . . . . .	18
6.4 Silica-Surfactant interaction . . . . .	21
<b>7 Conclusion</b>	<b>21</b>
<b>8 Nomenclature</b>	<b>21</b>
<b>9 References</b>	<b>22</b>
<b>10 Appendix</b>	<b>1</b>

## 2 Acknowledgements

...

### 3 Introduction

Molecular Dynamics (MD) and Monte Carlo (MC) are powerful tools to simulate molecular interactions of surfactants in solvent systems, allowing a deeper understanding of their self-assembly process (Pérez-Sánchez, Gomes, & Jorge, 2013). This process results in several types of surfactant mesoscale conformations that are specially useful to design bio-inspired silica materials (Patwardhan, 2011). With the addition of silica, these structures behave as scaffolds to mesoporous or nanoporous structures that are maintained even after surfactant removal, silica oligomer polymerization and calcination (Tolbert, Firouzi, Stucky, & Chmelka, 1997). A vast range of silica materials are examples of this phenomena, such as MCM-41 as reported by Kresge and Roth (2013), SBA-15 (Zhao et al., 1998), MSU-V (Taney, Liang, & Pinnavaia, 1997) and many others, in which the self-assembled structure depends on the type of surfactant, concentration of the substances involved and synthesis conditions such as temperature and pH. It should be noted that most of the experimental methods used to obtain data are based on observation and interpretation of final silica structure by using X-ray diffraction (XRD) and transmission electron microscopy (TEM). It follows that initial self-assembled conformations are predicted as a reflex of the final results and little is known about the mechanistic of this process. However, with MD simulations it is possible to observe and analyse these initial steps of self-assembly and predict, with more accuracy, properties and frameworks provided by surfactants (Tresset, 2009).

On the other hand, a major concern is that even though MD uses sophisticated software prepared to simulate systems with thousands of atoms, using all capacities of hardware available, such as high-speed multi-core processors in conjunction with GPUs designed specifically to process data from arrays (Pronk et al., 2013), they can hardly achieve long time horizons and are commonly limited to a few microseconds depending on the size of the system. For this reason, several techniques have been developed to optimize the performance of the simulations such as coarse-grain methods (Pérez-Sánchez et al., 2013). The basic idea of this technique is to fit parameters of atom groups with similar properties in a bigger "bead", which includes atomic masses and electrostatic charges lumped in approximated values. For example, given a simulation of an arbitrary surfactant with a long hydrophobic tail, it is possible to merge three or four carbon atoms (and its hydrogen atoms) in groups since they have similar hydrophobic properties, by this means reducing the number of particles in the system and speeding up the simulation.

In order to provide molecular topologies to simulations software packages several force-fields, for example MARTINI (Marrink, Risselada, Yefimov, Tieleman, & de Vries, 2007), uses Lennard-Jones potentials fitted to a range of pre-defined bead types to describe coarse-grain models. Furthermore, not only intramolecular beads are possible, but also intermolecular beads can be specified, such as multiple solvent molecules merged in a single bead or ions surround by water molecules. Previous works conducted by (Pérez-Sánchez et al., 2013) with this method recreated a model of surfactant in the presence of silica with explicit water that was successful in describing rod-like self-assembly structures detected on MCM-41 materials, demonstrating the capacities of up-scaling this type of systems. Nevertheless, solvent presence demands most of the computational resources, hence implicit solvent scheme has been the focus of many studies (Pronk et al., 2013).

Different concepts have been applied to develop a suitable model for implicit solvents; For example, Arnarez et al. (2015) developed the Dry MARTINI force-field by modifying parameters of its predecessor, in such manner that solvent interactions became incorporated in these values and then solvent beads are no longer necessary. Another method, described by Mirzoev and Lyubartsev (2013) is able to recreate an implicit solvent system from interaction potentials generated from a bottom-up approach, that means by using an all-atoms simulation to generate parameters for the coarse-grain model. It is supposed that these approximated potentials can be refined iteratively by this systematic method based on averages calculated from Monte Carlo simulations. During this process thermodynamic changes originated from solvent interaction with amphiphilic molecules or even electrostatic interactions can be incorporated in the potentials derived, therefore providing a flexible methodology to up-scale systems allowing the simulation of mesoscale self-assembly structures using coarse-graining.

The experiments presented in this work are an attempt to develop a method to upscale silica-surfactant interactions (Tolbert et al., 1997) from the atomistic model to a mesoscale model with the advent of this later coarse-grain technique. The methodology applied to reach the desired model is based on the Magic software package (Mirzoev & Lyubartsev, 2013) that in conjunction with a MD simulation software, in this case GROMACS (Pronk et al., 2013), will provide a suitable approximation to self-assembly of amphiphilic molecules. Further explanations of the process are described in the Experimental Methods section. For the scope of this project, a bolaamphiphilic molecule called 1,12-diaminododecane (DMDD) has been chosen as surfactant because, as seen in previous research by

Tanев et al. (1997), it self-assembles in multilamellar vesicles that in the presence of a silica precursor are capable of generating a mesoporous structure with remarkable properties. In order to validate this structure formation and framework formation for silica oligomers, a coarse-grain approximation is a suitable option since amphiphilic molecules interaction with solvents can be described efficiently with potentials derived from the chosen methodology (?). As a final objective at the end of this project, an implicit water coarse-grain model for DMDD and Silica will be generated and properly validated based on MD simulations and thermodynamic properties, in order to provide a satisfactory approximation of self-assembled silica-surfactant structures in mesoscale.

## 4 Theoretical Basis

This work is based on two molecular simulation techniques. The former, called Molecular Dynamics, accounts for integration of Newton forces in order to describe positions and velocities of particles in the system. The latter, called Monte Carlo method, is a powerful statistical tool based on stochastic inputs to measure properties from many systems, in this case molecular systems in thermodynamic equilibrium. In the following sections the mechanism behind each technique will be described, including additional explanations on how they are applied in simulation software.

### 4.1 Simulation Methods

#### 4.1.1 Molecular Dynamics simulation

Molecular Dynamics is a simulation method that originates from the dynamic nature of atomic interactions. Generally, a system of atoms can be treated as a multi-particle system ruled by Newton's Law (Frenkel & Smit, 2001). Considering a system with  $\mathcal{N}$  atoms, for each  $i$ th particle of the system the following differential equation will determine its dynamic behaviour:

$$m_i \frac{d^2 \vec{r}_i(t)}{dt^2} = \vec{F}_i(t) \quad (1)$$

Where  $\vec{r} = (x, y, z)$  is the position vector and  $F = (F_x, F_y, F_z)$  are the force components. Therefore, for this  $\mathcal{N}$  particle system, it is necessary to solve  $3\mathcal{N}$  differential equations in order to fully describe it analytically at any  $t$ . Since any molecular system involves an enormous number of molecules, solving these equations analytically is impracticable and simulation methods are necessary. The derivative terms need to be numerically calculated and as soon as simulation efficiency is a major concern, a suitable approximation to the second derivative of position is the Taylor's expansion, taking the  $x$  coordinate as an example, for a given  $\Delta t$  is:

$$x(t + \Delta t) = x(t) + \Delta t \frac{dx(t)}{dt} + \frac{1}{2!} \Delta t^2 \frac{d^2 x(t)}{dt^2} + \frac{1}{3!} \Delta t^3 \frac{d^3 x(t)}{dt^3} + \mathcal{O}(\Delta t^4) \quad (2)$$

$$x(t - \Delta t) = x(t) - \Delta t \frac{dx(t)}{dt} + \frac{1}{2!} \Delta t^2 \frac{d^2 x(t)}{dt^2} - \frac{1}{3!} \Delta t^3 \frac{d^3 x(t)}{dt^3} + \mathcal{O}(\Delta t^4) \quad (3)$$

$$\frac{d^2 x(t)}{dt^2} = \frac{x(t + \Delta t) - 2x(t) + x(t - \Delta t)}{\Delta t^2} + \mathcal{O}(\Delta t^2) \quad (4)$$

Hence, by neglecting the error of order  $\mathcal{O}(\Delta t^2)$ , it is possible to use equations (1) and (4) to derive the called "Verlet method" where position and velocity vectors for each  $i$ th atom on next step ( $t + \Delta t$ ) are calculated using previous and current step:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \vec{F}_i(t) \quad (5)$$

$$\vec{v}_i(t) = \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t} \quad (6)$$

In order to improve accuracy of the Verlet method, another approach can be made by accounting for a new force term in the velocity. This term is calculated from the updated position vector, and this improved velocity term is used to calculate the new position in the next step, creating a more precise step cycle with more stability (Satoh, 2010). This set of equations is called "Velocity Verlet method" and it is described as:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t)\Delta t + \frac{\vec{F}_i(t)}{2m_i}\Delta t^2 \quad (7)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{\vec{F}_i(t + \Delta t) + \vec{F}_i(t)}{2m_i} \Delta t \quad (8)$$

As a final variation for Verlet method one can derive the "Leap frog method", in which one considers half-step when calculating the velocity term. Even though the use of this technique leads to a more stable and accurate behaviour when compared to Verlet method, it is noticeable that velocity and position are not in the same time steps, therefore it is not possible to calculate the total energy at a given  $t$ , just kinetic or potential energy separately (Frenkel & Smit, 2001). By applying first-order derivatives with half-steps, one can obtain the following equations:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t + \Delta t/2) \Delta t \quad (9)$$

$$\vec{v}_i(t + \Delta t/2) = \vec{v}_i(t - \Delta t/2) + \frac{\vec{F}_i(t)}{m_i} \Delta t \quad (10)$$

The force term for each step is a key factor to define whether Molecular Dynamics are realistic or not. They are strictly related to the force-field adopted to describe molecular interactions, since they provide parameters to specify intra and intermolecular potentials from which forces can be calculated at any system configuration. Eventually, as seen in innovative methods such as the used on this project, the potential values can also be provided by tables generated specifically for each molecular interaction, explanations about this technique will be given on further sections (See Section 4.2.2). Molecular dynamics technique enables the use of thermostats and barostats to control dynamic behaviour of temperature and pressure of the system, recreating different types of thermodynamic ensemble. Hence, at equilibrated states, system properties can be obtained as a temporal average of instantaneous values and those averages can be compared to real experimental data in order to validate simulations.

#### 4.1.2 Monte Carlo simulation

Monte Carlo is a simulation method based on the statistical probability of existence of a system at thermodynamic equilibrium. The equilibrium condition happens when the free energy reaches a minimum by calculating the energy for each particle's microscopic state. Given a system of  $\mathcal{N}$  particles, temperature  $T$  and volume  $V$  (thus it can be denoted as a canonical ensemble), it is possible to assume that the Helmholtz free energy of this system is:

$$A = U - TS \quad (11)$$

where  $S$  is the entropy and  $U$  is the internal energy, meaning that not only a minimum in free energy can arise from a minimum in total energy, but also an increase in entropy of the system can contribute to this energy minimization. A system can be described by a group of coordinates, in this case for clarity, a group of distance vectors  $\lambda = (\vec{r}_1, \vec{r}_2, \dots, \vec{r}_{\mathcal{N}})$  from system origin. For an arbitrary  $\lambda$  the probability of a single particle of the system to statistically occupy that position is described by a probability distribution function (Satoh, 2010):

$$\rho(\lambda) = \frac{\exp\left(-\frac{U(\lambda)}{kT}\right)}{\int_V \dots \int_V \exp\left(-\frac{U(\lambda)}{kT}\right) d\vec{r}_1 d\vec{r}_2 \dots d\vec{r}_{\mathcal{N}}} \quad (12)$$

If the system configuration is generated with a considerable number of microstates that satisfies this probability, then the final configuration will have a real physical meaning. But, as soon as it is almost impossible to define an analytical solution to this equation when  $\mathcal{N}$  is too large, another method became necessary to use Monte Carlo simulations in molecular dynamics. The Metropolis method (Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) allowed the use of Monte Carlo technique by introducing the following concept: given 2 different microstates, the probability to change from state 1 to 2 is defined by:

$$P_{1 \leftrightarrow 2} = \begin{cases} 1 & \text{for } \frac{\rho(\lambda_2)}{\rho(\lambda_1)} \geq 1 \\ \frac{\rho(\lambda_2)}{\rho(\lambda_1)} & \text{for } \frac{\rho(\lambda_2)}{\rho(\lambda_1)} < 1 \end{cases} \quad (13)$$

Therefore, the integral term in eq.(12) vanishes and it is clear to observe that when  $U(\lambda_1) \geq U(\lambda_2)$  the system will certainly change to this new microstate since it has lower energy, however in the case of  $U(\lambda_1) < U(\lambda_2)$  the system has a certain probability of changing to this new microstate indicating

an increase of entropy in the system. So, by applying stochastic inputs for particle displacements and acceptance values for eq.(13), with a large number of Monte Carlo steps the system will eventually reach a minimum in free energy. Moreover, when this system reaches equilibrium it is possible to calculate accurately microstate dependent properties via ensemble averages. That means with  $n$  samples of Monte Carlo steps, the average of a  $\xi$  property can be obtained by:

$$\langle \xi \rangle = \sum_{i=1}^n \frac{\xi_i}{n} \quad (14)$$

It is important to notice that Monte Carlo simulations do not consider dynamic properties of the system such as kinetic energy, and therefore only Molecular Dynamics can account for such things (Satoh, 2010). Finally, now that both molecular simulation methods have been introduced it is possible to proceed to an overview of both simulation software used on this project. Together they can unite the benefits of both simulation methods, in order to recreate a suitable coarse-grain model for Molecular Dynamics.

## 4.2 Software Description

### 4.2.1 About GROMACS

GROMACS (**G**roningen **M**achine for **C**hemical **S**imulations) (van der Spoel et al., 2013) is a Molecular Dynamics and energy minimization software created at the University of Groningen (the Netherlands), and currently has been developed and updated by Royal Institute of Technology (Sweden) and Uppsala University (Sweden). GROMACS is a widely used tool in the branch of computational chemistry because it is flexible and efficient. Its flexibility comes from the capacity of simulating multiscale molecular models, from atomistic to mesoscale depending on topology described by the user. The efficiency originates from the computational optimisation, by this meaning not only that GROMACS can achieve impressive simulation speed in large systems using supercomputing or clusters, but also it can run small simulations in any ordinary computer with great performance.

In order to interpret interatomic interactions, that means forces between each pair of atoms, GROMACS uses input files called Topologies in which the user can describe almost any molecule using force fields or tabulated potentials. Data given in this description includes information regarding atoms' type, mass and charge. Moreover, it includes a detailed description of each bond, angle and dihedral for the molecule. There are several Force-fields available for usage depending on the desired objective of the simulation, their role is to be a database that describes interaction potentials. Hence their applicability is suitable to almost any system from all-atoms to coarse-grain descriptions, but it is entitled to the user to choose the most appropriate to obtain a realistic model.

When dealing with simulations of complex systems, realism means more computational power expended trying to achieve it (Satoh, 2010), therefore a major concern is how to achieve longer time horizons without loosing precision. A good example is the Lennard-Jones potential for Van der Waals forces, used by OPLS-AA Force-field (Jorgensen, Maxwell, & Tirado-Rives, 1996), which is the one chosen for all-atom simulations in this project. This type of potential approximation for a given atom  $i$  in relation to atom  $j$  is defined simply by two parameters  $\epsilon$  and  $\sigma$  as it can be seen in Eq.15:

$$\mathcal{U}_{LJ}(r_{ij}) = 4\epsilon_{ij} \left( \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right) \quad (15)$$

Where the forces used in Molecular Dynamics equations come from the gradient of this potential for each pair of particles:

$$\vec{F}_{ij} = \nabla \mathcal{U}_{LJ}(r_{ij}) = \left( \frac{\partial \mathcal{U}_{LJ}(r_{ij})}{\partial x} \hat{x} + \frac{\partial \mathcal{U}_{LJ}(r_{ij})}{\partial y} \hat{y} + \frac{\partial \mathcal{U}_{LJ}(r_{ij})}{\partial z} \hat{z} \right) \quad (16)$$

For every simulation time-step, GROMACS recalculates bonded potentials from topology, non-bonded potentials from the summation of Lennard-Jones potentials and Coulomb potentials. As shown in Figure (1) if  $r_{ij} \rightarrow \infty \Rightarrow \mathcal{U}_{LJ} \rightarrow 0$  and a similar behaviour occurs for Coulomb forces. Hence, GROMACS generates a neighbour list of atoms determined by a cut-off radius for each particle, by this means avoiding having to spend a lot of computational power calculating numbers with negligible value. Moreover, this feature allows the use of periodic boundary conditions that recreate a continuous media environment for simulation boxes bigger than two times the cut-off radius, since each particle cannot interact with itself.

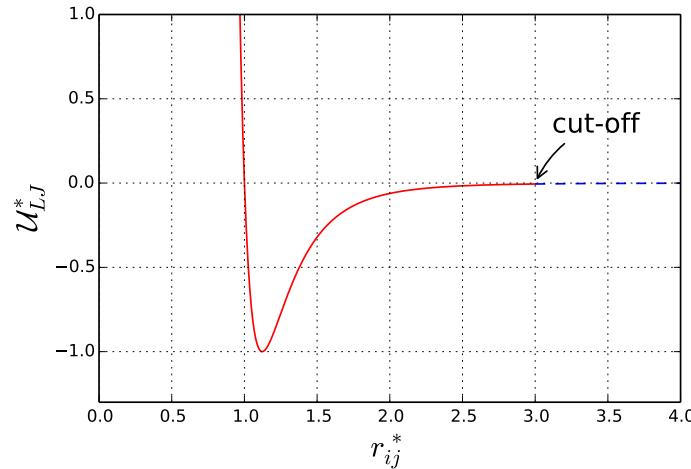


Figure 1: Example of Lennard-Jones Potential (with reduced units). For this plot  $\epsilon = 1$  and  $\sigma = 1$

Although speed efficiency is a concern, it is not the only preoccupation. In fact, it is also necessary to control environment variables such as temperature and pressure, in order to maintain the simulation within physical limits, which demands computational resources. GROMACS is capable of using three types of thermostat: Berendsen, v-rescaling and Nosé-Hoover and three types of barostat: Berendsen, Parrinello-Rahman and MTTK (van der Spoel et al., 2013). Furthermore, other computational expensive technique is the use of Particle-mesh Ewald (PME) summation method (Darden, York, & Pedersen, 1993) to deal with long-range electrostatic interactions, which is indispensable in ionic environments. All these tools makes GROMACS not only fast but also reliable for Molecular Dynamics simulation, moreover this project will as well demonstrate the capacity of this software to simulate realistic multi-scale molecular models.

#### 4.2.2 About MagiC

MagiC is a software package developed by Mirzoev and Lyubartsev (2013), which is a systematic method to generate coarse-grain models from all-atomic simulations using Iterative Boltzmann Inversion (IBI) and Inverse Monte Carlo (IMC). As described in Figure (2), this method consists of a three step process where the input is the trajectory file of an equilibrated all-atoms simulation from GROMACS, and outputs are a set of tabulated potentials and a Topology for a coarse-grained model that uses these potentials to run Molecular Dynamics simulations using GROMACS.

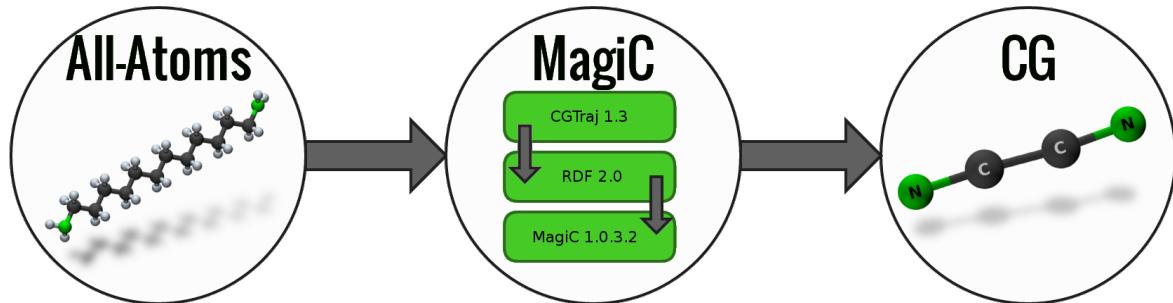


Figure 2: Graphical representation of MagiC's systematic process. The central circle represent the process itself while the other two are input and output. The green boxes indicates each step inside MagiC process.

The first step uses "CGTraj 1.3" (MagiC package utility written in Fortran 90) (Mirzoev & Lyubartsev, 2014), where the trajectory of the all-atoms simulation is recalculated based on a bead mapping described by the user. Basically, it analyses the group of atoms in each bead and assigns the position of this bead as the center of mass, charge as summation of all charges and mass as

summation of atomic masses, thus giving as output a representation of the input system as if it was an "ideal" coarse-grain model. Next, this rewritten trajectory is used as input to rdf-2.0 (MagiC package utility written in Python) (Mirzoev & Lyubartsev, 2014), where several reference Radial Distributions Functions are created for each bond, angle and pairwise intermolecular interaction, as specified by the user. These reference RDFs in conjunction with the coarse-grain topology are the inputs to the MagiC kernel (MagiC package utility written in Fortran 90) (Mirzoev & Lyubartsev, 2014), which is the key step of the process, where the final output might be a suitable coarse-grain model ready for GROMACS simulation.

The kernel is where the actual Inversion process occurs, as schematised in Figure (3). Whether using IMC or IBI, a couple of millions of Metropolis Monte Carlo steps are simulated using a set of trial potentials to create several samples of averaged thermodynamics properties. Then, these averages are applied in the equations of the chosen inversion method in order to refine the trial potentials that will be used in the next inversion step.

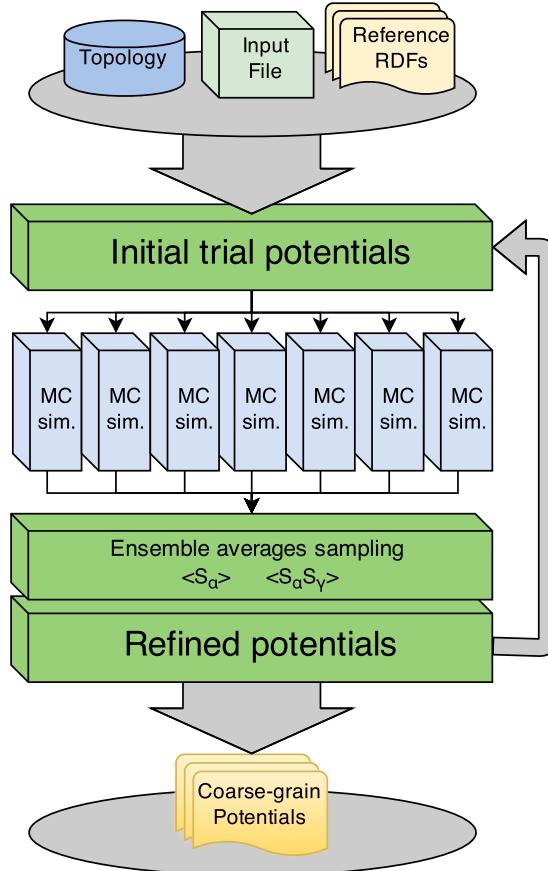


Figure 3: MagiC kernel work-flow scheme

The Iterative Boltzmann Inversion is a reasonable methodology to apply in the initial iterations, since it has fast convergence. The process derives from probability distribution function (Eq.(12)), thus for a pairwise particle interaction, the relationship between the reference RDF and Potential Mean Force (PMF) enables the use of an iterative method shown in Eq.(17) (Moore, Iacovella, & McCabe, 2014):

$$\mathcal{U}^{(i+1)}(r) = \mathcal{U}^{(i)}(r) - \eta k_B T \ln \left( \frac{\rho^{(i)}(r)}{\rho_{ref}(r)} \right) \quad (17)$$

Where  $\eta$  is a regularization parameter to avoid excessive variations. For a system in a canonical ensemble, temperature will be constant, hence the potential is corrected for each iteration using the  $\rho_{ref}(r)$  calculated from reference RDF in comparison to  $\rho^{(i)}(r)$  from RDF generated by Metropolis Monte Carlo simulation. Although this method is efficient to achieve small deviations from references, it does not guarantee that the generated potential represents one that will reproduce a behaviour similar to atomistic model because IBI does not consider cross-correlation terms between bonds and angles (Mirzoev & Lyubartsev, 2013).

Regarding the Inverse Monte Carlo process, it follows the methodology described by Lyubartsev and Laaksonen (1995). Given a Hamiltonian, that means the total energy of the system:

$$\mathcal{H}(q) = \int \mathcal{U}_\alpha \mathcal{S}_\alpha(q) d\alpha \quad (18)$$

For pair interactions, the degree of freedom  $q$  in Eq.(18) becomes the distance  $r$  between a pair of atoms. Therefore, for a real system the Hamiltonian would be the summation of an infinite set of  $\mathcal{U}_\alpha$ . However, a good approximation for numerical methods is to assume a cut-off radius  $r_{\text{cut}}$ , where potentials no longer have relevant value after this point. One can define a  $\Delta r = \frac{r_{\text{cut}}}{M}$ , obtaining  $M$  subdivisions and thus a finite set of  $\mathcal{U}_\alpha$  in which  $\mathcal{S}_\alpha$  is the number of pairs within interval  $[r_\alpha, (r_\alpha + \Delta r)]$ . Since  $\mathcal{S}_\alpha$  is function of  $\mathcal{U}$ , one can derive the following Taylor expansion:

$$\Delta \langle \mathcal{S}_\alpha \rangle = \sum_{\phi=\alpha}^M \frac{\partial \langle \mathcal{S}_\alpha \rangle}{\partial \mathcal{U}_\phi} \Delta \mathcal{U}_\phi + \mathcal{O}(\Delta \mathcal{U}_\phi^2) \quad (19)$$

$$\frac{\partial \langle \mathcal{S}_\alpha \rangle}{\partial \mathcal{U}_\phi} = \frac{\langle \mathcal{S}_\alpha \rangle \langle \mathcal{S}_\phi \rangle - \langle \mathcal{S}_\alpha \mathcal{S}_\phi \rangle}{k_B T} \quad (20)$$

For each inversion step  $k$ , Monte Carlo simulations are run using a set of potentials  $\mathcal{U}_\alpha^{(k)}$ , then as explained previously it is possible to calculate ensemble averages of the cross-correlation terms  $\langle \mathcal{S}_\alpha \mathcal{S}_\phi \rangle^{(k)}$  and averages  $\langle \mathcal{S}_\alpha \rangle^{(k)}$ ,  $\langle \mathcal{S}_\phi \rangle^{(k)}$ . At this point, the reference RDFs become useful for inverse Monte Carlo. Since this method deals with pairwise interactions the values of  $\mathcal{S}_\alpha$  and  $\rho(r)$  are directly proportional following the rule  $\mathcal{S}_\alpha^* = \rho(r_\alpha) \Delta r$  (Mirzoev & Lyubartsev, 2013) because both measure the quantity of one type of coarse-grain bead in relation to another. Hence, by using  $\Delta \langle \mathcal{S}_\alpha \rangle = \langle \mathcal{S}_\alpha \rangle^{(k)} - \mathcal{S}_\alpha^*$  in conjunction with Eq.(19) and Eq.(20) and neglecting the error  $\mathcal{O}(\Delta \mathcal{U}^2)$  the potential correction for next step  $\Delta \mathcal{U}_\alpha^{(k)}$  can be obtained and finally potentials are updated using the following equation:

$$\mathcal{U}_\alpha^{(k+1)} = \mathcal{U}_\alpha^{(k)} + \Delta \mathcal{U}_\alpha^{(k)} \quad (21)$$

All of the previous methods described rely on a considerably large amount of Monte Carlo simulation steps. That is due to the fact that the closer the set of potentials gets to solution, the harder it becomes to distinguish between statistical noise and a suitable potential correction. For this reason, as it can be seen in Figure (3), MagiC has a built-in capacity of running parallel Monte Carlo simulation, which enables the user to reach the order of billions of simulation steps in a feasible timespan. Moreover, there is not a unique solution for a given RDF (Moore et al., 2014) and eventually the solution can possible diverge at a certain point. Therefore, a satisfactory result depends on analysis of all outputs from inversion steps in order to guarantee convergence. Further explanations regarding functionalities and parameters for MagiC and GROMACS will be provided in the next section.

## 5 Methods Description

The main objectives of the experiments of this project can be characterized in three stages: the first is to develop and simulate a suitable model for the silica oligomers and the surfactant in an atomistic scale considering a high pH system. Secondly, the aim is to use MagiC software package to develop potentials for the coarse-grain model proposed using a variety of techniques to improve efficiency of the process. Lastly, the focus will be the evaluation and validation of the self-assembly structures simulated using the coarse-grain models that were generated from the previous experiments.

### 5.1 Atomistic Molecular Model

A initial concern for the atomistic simulations was the design of the water, surfactant and silica molecules. Beginning with the water molecule, the model used was the TIP4P water (Figure (4a)) created by (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983) which is a widely used water model that uses OPLS-AA forcefield (Jorgensen et al., 1996) parameters. For the DMDD surfactant, a model was created with the aid of co-workers in OPLS-AA forcefield which increases the reliability of the atomic interactions with the water models, since they use the same forcefield. This surfactant model includes 3 variations as a result of the high pH environment, because as described by (?, ?) the hydrophilic Nitrogen head receives an extra hydrogen thus in a solution it may be found the Neutral DMDD molecule (DMDDn), the Single charged DMDD molecule (DMDDs) and the Double charged

DMDD molecule (DMDDd). Since atomistic simulations need to be preferably set in a neutral charge environment, an inert anionic Bromide counter-ion was selected to equilibrate the charges.

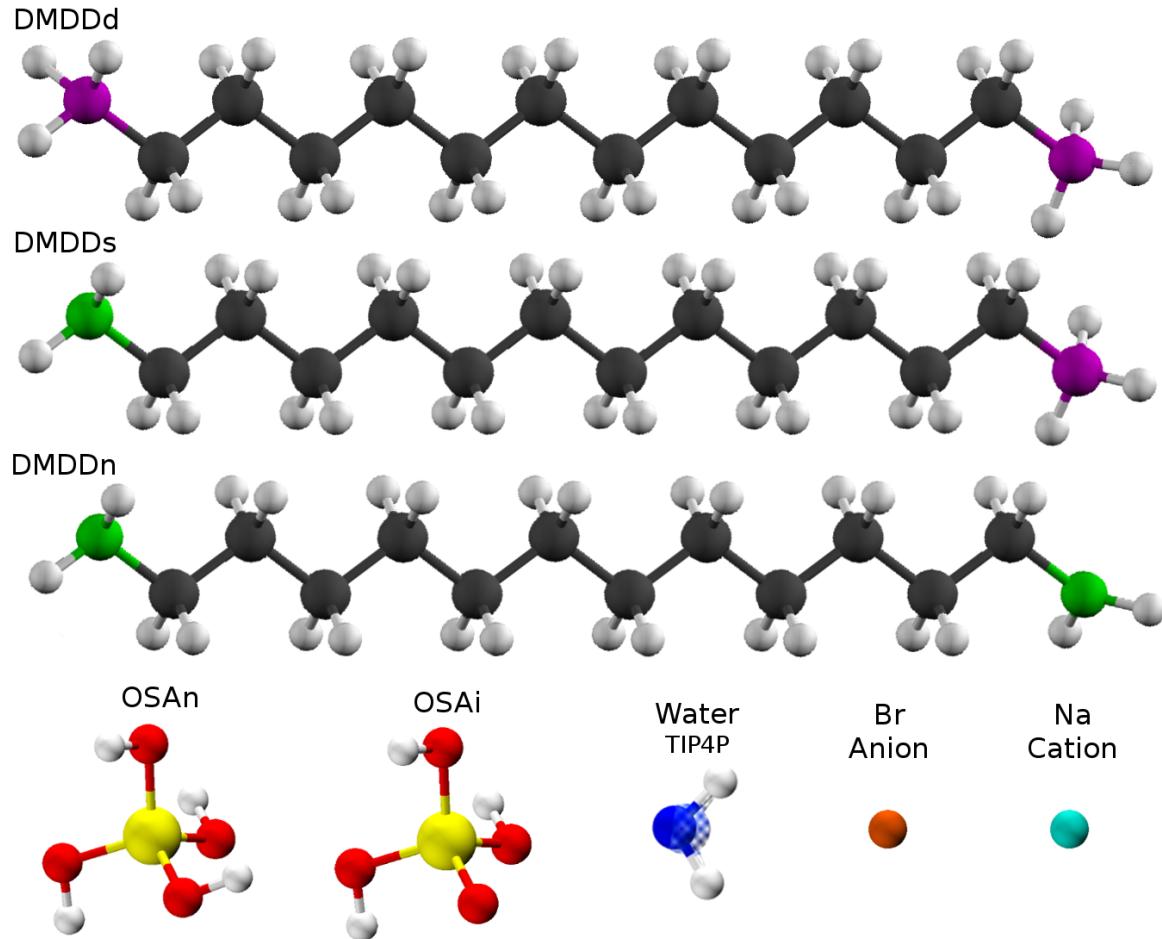


Figure 4: Graphical representation of the Atomistic models.

When it comes to the description of the silica model, it is important to notice that the in solution substances should be treated differently from pure liquid silica precursor. As described in (Tanev et al., 1997) the silica precursor tetraethoxysilane (TEOS =  $Si(OCH_2CH_3)_4$ ) is mixed in solution in a ratio of 1:2 in relation to the DMDD surfactant, that means in an atomistic level, one silica to each surfactant head. However, due to high pH condition a substitution process that might occur in TEOS where it is forced to replace its Ethanol ramifications by hydrogen atoms forming an temporary unstable substance, the orthosilicic acid (OSA). As stated by (Pérez-Sánchez et al., 2013) this unstable silica oligomer is the major representative of the silica portion in the system at early moments in solution state, which in later stages may interact with surfactant structures and polymerise around them in a variety of configurations, turning into a nanoporous and highly organized silica structure. For the scope of this project only the orthosilicic acid model will be considered, since the simulations will take place during the formation of the surfactant self-assemble structure which occurs in the early moments of the solution stabilization. Therefore, as it can be seen in Figure (4c) it is possible to consider two principal representations of the orthosilicic acid: the neutral molecule (OSAn) and the ionized molecule (OSAi). Similarly to Bromide, cationic Sodium counter-ion will be used to equilibrate the system charge.

## 5.2 Atomistic Simulation

Regarding the composition of each simulation box, the main idea was to work upon the pH value of the solution. The fact of having multiple molecular models allows the creation of simulation boxes with a varying pH, in which each one represent a different stage of deprotonation, that means the

higher the number of molecules with missing hydrogen the higher the pH that is represented on the system. It is difficult to determine exactly what is the number of molecules of each type to reach an specific pH, since the molecular types which are deprotonated should vary dynamically (Wallace & Shen, 2012). This means that even when the ratio of deprotonated and protonated molecules is in an adequate value to represent a certain pH, it might not be possible to guarantee that an specific molecule will remain at the same level of deprotonation at all times. If a molecule stays at the same state during the whole simulation it will be not representing the desired pH in the system, instead it could create an ambiguity where different regions of the box may represent a different pH value.

There are several techniques such as represented in work done by (Lee, Salsbury, & Brooks, 2004) which are capable to recreate the dynamic behaviour of the pH in solution, which probably can aid to obtain a more accurate results in when it comes to the self-assembled structures in an atomistic level. However, the adoption of such methods goes far beyond the scope of this project's main objectives, and instead in order to maintain the system with a less complex structure only high pH values were represented in simulation boxes. The types of boxes that were chosen used a single state of the molecule to represent a pseudo-high pH solution, because most of them are extreme cases that may not exist in reality but they are good approximations of real systems.

In previous works, the main subject of study was the DMDDn model, because even though it represent the most extreme case in which all polar heads are deprotonated, indicating a very high pH, it still a feasible approach to real high pH solution and furthermore provides the advantage of not presenting coulomb forces in the coarse-grain model. However, when the silica precursor is introduced in the system it might be unrealistic to assume that the OSA model will ionize all its hydrogens, since it only polymerizes in further stages after the interaction with the surfactant. Hence, the adopted model was the OSAi molecular type that when used together with DMDDn still a system with elevated pH but slightly reduced when compared to pure surfactant solution, since it has non-ionized hydrogens. For this kind of systems two approaches were adopted in order to recreate the simulation boxes. The first method was to randomly insert all molecules inside the box at adequate ratios. The second was to use simulation boxes that were generated based in DMDDn simulations that already had been equilibrated into a layer just by removing the water and putting the right amount of OSAi and then resolvating the box. The specific number of molecules of each type is described in Table (1), furthermore the implications concerning the use of both methods of box creation will be discussed in Section (6.1).

The second molecular model chosen to represent the surfactant was DMDDs. It was chosen because not only it represents a mid-state between protonation and deprotonation in the polar heads, but also it presents differentiated behaviour when compared to the previous model due to coulombic interactions within the assembled layer. Moreover, the DMDDs simulation boxes can be considered as slightly elevated pH systems when used in conjunction with ionic silica model, since it has just only one ionized head extra per surfactant. In the case of DMDDs, two types of systems were studied: one containing only DMDDs surfactant with Bromide counter-ions to equilibrate the charges, and another with DMDDs and OSAi using Sodium as counter-ion. All boxes were generated using by random placement of the molecules, to ensure whether or not it occurs layer self-assembly in both cases, and as the same information about the composition is shown in Table (1).

Table 1: Description of all-atom simulation boxes

Box	Surfactant Type	no.	Silica Type	no.	Ion Type	no.	Water no.	Avg. Size (Å)	Concentration (mM)
1 <sup>a</sup>	DMDDn	45	-	-	-	-	240	28.3	3.3
2 <sup>b</sup>	DMDDn	169	-	-	-	-	905	43.9	3.3
3 <sup>a</sup>	DMDDn	30	OSAi	60	Na <sup>+</sup>	60	343	30.3	1.8
4 <sup>b</sup>	DMDDn	100	OSAi	200	Na <sup>+</sup>	200	520	42.2	2.2
5 <sup>a</sup>	DMDDs	45	-	-	Br <sup>-</sup>	45	475	31.7	2.3
6 <sup>b</sup>	DMDDs	100	-	-	Br <sup>-</sup>	100	1282	42.6	2.1

<sup>a</sup> Short-range box.

<sup>b</sup> Long-range box.

It is noticeable that most atomistic boxes have a different concentration. This fact derives from the idea of providing flexibility to the coarse grained model, because they will be used as references to the coarse-grain model. The analysis of the results from previous experiences (?, ?), it is possible to observe that the coarse-grain model can represent and adapt to different apparent concentrations, this fact also means that the coarse-grain potentials obtained from these atomistic simulation boxes

contains a certain degree of similarity between them, at least when it refers to the position of the peaks and valleys. Therefore, having multiple similar concentrations within a short range can only benefit the final result obtained since flexibility is a desirable property. Another important point highlight is the box size. For all cases we have two ranges of size: the first is a small box with side bigger than the DMDD molecule (about 16 Å) used to obtain initial short range potentials; and a big size box with side at least over 40 Å to obtain long range potentials. A more detailed explanation about the difference of long range and short range potentials will be given in Section (6.2)

The simulation boxes generated for all experiments were created using "genbox" (GROMACS package utility) following a standard procedure in order to ensure the consistency of the experiments, moreover all the atomistic simulations that were used as reference for coarse-graining process were run in NPT ensemble. In order to reach this NPT condition, all boxes received an energy minimization step in order to avoid system blow-up due to possible extreme potential energy spots created during box generation. That is because during random displacement of molecules some of them could overlap others and then cause a destabilization of the system. Afterwards, an equilibration step was necessary to reach the desired initial temperature and pressure conditions for the MD simulation. During equilibration, temperature coupling was kept at 298 K using a v-rescaling thermostat (Bussi, Donadio, & Parrinello, 2007) with time constant of 0.01 ps and pressure coupling was kept at 1 bar using a Berendsen barostat (Berendsen, Postma, van Gunsteren, DiNola, & Haak, 1984) with time constant of 0.5 ps. Every equilibration simulation was run for 200 ps with time step of 0.5 fs using leap-frog algorithm.

When it comes to the final simulation run, even though it have different structural parameters for each box content, all of them followed the same ensemble with temperature kept at 298 K by a Nosé-Hoover thermostat (Nosé, 1984) (Hoover, 1985) and pressure kept at 1 bar by a Parrinello-Rahman barostat (Parrinello & Rahman, 1981). Moreover, simulation time was also standardized. Every simulation lasted 150 ns with a time step of 1 fs using leap-frog algorithm. But for the sake of experiments using MagiC, the first 50 ns of the simulations were disregarded since during this time the system was still in equilibration, and therefore that would affect reference RDF values.

### 5.3 Coarse-grained molecular Model

Regarding the coarse-graining process of each system described previously, it is possible to say that much effort was put in transforming the application of MagiC software technique in the most systematic way as possible. Base on the analysis done in the previous works, it was supposed that the use of even more simplified coarse-grain model for the surfactant should be beneficial for the whole MagiC process, in the case the adoption of 3 bead model reduced the complexity in such a manner that it only includes bond interactions and angular interactions, which is good since the MagiC process do not account for 1-4 bead interactions. As it can be seen in Figure (5,6) the 3 bead model includes 3 types of bead only: the neutral Polar head (Nn), the ionic Polar head (Ni) and the Apolar body (Cn). Therefore, combinations of them creates the 2 different types of DMDD coarse-grained molecule used in the following experiments. Moreover, for the silica molecule and both counter-ions a single bead model was accepted since reduces drastically the number of intermolecular interactions. Finally, it is important do highlight that all coarse-grain models adopted in these experiments assume implicit solvent condition, fact that reduces the complexity of the system when applying magic method and increases simulation speed.

During the usage of MagiC package, the model description begun at the "CGtraj" step, where not only the bead types for the model were described, but also which atoms would be inside of each bead. These informations must be part of input file for "CGtraj", following the strict guidelines of MagiC Manual (Mirzoev & Lyubartsev, 2014). In addition, it was at this moment that occurred the determination of implicit water condition, therefore during the recalculation of trajectories the water molecules were removed from the box. At the second step is where bonds, angles, and interatomic interactions were described in the input file for the "rdf-2.0" utility, following as well MagiC Manual guidelines (Mirzoev & Lyubartsev, 2014). That is because at this point MagiC generated reference RDF's used in the Inversion process for each potential table. Additional information regarding model description in input files can be found in Appendix (see Section 10).

Once reference RDFs and a Topology were specified, the inversion process started. The parameters of each inversion step were based on work done previously (? , ?), where multiple combinations of inputs were tested. Nevertheless, considering that this model owns a less complex molecular structure, the number of Monte Carlo simulation steps were adapted to fit this new model. In order to increase the convergence speed, an improvement was applied to the Monte Carlo simulations in which instead of beginning from a random state, the initial configuration was extracted from snapshots of pre-assembled

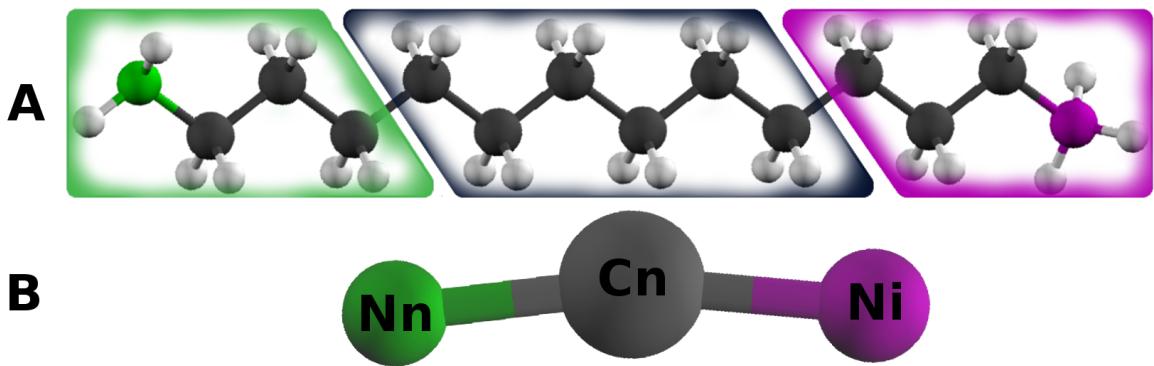


Figure 5: A) Schematic of DMDDs molecule split plan, Carbon atoms represented in black, Neutral Nitrogen in green, Charged Nitrogen in purple and Hydrogen in gray. B)Final coarse-grain model for DMDDs. This model aggregated the three most polar heavy atoms on the tips and six apolar heavy atoms at the center.

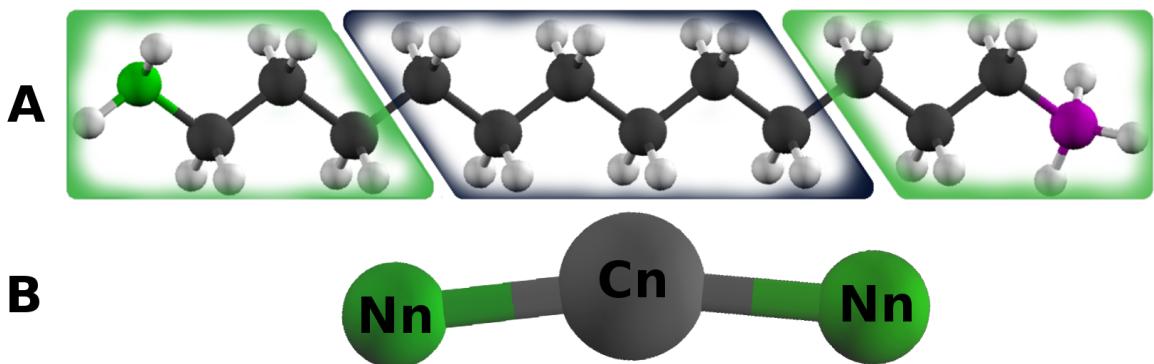


Figure 6: A) Schematic of DMDDn molecule split plan, Carbon atoms represented in black, Neutral Nitrogen in green and Hydrogen in gray. B)Final coarse-grain model for DMDDn. This model aggregated the three most polar heavy atoms on the tips and six apolar heavy atoms at the center.

layers obtained from "CGtraj" step. Furthermore, when dealing with inversion process of the long-range boxes, the initial potentials were estimated using the potential obtained from the short-range boxes, avoiding complete divergence in the early stages of inversion process. With the aid of these two techniques it became possible to standardize the whole inversion process as it can be seen in Table (2). The simulations were run in parallel using up to 20 cores. Table (2) exposes the specific number of steps per core and whether it used IBI or IMC method for inversion for each type of box type. The ensemble averages where calculated every 1000 Monte Carlo steps, after equilibration. Since the water is implicit, the dielectric constant for Monte Carlo simulations was set to the experimental dielectric constant of water at 298 K and 1 bar which is  $\varepsilon = 78$  (Fernández, Mulev, Goodwin, & Sengers, 1995).

After each inversion step MagiC generates RDF based on ensemble averages calculated during Monte Carlo simulations and then it calculates the deviation between reference RDFs and inversion RDFs. When the deviation reached the desired convergence value, that means in the order of  $\mathcal{O}(\Delta 10^3)$ , the potentials and topology were exported to GROMACS format. Depending on the type of system, each simulation using GROMACS had different input parameters when it comes to the timestep, cut-off range and electrostatic cut-off range and type. further informations about which values where chosen and what were the implications of this modified parameter will be given in Section (6.3). As a general rule, the systems had been up-scaled maintaining concentration of the original atomistic reference simply by following a formula, in which the size of the box's edge was determined by rescaling average

Table 2: Inversion Process description for short-range and long-range

S-R	$N_{Inv}$	10	10	10	10
	Method	IBI	IMC	IMC	IMC
	$N_{Cores}$	20	20	20	20
	$N_{MCe}$	0.5	1	2.5	10
	$N_{MC}$	2	3	10	30
L-R	$N_{Inv}$	10	10	10	10
	Method	(No IBI)	IMC	IMC	IMC
	$N_{Cores}$	Initial trial	20	20	20
	$N_{MCe}$	potentials from S-R	1.5~3	5~6	7.5~10
	$N_{MC}$		5~10	20	30
				10~12.5	$\sim N_{MC}^{(i)}/4$
				50	$\sim N_{MC}^{(i-1)} + 20$

<sup>a</sup>  $N_{MCe}$  and  $N_{MC}$  in millions of steps.

<sup>b</sup> Only reference box needed over 4 iterations.

box size from Table (1) using the following equation:

$$L_{CG} = \sqrt[3]{\frac{N_{CG} L_{AA}^3}{N_{AA}}} \quad (22)$$

where  $N$  is number of molecule and  $L$  is the edge side. Therefore, every coarse-grain simulation has almost the exact same concentration as their original reference. Regarding the simulation time of coarse-grain simulations, it depends directly on how long it took the system to self-assemble in a constant structure, then as soon as this structure remains stable for considerable time when compared to the equilibration time the simulations where stopped. Seldom it was possible to set a fixed time horizon, thus the simulation time for each coarse-grain simulation will be given for each box in the image description.

## 6 Results and Discussion

### 6.1 Atomistic Simulations

The first point of interest for this project are the structures observed in the atomistic simulations. Since there are three distinct types of systems under analysis, the main objective is to compare the self-assemble structures in all cases, making an qualitative evaluation by observing the simulation boxes and as well as a quantitative evaluation by comparing the RDFs of each system. However, before entering in the analysis itself it is important to discuss the reason for the creation of each simulation box. The first highlight topic is the two different categories of simulation box the short ranged and the long ranged. Theoretically, the only box type necessary to derive potentials for coarse-grain models using MagiC is the one with the biggest box size, because it will include considerably more information regarding the self-assembled structures. Nevertheless, it is observable that the larger the reference system the harder it becomes to achieve a great convergence value using MagiC methodology, because it requires a great number of Monte Carlo simulations. In complex systems, a high number of Monte Carlo steps in the early stages of inversion using Boltzmann Inversion rarely produces initial potentials that will result in convergent behaviour during Inverse Monte Carlo procedure. For this reason the use of Short-range references box was introduced, since simpler systems can reach convergence very fast and the final potentials generated can be used as an approximation for initial potentials for the Long-range boxes.

As it is possible to see in Figure (??), the first type of box that contains only DMDDn is an attempt to study the layer formation characteristic of this type of surfactant, in which the concentration was fixed for both, but the Short-range box contains a single periodic layer while the Long-range contains a double-layer periodic structure. Another characteristic of the system is the Radial Distribution Function. The RDF of N-N interactions are a great representatives of the system because the N atoms, which are the polar heads, can provide informations about the position of each molecule in relation their neighbours. In this case, as it can be seen in Figure (??), the RDFs of both boxes are similar since the aggregation phenomena inquires that proximity between polar and apolar parts, remains at a distance where free energy is the lowest. The different values to the peaks, does not means that there are more N heads within a given radius. This value varies according to the bulk density of the box

because the RDF values are normalized, that means a higher concentration implies in a higher number of atoms within a given range therefore, a lower normalized peak. The only difference appears in the second peak due to the proximity of the double-layers, and this difference may impact the MagiC process since it is based in the reference RDF.

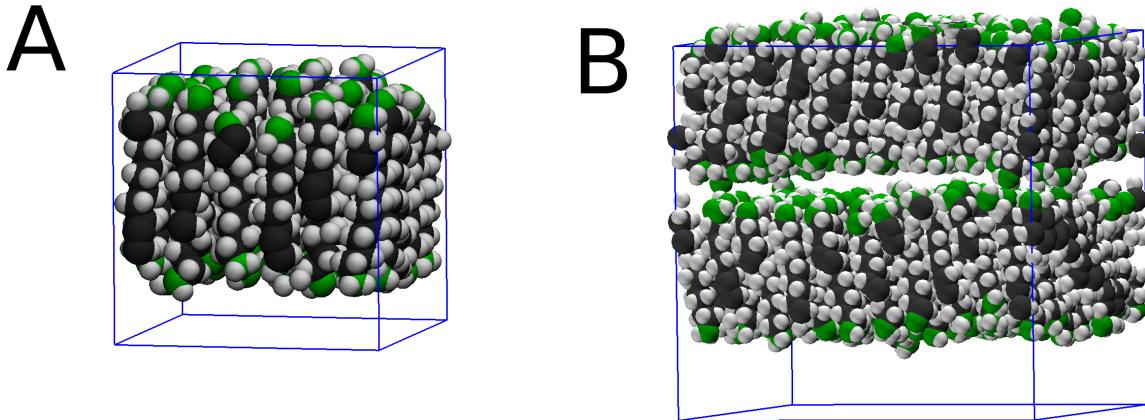


Figure 7: Work in Progress

The other type of system is the DMDDs with Bromide counter-ions which introduced two new features to the reference system: the addition of electrostatic charges and the use of multiple molecular types. As demonstrated in Figure (??), the DMDDs self organize in a slightly different structure when compared to the DMDDn, presenting alternated head position in which the charged polar head is more exposed to the ion-solvent solution and the non-charged head right below alternation positions. Comparing the RDFs between Ni-Ni (Charged) heads and Nn-Nn (Neutral) heads in Figure (??), it is possible to see that the distance is between heads is similar in both cases. Furthermore, it is possible to see that this type of layer preferentially remains positioned diagonally in the periodic box. Implications of the coarse-graining procedure of this type of system will be more detailed in further sections.

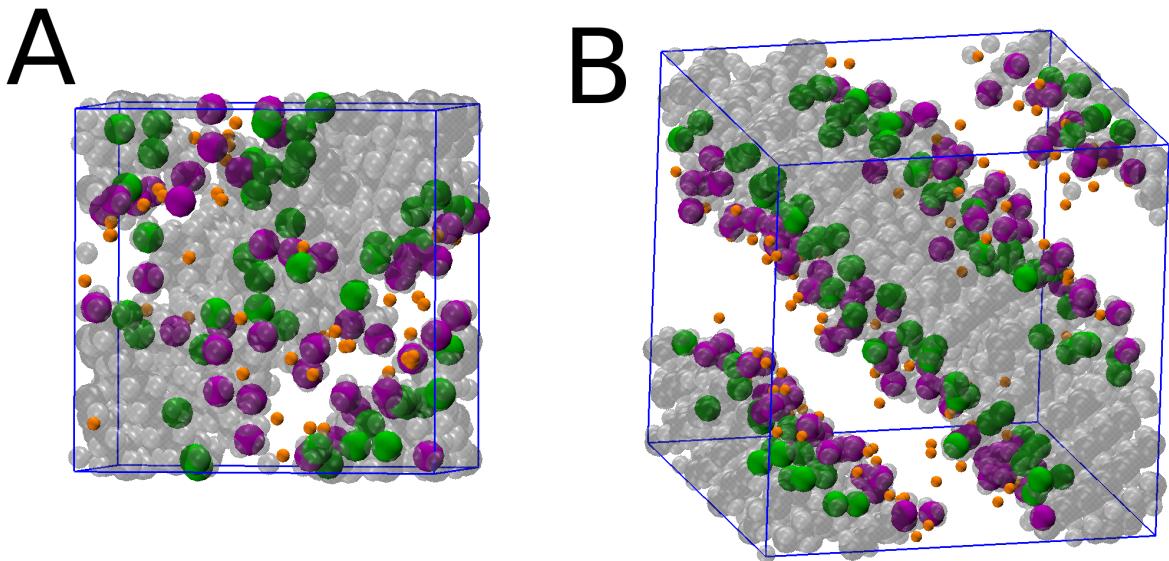


Figure 8: Work in Progress

The last type of systems under analysis are the ones with OSAi, DMDDn and Sodium counter-ion, since they are much more complex systems the simulation boxes took longer to equilibrate in a stable

configuration. According to Tanev et al. (1997) the expected structure in this case is small surfactant layers surrounded by polymerized silica, however as soon as there is no silica oligomers in the model chosen for simulation it is expected to observe a similar but not identical behaviour. As it can be seen in Figure (??), the layer formation phenomenon still observable, but without a regular shape and periodicity that occurs when there is only DMDDn surfactant. Another important factor that becomes clear in Figure (??) when measuring the RDF for Si-Si and Si-Na atoms, it that it seems the silica phase self-organizes in a rigid structure similar to ion-ion crystal, which may be due to the fact that Sodium interacts too strongly with ionic silica and the use of a more inert counter-ion such as tetramethylammonium (TMA) would be more suitable (Pérez-Sánchez et al., 2013).

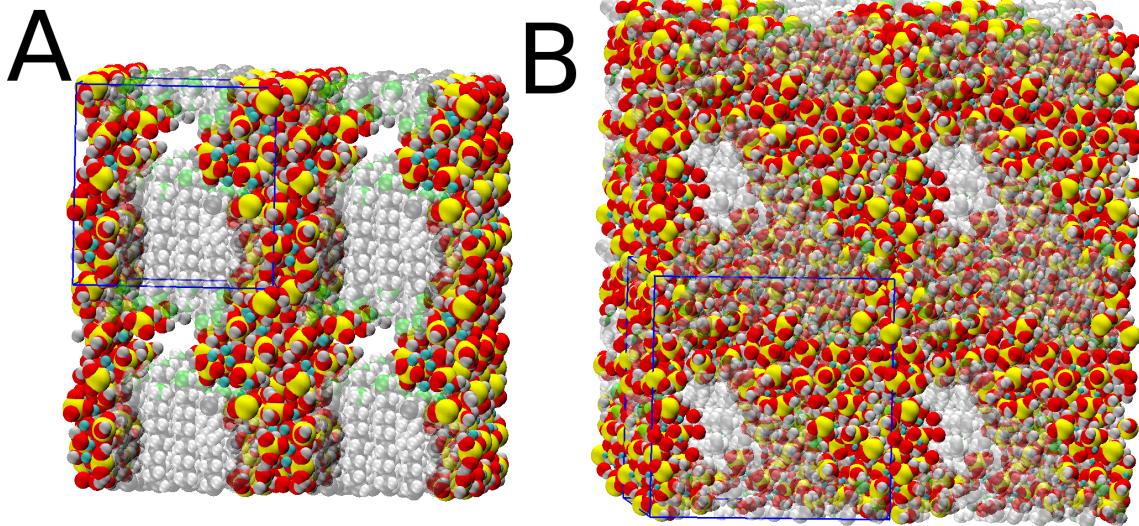


Figure 9: Work in Progress

## 6.2 Coarse-Graining Process: Convergence Analysis

Concerning the coarse-grain process using MagiC, an important factor focused during the whole process, with all the reference atomistic simulations was that the methodology should be maintained as systematic as possible, that means making the less number of alterations in the input files during inversion process. Therefore, this approach is also an attempt to validate whether it is possible to generalize any type of system using one single methodology or if it is necessary to treat each of them specially. The introduction of the usage of pre-equilibrated structures as initial configurations for the Monte Carlo simulations reflected in a new behaviour for the convergence curves in all references when comparing to previous studies, it not only diminished the number of simulation steps required during inversion process but also reduced the number of iteration cycles to achieve the desired deviance value. The Coarse-grain method adopted in this project rely deeply on visual and numerical analysis of the convergence during Inversion process. For all experiments each inversion step was validated by comparing it with previous steps and with the behaviour observed in atomistic simulations. There is not any guarantee that the set of potentials generated after each step will necessarily reproduce a physically meaningful system. That is because these potentials are made to reproduce the RDF obtained after inversion, and since the configuration of the system is a probabilistic event there is more than one way to represent a given RDF. Therefore, the smaller the deviation from reference RDFs, the greater the odds that the potential set will reproduce the desired behaviour with acceptable reliability.

The following graphs represent the whole inversion process for each reference system determined. They show the potential obtained and the RDF generated by this set of potentials after each inversion step. When comparing the short-range and the long-range cases, it is possible to visualize clearly the progression for both case. Moreover the new techniques adopted as described previously resulted in a faster progression towards the convergence region when comparing to the inversion process made without them. As specified in Table (2) after 10 inversion steps the number of sampled Monte Carlo simulations was increased, therefore the deviance curves have this cyclic oscillatory behaviour every 10 steps. Nevertheless, in all cases even if it slightly diverges inside each 10 step cycle, normally they show

a convergent behaviour, where the potential graphs were confined within a certain convergence region. The plots does not seem to show a significant difference between the short-range models and the long-range models, fact that is a good indicative that the methodology adopted may result in a more predictable behaviour for the convergence curves. However, there where some difficulties to obtain the Long-range potentials for the Box 2 with the pure DMDD type. When it comes to limitations of the techniques adopted it is possible to say that the imported potentials from short-ranges might need a more gradual step, because the big gap between the complexity of the Short-range case in this system type resulted in the same behaviour as like beginning from a raw potentials in the Long-range box, in which later proceeded to the same complete divergence situation.

### 6.2.1 Pure DMDDn and water

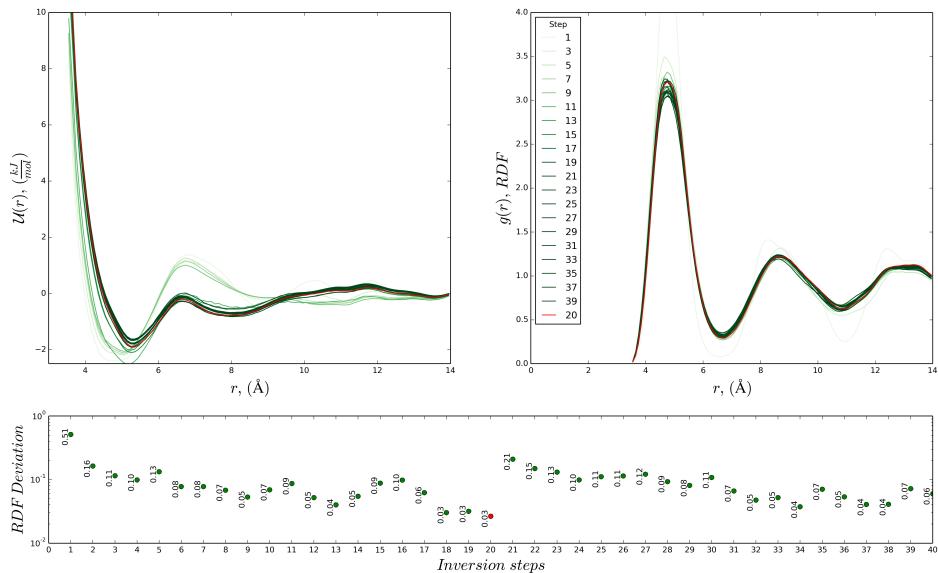


Figure 10: Graphical representation of the inversion process for Short-Range Box 1. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing N-N interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

### 6.2.2 DMDDs and Bromide

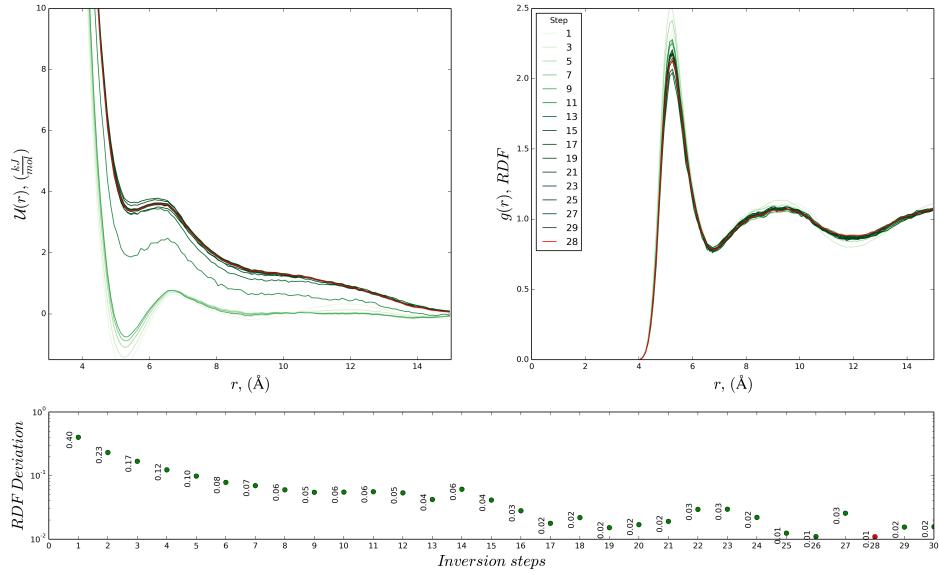


Figure 11: Graphical representation of the inversion process for Short-Range Box 5. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing Ni-Ni interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

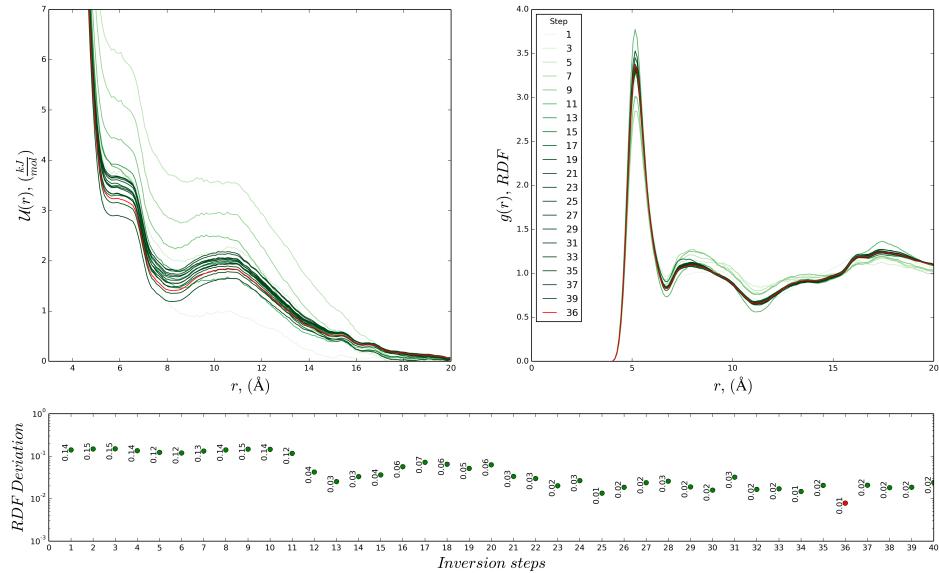


Figure 12: Graphical representation of the inversion process for Long-Range Box 6. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing Ni-Ni interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

### 6.2.3 Pure DMDDs and Bromide (Integrated Potentials)

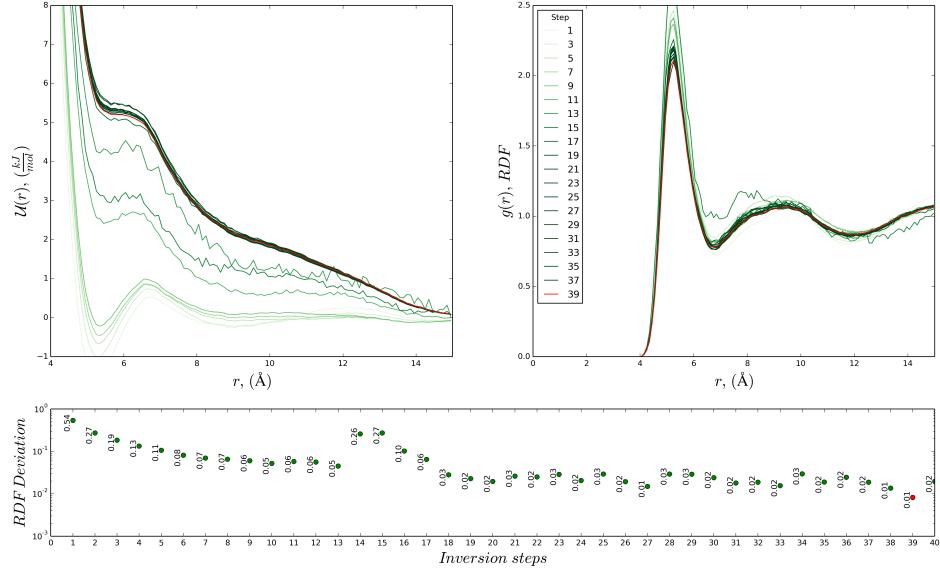


Figure 13: Graphical representation of the inversion process for Short-Range Box 5. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing Ni-Ni interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

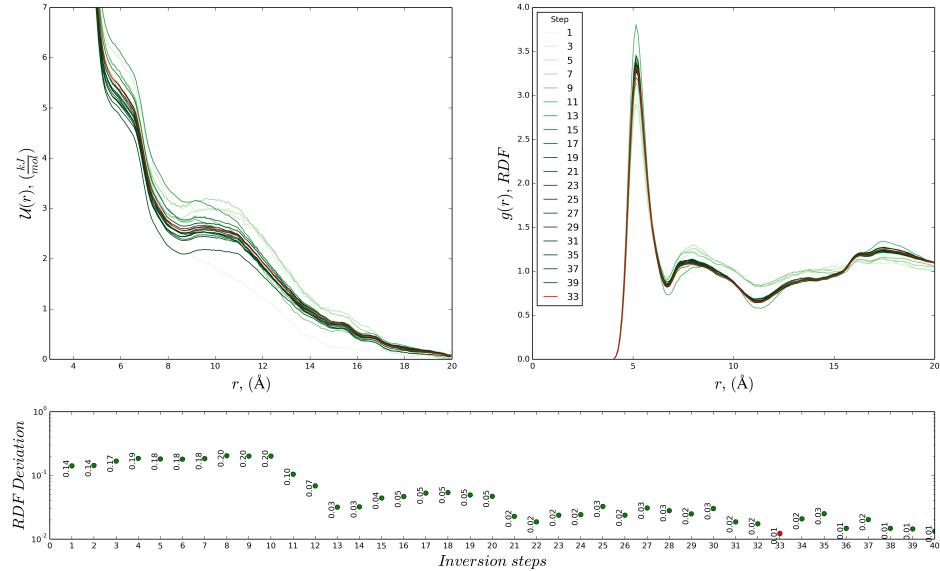


Figure 14: Graphical representation of the inversion process for Long-Range Box 6. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing Ni-Ni interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

### 6.2.4 DMDDn and Silica

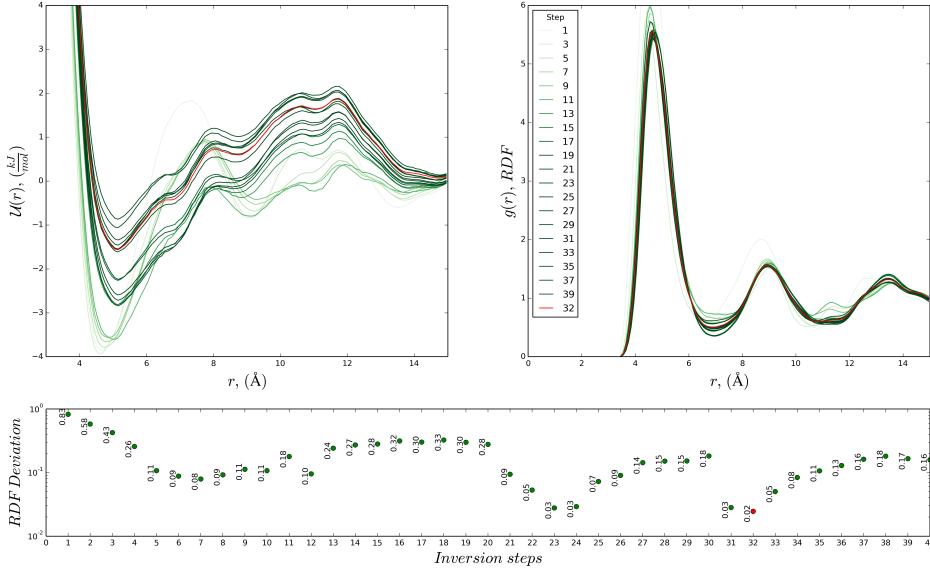


Figure 15: Graphical representation of the inversion process for Short-Range Box 3. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing N-N interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

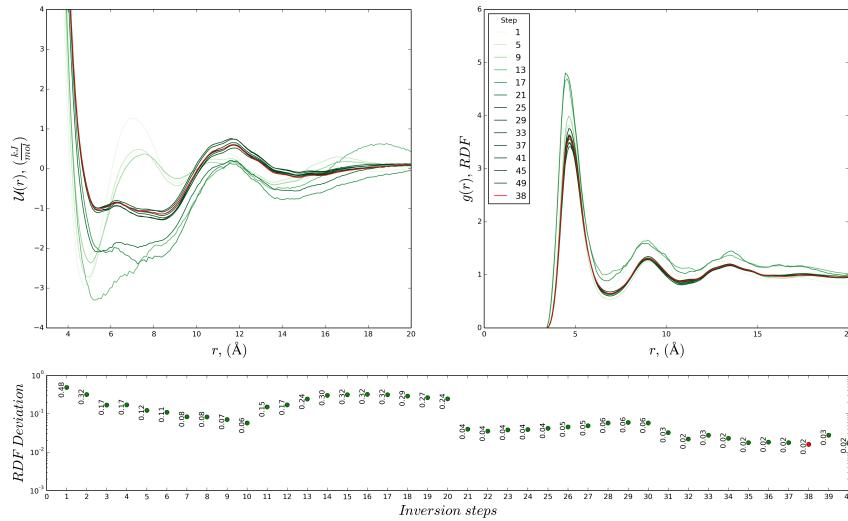


Figure 16: Graphical representation of the inversion process for Long-Range Box 4. Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing N-N interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. The Red line indicate the final potential chosen. Graph 3 (bottom) shows the root mean square deviation between reference RDF and RDF of a given step (for all interactions). Notice that this graph is using logarithmic scale.

## 6.3 Charged Systems

The efficiency of charged systems by using MagiC coarse-grain technique will be evaluated in this section. It has been seen that this technique is capable of reproducing non-charged atomistic reference

systems with a high degree of fidelity, therefore with the introduction of charges in the system it is expected to obtain reasonable results as well. That is because MagiC is capable of account for electrostatic forces and differentiate it from other kinds of forces when deriving the potentials for each bead type. In order to verify this capability, as previously described, it was proposed 2 ways of obtaining potentials to this type of coarse-grain model. The first way is simply assume a charged system, where the beads itself will have a real charge and the electrostatic force will be not included in the potentials. Hence, when using GROMACS to simulate using these potentials, the electrostatic forces will be calculated by coulomb equations or PME electrostatics. The second approach planned was to consider the electrostatic interaction merged together with all other types of forces in the same potential derived by MagiC methodology. However in this case, all beads will be considered as non-charged particles in GROMACS simulations avoiding the necessity of calculate electrostatic forces which can increase simulation speed drastically. It is clear that the most realistic scenario is represented by the first option proposed, nevertheless since coarse-graining is all about increasing simulation speed and scale the second option should not be promptly discarded.

A major issue expected when dealing with charged systems was during the coarse-grain process, since it was necessary to differentiate between charged and non-charged bead types, the complexity of the systems increased drastically. Therefore, this fact encouraged the creation of the techniques mentioned in previous section. Moreover, there was not too much discrepancy between the time expended to obtain potentials in both scenarios proposed, thus the adoption of charges during MagiC methodology was not a factor that increased the complexity of the system by itself.

Now turning to the analysis of potentials obtained for Long-Range systems, Figure (17) reveals interesting features regarding the interaction with charged coarse-grain beads. The two different types of polar heads of the molecule show the 3 possibilities for interactions: neutral-neutral, charged-neutral and charged-charged. Both non-charged interactions have the peculiarity of presenting regions with negative potential where there are predominant regions of low energy that maintain the distance between beads within a specific range as like a customized Lennard-Jones potential. Contrastingly, the interaction between charged beads show a high energy potential with mostly repulsive forces as the Ni-Ni incorporated potential shows, which is an expected behaviour for electrostatic forces. When considering electrostatic forces separated from the others, it is expected to see a potential similar to the neutral heads where the Van der Waals forces are predominant over solvent interaction forces. But instead, the graph shows new scenario where the potential remains with high energy and mostly repulsive forces and in addition it is possible to observe also lower energy pits in almost the same region as the Nn-Nn pits, indicating that the beads have a preferred structure even with high energy.

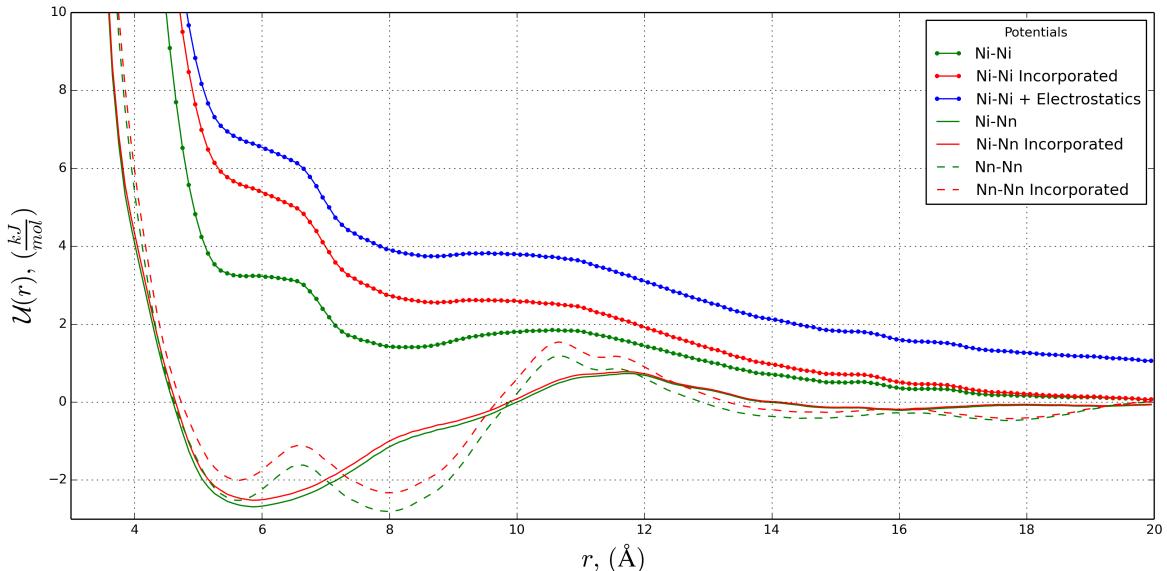


Figure 17: Comparison between potentials for different types Polar head. The color shown in the label indicates which type of electrostatic methodology was applied, in green considering charges separately and in red all potentials incorporated as a whole and unique potential. The blue line is just a theoretical curve representing a summation of the green line and a calculated Coulomb's potential.

The reasons for this phenomenon can possibly be explained by the forces originated from solvent interaction. The sources can be related to polarization effects that were intensified by the presence of charges, since the solvent will hydrate the charged polar head. This hydration probably is the major responsible for the increased repulsion between heads. The same effect is present in the counter-ion potentials where repulsion forces are also. Another option is possible related to dielectric permittivity adopted for the coarse-graining experiments because the adoption of implicit solvent requires the consideration of a relative permittivity, but there is no guarantee that the experimental permittivity for water is suitable for coarse-graining techniques and therefore this point should be analysed more deeply. As it is possible to see in Figure (17), the sum of the electrostatic potential calculated assuming  $\epsilon = 78$  and the "charged" potential (green line) originates the blue line which theoretically should be the same as the integrated potential (red line), however even though they have almost the same shape the blue line seems to be shifted to a slightly higher energy state.

Next, once the potentials were exported to GROMACS, the initial concern was which cut-off range and type of electrostatic parameters should be used for this coarse-grain model. For this a 1000 DMDDs molecules system was simulated using the most refined potential set from the long-range model. The parameters evaluated were first which type of potential should be used: Plain Cut-off or PME electrostatics, and the second test was to determine to what extent the the cut-off range affects the results. As determined during MagiC process, the electrostatic potential portion removed from the potential curve was set to be a  $15 \text{ nm}$  cut-off range, thus this value should be the optimal value to use in GROMACS simulations. Figure (18) reveals that when this cut-off range is used with Plain Cut-off electrostatic type it seems that the self assembly structure resulted in an artifact, since there is too much apolar parts exposed to the implicit solvent. Apparently, the molecules are attempting to replicate the atomistic reference rather than organizing in a real structure where the apolar part is as less as possible exposed to the solvent. Nevertheless, when the PME electrostatic type was used with the same cut-off the layer self assembled in a much more physically meaningful structure. The structure formed were two semi-planes with infinite periodicity in one direction, moreover it seems to change the orientation of some polar heads to reduce the contact with the solvent at the exposed edges of the plane. Therefore, this second structure seem to be far more reasonable and probably is the best parameters to use when doing GROMACS simulations. When the Cut-off range was increased to  $20 \text{ nm}$  it is possible to identify again an artifact-like structure, however now the preferred organization was to form a stack of circular layers, tending to organize in a sphere. This behaviour may be a result of the influence from forces that should not be present and then the molecules are trying to maintain a compact structure around a center of mass. For the integrated potential, the electrostatic cut-off and type did not make any difference since there were no charges. However as it can be seen in Figure (18), even though the layer formed is very similar to the ideal format it is possible to observe that the polar heads are not organizing in a suitable configuration, tending to agglomerate independently of the type and in some areas there is no distinction between charged heads and neutral heads. This fact is a good indicative that electrostatic forces should be treated separately from all other types, since it may generate very erroneous results.

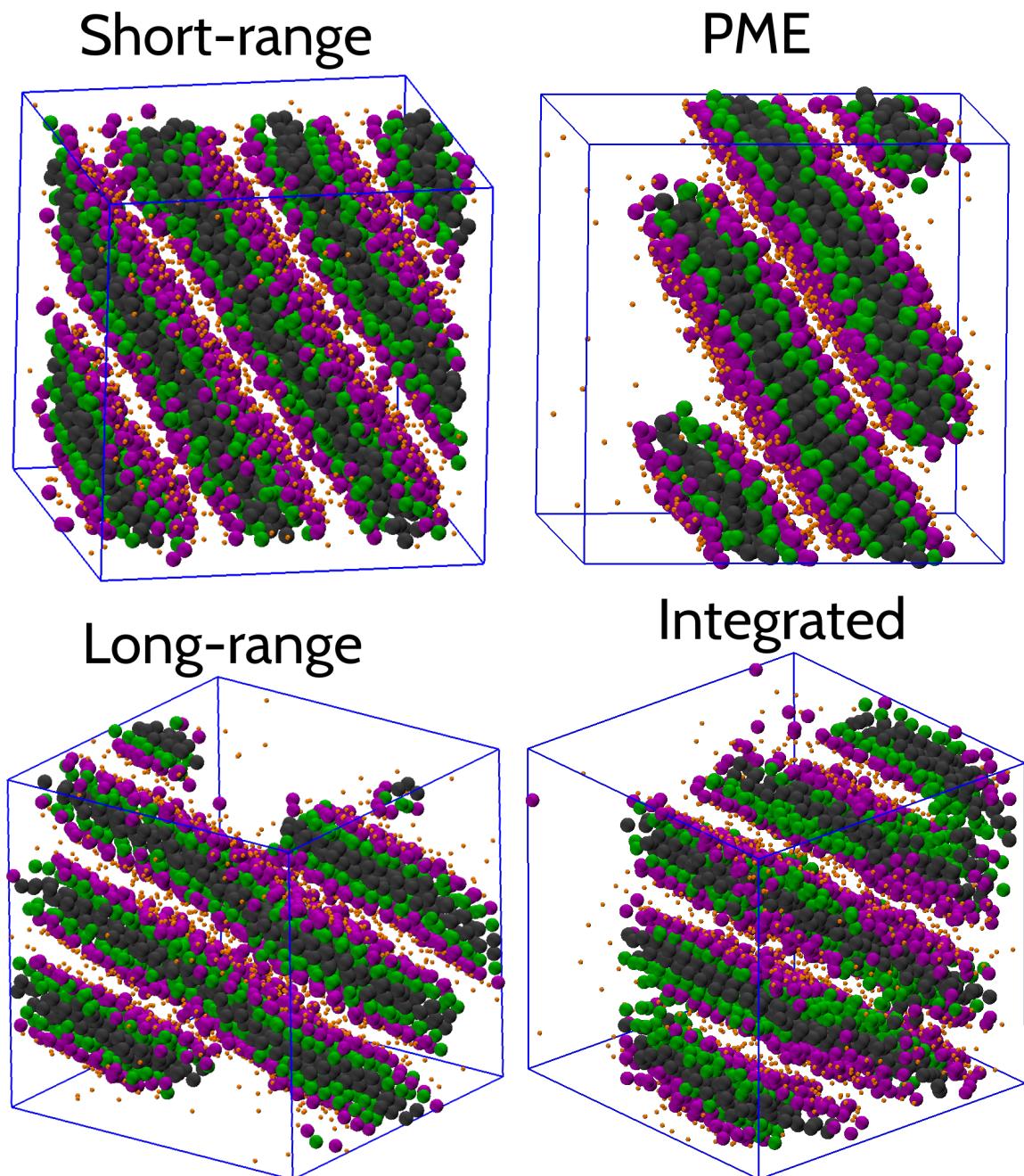


Figure 18: Work in Progress

#### 6.4 Silica-Surfactant interaction

### 7 Conclusion

### 8 Nomenclature

$k_B$	Boltzmann's Constant
$C_n$	Molar concentration of species n ( $ mM$ )

## 9 References

- Arnarez, C., Uusitalo, J. J., Masman, M. F., Ingólfsson, H. I., de Jong, D. H., Melo, M. N., ... Marrink, S. J. (2015). Dry martini, a coarse-grained force field for lipid membrane simulations with implicit solvent. *Journal of Chemical Theory and Computation*, 11(1), 260-275. Retrieved from <http://dx.doi.org/10.1021/ct500477k> doi: 10.1021/ct500477k
- Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A., & Haak, J. R. (1984). Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8), 3684-3690. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/81/8/10.1063/1.448118> doi: <http://dx.doi.org/10.1063/1.448118>
- Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), -. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/126/1/10.1063/1.2408420> doi: <http://dx.doi.org/10.1063/1.2408420>
- Darden, T., York, D., & Pedersen, L. (1993). Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12), 10089-10092. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/98/12/10.1063/1.464397> doi: <http://dx.doi.org/10.1063/1.464397>
- Fernández, D. P., Mulev, Y., Goodwin, A. R. H., & Sengers, J. M. H. L. (1995). A database for the static dielectric constant of water and steam. *Journal of Physical and Chemical Reference Data*, 24(1), 33-70. Retrieved from <http://scitation.aip.org/content/aip/journal/jpcrd/24/1/10.1063/1.555977> doi: <http://dx.doi.org/10.1063/1.555977>
- Frenkel, D., & Smit, B. (2001). *Understanding molecular simulation: From algorithms to applications*. Elsevier Science.
- Hoover, W. G. (1985, Mar). Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A*, 31, 1695–1697. Retrieved from <http://link.aps.org/doi/10.1103/PhysRevA.31.1695> doi: 10.1103/PhysRevA.31.1695
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2), 926-935. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/79/2/10.1063/1.445869> doi: <http://dx.doi.org/10.1063/1.445869>
- Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45), 11225-11236.
- Kresge, C. T., & Roth, W. J. (2013). The discovery of mesoporous molecular sieves from the twenty year perspective. *Chem. Soc. Rev.*, 42, 3663-3670. Retrieved from <http://dx.doi.org/10.1039/C3CS60016E> doi: 10.1039/C3CS60016E
- Lee, M. S., Salsbury, F. R., & Brooks, C. L. (2004). Constant-ph molecular dynamics using continuous titration coordinates. *Proteins: Structure, Function, and Bioinformatics*, 56(4), 738-752. Retrieved from <http://dx.doi.org/10.1002/prot.20128> doi: 10.1002/prot.20128
- Lyubartsev, A. P., & Laaksonen, A. (1995, Oct). Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys. Rev. E*, 52, 3730–3737. Retrieved from <http://link.aps.org/doi/10.1103/PhysRevE.52.3730> doi: 10.1103/PhysRevE.52.3730
- Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P., & de Vries, A. H. (2007). The martini force field: Coarse grained model for biomolecular simulations. *The Journal of Physical Chemistry B*, 111(27), 7812-7824. Retrieved from <http://dx.doi.org/10.1021/jp071097f> (PMID: 17569554) doi: 10.1021/jp071097f
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., & Teller, E. (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6), 1087-1092. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/21/6/10.1063/1.1699114> doi: <http://dx.doi.org/10.1063/1.1699114>
- Mirzoev, A., & Lyubartsev, A. P. (2013). Magic: Software package for multiscale modeling. *Journal of Chemical Theory and Computation*, 9(3), 1512-1520. Retrieved from <http://dx.doi.org/10.1021/ct301019v> doi: 10.1021/ct301019v
- Mirzoev, A., & Lyubartsev, A. P. (2014). Magic: User manual. version 1.0.3 [Computer software manual].
- Moore, T. C., Iacovella, C. R., & McCabe, C. (2014). Derivation of coarse-grained potentials via multistate iterative boltzmann inversion. *The Journal of Chemical Physics*, 140(22), -. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/140/22/10.1063/1.4880555> doi: <http://dx.doi.org/10.1063/1.4880555>

- Nosé, S. (1984). A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics*, 52(2), 255-268. Retrieved from <http://dx.doi.org/10.1080/00268978400101201> doi: 10.1080/00268978400101201
- Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12), 7182-7190. Retrieved from <http://scitation.aip.org/content/aip/journal/jap/52/12/10.1063/1.328693> doi: <http://dx.doi.org/10.1063/1.328693>
- Patwardhan, S. V. (2011). Biomimetic and bioinspired silica: recent developments and applications. *Chem. Commun.*, 47, 7567-7582. Retrieved from <http://dx.doi.org/10.1039/C0CC05648K> doi: 10.1039/C0CC05648K
- Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., ... Lindahl, E. (2013). Gromacs 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), 845-854. Retrieved from <http://bioinformatics.oxfordjournals.org/content/29/7/845.abstract> doi: 10.1093/bioinformatics/btt055
- Pérez-Sánchez, G., Gomes, J. R. B., & Jorge, M. (2013). Modeling self-assembly of silica/surfactant mesostructures in the templated synthesis of nanoporous solids. *Langmuir*, 29(7), 2387-2396. Retrieved from <http://dx.doi.org/10.1021/la3046274> (PMID: 23343439) doi: 10.1021/la3046274
- Satoh, A. (2010). *Introduction to practice of molecular simulation: Molecular dynamics, monte carlo, brownian dynamics, lattice boltzmann and dissipative particle dynamics*. Elsevier Science.
- Tanев, P. T., Liang, Y., & Pinnavaia, T. J. (1997). Assembly of mesoporous lamellar silicas with hierarchical particle architectures. *Journal of the American Chemical Society*, 119(37), 8616-8624. Retrieved from <http://dx.doi.org/10.1021/ja970228v> doi: 10.1021/ja970228v
- Tolbert, S. H., Firouzi, A., Stucky, G. D., & Chmelka, B. F. (1997). Magnetic field alignment of ordered silicate-surfactant composites and mesoporous silica. *Science*, 278(5336), 264-268. Retrieved from <http://www.jstor.org/stable/2894713> doi: 10.2307/2894713
- Tresset, G. (2009). The multiple faces of self-assembled lipidic systems. *PMC Biophysics*, 2(1). Retrieved from <http://dx.doi.org/10.1186/1757-5036-2-3> doi: 10.1186/1757-5036-2-3
- van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A. R., Apol, E., Meulenhoff, P. J., ... Berendsen, H. J. C. (2013). Gromacs user manual version 4.6.5 [Computer software manual].
- Wallace, J. A., & Shen, J. K. (2012). Charge-leveling and proper treatment of long-range electrostatics in all-atom molecular dynamics at constant ph. *The Journal of Chemical Physics*, 137(18), -. Retrieved from <http://scitation.aip.org/content/aip/journal/jcp/137/18/10.1063/1.4766352> doi: <http://dx.doi.org/10.1063/1.4766352>
- Zhao, D., Feng, J., Huo, Q., Melosh, N., Fredrickson, G. H., Chmelka, B. F., & Stucky, G. D. (1998). Triblock copolymer syntheses of mesoporous silica with periodic 50 to 300 angstrom pores. *Science*, 279(5350), 548-552. Retrieved from <http://www.sciencemag.org/content/279/5350/548.abstract> doi: 10.1126/science.279.5350.548

## 10 Appendix