

UNIVERSITY OF STRATHCLYDE

DEPARTMENT OF CHEMICAL & PROCESS ENGINEERING

M.ENG CHEMICAL & PROCESS ENGINEERING 18530

Introduction to Computer-aided Design of Bio-inspired Nanoporous Silica Materials

Author:

André CREScenzo

Supervisors:

Miguel JORGE

Alessia CENTI

Carlos F. RANGEL

May 4, 2015

1 Summary

This project presented the design of a Coarse-grain model by using MagiC software package. MagiC is a systemic tool to develop Coarse-grain models based on an atomistic simulation by applying two techniques: Iterative Boltzmann Inversion and Inverse Monte Carlo. Using GROMACS software package, several All-atoms simulations were run in order to create multiple Coarse-grain models, using different designs and concentrations. The characteristics of Inversion process for each model were described highlighting common points. Afterwards, up-scaled simulations of these models were compared and evaluated revealing interesting features of self-assembly process of a bolaamphiphilic surfactant called 1,12-diaminododecane (DMDD). All the work done created a solid basis to future studies of more complex systems.

Contents

	Page
1 Summary	i
2 Acknowledgements	iii
3 Introduction	1
4 Theoretical Basis	2
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
5 Methods Description	7
5.1 GROMACS All-Atom simulations	7
5.2 Experiment 1: Model Creation	8
5.3 Experiment 2: Concentration Analysis	10
6 Results and Discussion	11
6.1 GROMACS simulations	11
6.2 Experiment 1: Model Creation	13
6.2.1 Convergence Analysis	13
6.2.2 Reproduction Tests	15
6.2.3 System up-scaling	16
6.3 Experiment 2: Concentration Analysis	17
6.3.1 Systematic Replication	17
6.3.2 Potential Comparison	18
6.3.3 Interchangeability Test	19
7 Conclusion	21
8 Nomenclature	21
9 References	22
10 Appendix	1

2 Acknowledgements

I would like to express my sincere gratitude to my three supervisors: Miguel Jorge, Alessia Centi and Carlos Rangel for providing me an outstanding support throughout the whole project and for introducing me to the adventurous field of Molecular Dynamics. A special thanks to Alessia who kindly helped me with the writing and shared fundamental informations from her current works, by this means enabling the realization of this project. Last but not least, I would like to thank my fellow friends and flat-mates who always encouraged me to keep writing, with special thanks to João Luís not only for the great cooperative work during GROMACS learning, but also with the text revisions.

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 thermodynamic changes on the system, originated from solvent interaction with amphiphilic molecules, are incorporated in these approximated potentials. Therefore, changes in parameters as concentration may not affect coarse-grained model performance in recreating self-assembly structures (Mirzoev & Lyubartsev, 2014b).

The work presented in the following experiments is an attempt to create a flexible 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 Taney 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 tabulated Lennard-Jones potential interactions. As a final objective at the end of this project, an implicit water coarse-grain model for DMDD will be generated and properly validated based on MD simulations and thermodynamic properties, in order to provide a satisfactory approximation to interactions with silica oligomers in a mesoscale model.

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 Δt 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 λ 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 ξ 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 ϵ and σ 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.

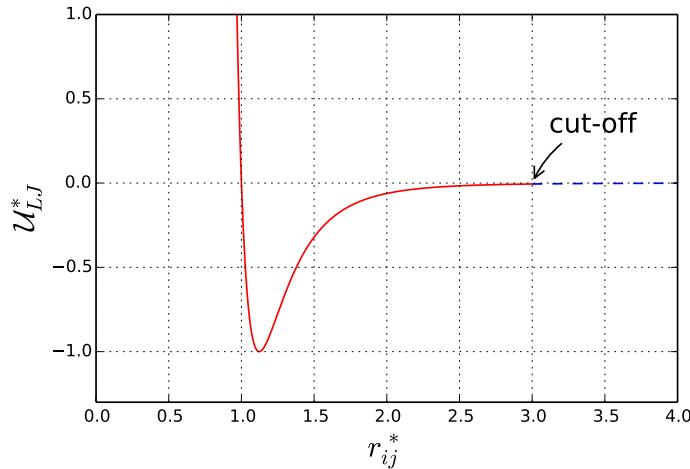


Figure 1: Example of Lennard-Jones Potential (with reduced units). For this plot $\epsilon = 1$ and $\sigma = 1$

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.

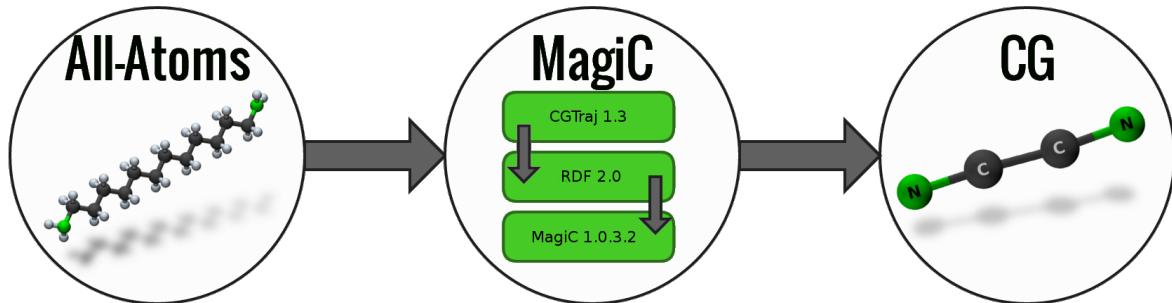


Figure 2: Graphical representation of MagiC's systematic process. The central circle represent the process itself while the other two are input and output. The green boxes indicates each step inside MagiC process.

The first step uses "CGtraj 1.3" (MagiC package utility written in Fortran 90) (Mirzoev & Lyubartsev, 2014a), 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, 2014a), 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, 2014a), 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.

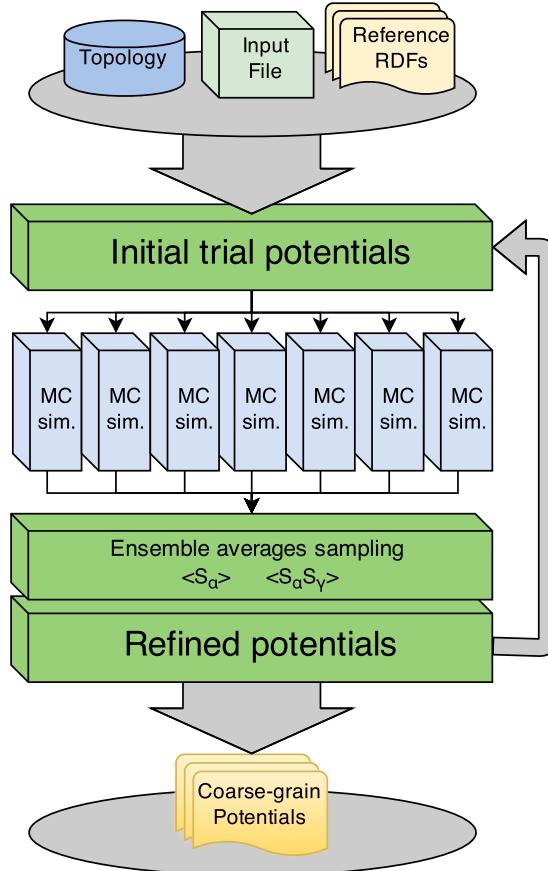


Figure 3: MagiC kernel work-flow scheme

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 η 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}_α . However, a good approximation for numerical methods is to assume a cut-off radius r_{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}_α in which \mathcal{S}_α is the number of pairs within interval $[r_\alpha, (r_\alpha + \Delta r)]$. Since \mathcal{S}_α 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}_α 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:

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

Both 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

5.1 GROMACS All-Atom simulations

The experimental process of this project was divided in two main parts: Firstly, develop an efficient coarse-grain model and validate its use. Secondly, analyse the relationship between concentration of all-atoms simulation and concentration of coarse-grained simulation.

In order to maintain standard parameters for all GROMACS simulations, all-atoms simulations that were used for coarse-grained reference were run in NPT ensemble. All simulation boxes were created using "genbox" (GROMACS package utility), first by inserting the desired number of surfactant molecules and then adding the solvent at the adequate ratio. All molecules were designed based on OPLS-AA forcefield (Jorgensen et al., 1996) parameters. The surfactant DMDD was created with aid of co-workers and water molecules type was TIP4P (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). Moreover, no initial configuration was determined, such as a pre self-assembled system, to ensure stochastic nature of the simulation.

Initially, 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.

Regarding the concentration of surfactant in each box, three all-atom boxes were created with composition and size described in Table (1):

Table 1: Description of all-atom simulation boxes

Box	DMDD (no. Molec.)	Water (no. Molec.)	Silica (no. Molec.)	Avg. size (Å)	Concentration (mM)
B1	30	1500	-	38.2646	0.89
B2	30	670	-	31.3201	1.62
B3	45 ^a	240	-	28.2972	3.30

^a Necessary change, otherwise box size would be too small for the experiments, however concentration is the same as a 30-160 box.

The utility of each box will be described further in each experiment section. Even though they have different structural parameters, 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 350 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.

5.2 Experiment 1: Model Creation

Since the main objective of this project is to probe the effectiveness of coarse-grain models generated using MagiC, this first experiment was an attempt to understand and analyse to what extent bead size influences the process of coarse-graining. Hence, two models were proposed: Model 1 (M1) described in Figure (4) and Model 2 (M2) described in Figure (5).

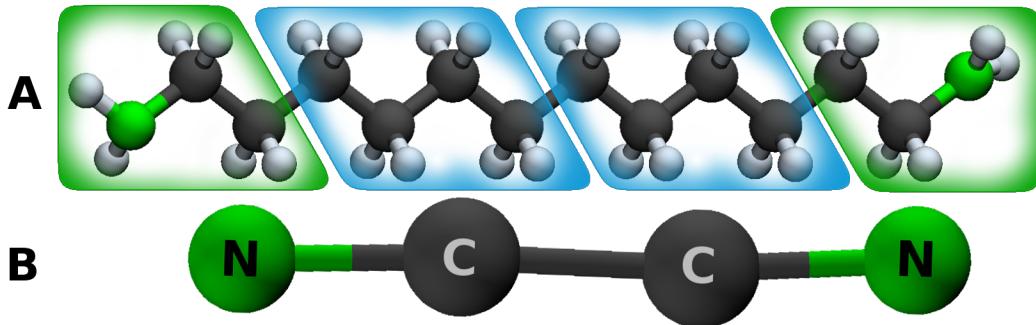


Figure 4: A) Schematic of M1 molecule split plan, Carbon atoms represented in black, Nitrogen in green and Hydrogen in gray. B)Final coarse-grain model for M1. This model aggregated the three most polar heavy atoms on the tips and four apolar heavy atoms at the center.

The differences between M1 and M2 were an effort to distinguish at different level of detail the amphiphilic nature of the DMDD. In order to avoid electrostatic interactions on the IMC process and further GROMACS simulations, the heavy atom groups (heavy atom + hydrogen) were kept together, thus the final charge remains neutral in each bead. Hence, forces will only arise from Van der Waals interactions and implications of inverse process might originate from degrees of freedom due to number of beads. The All-Atom simulation box chosen as reference for model creation was B1, because for this experiment reference concentration was not a major concern yet, more information regarding this point is given at Experiment 2 (see Section 6.3)

In order to fully describe the system, MagiC generates potential tables for interaction between each bead. The first type of tables are for each bond type, for example in M1, there are two different tables: one for N-C bond and another for C-C bond. However, because the molecule is symmetric

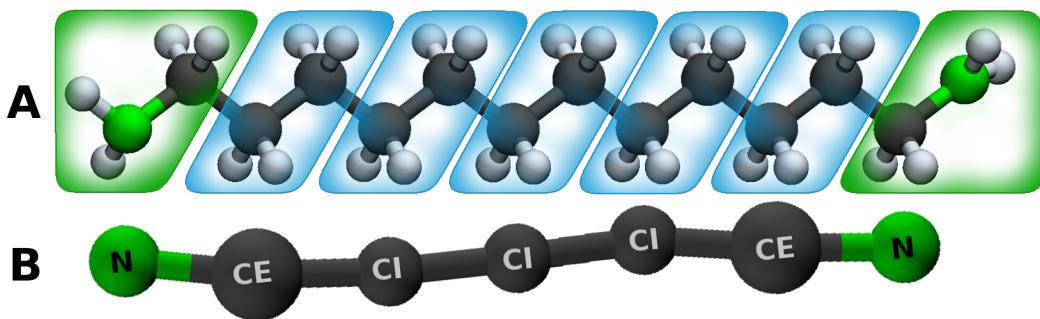


Figure 5: A) Schematic of M2 molecule split plan, Carbon atoms represented in black, Nitrogen in green and Hydrogen in gray. B)Final coarse-grain model for M2. Heavy atoms were grouped pairwise in such a manner that amphiphilic parts were kept as distinct as possible. It is important to notice the necessity of two carbon bead types CE and CI.

the two external N-C bonds use the same potential table. Secondly, it generates tables for each angle type, following the symmetry rule. And finally, each type of intermolecular pair has its own potential table, i.e. for M1 there are three: N-N, N-C and C-C. A more detailed description for both models is exposed in Table (2).

Table 2: Description of coarse-grain models

	Bonds	Angles	Interatomic
Model 1	[N-C] [C-C]	[N- \widehat{C} -C]	[N-N] [N-C] [C-C]
Model 2	[N-CE] [CE-CI] [CI-CI]	[N- \widehat{CE} -CI] [CE- \widehat{CI} -CI] [CI- \widehat{CI} -CI]	[N-N] [N-CE] [N-CI] [CE-CE] [CE-CI] [CI-CI]

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, 2014a). 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, 2014a). 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 for M1 and M2 can be found in Appendix (see Section 10).

Once reference RDFs and a Topology were specified, the inversion step started. The parameters of each inversion step were based on work done by Mirzoev and Lyubartsev (2014b), in which similarly to this project the authors also analyse the self-assembly of amphiphilic molecules. Nevertheless, considering that DMDD owns a less complex molecular structure, the number of Monte Carlo simulation steps were considerably modified for each model. The simulations were run in parallel using up to 24 cores. Table (5.2) exposes the specific number of steps per core and whether it used IBI or IMC method for inversion. The ensemble averages were 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 $\epsilon = 78$ (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, the potentials and topology were exported to GROMACS format. In order to compare both models to the atomistic model, a reproduction test was done using the same parameters of all-atom boxes, with the exception of dielectric constant that was changed to $\epsilon = 78$ and specification of each potential table required for GROMACS Simulation. Furthermore, since the charges of all beads are zero, PME long-range electrostatics were replaced by cut-off scheme which increased simulation speed and cut-off was increased to $R_{cut} = 1.8 \text{ nm}$. An example of reproduction test input parameters can be found in Appendix (see Section 10).

Table 3: Inversion Process description for M1 and M2

	N_{Inv}	10	10	10	5		
M1	Method	IBI	IMC	IMC	IMC		
	N_{Cores}	8	8	8	8		
	N_{MCe}	3000000	3750000	5625000	7500000		
	N_{MC}	12000000	15000000	22500000	30000000		
	N_{Avg}	72000	90000	135000	180000		
M2	N_{Inv}	10	10	10	5	5	14
	Method	IBI	IMC	IMC	IMC	IMC	IMC
	N_{Cores}	24	24	24	24	24	24
	N_{MCe}	1000000	10000000	20000000	30000000	40000000	30000000
	N_{MC}	4000000	30000000	60000000	90000000	120000000	250000000
	N_{Avg}	72000	480000	960000	1440000	1920000	5280000

Finally regarding up-scaling tests, all simulations were made with the same machine using up to 8 MPI threads, depending on the size of the system. All values were obtained from the prediction made by "mdrun" (GROMACS package utility) (van der Spoel et al., 2013), which provides an estimated average of simulation speed, however it is important to consider that it may vary throughout the whole simulation. The simulation parameters were the same as the reproduction tests, however now the simulations were run using a higher timestep of 20 fs for M1 and 10 fs for M2.

5.3 Experiment 2: Concentration Analysis

Subsequently, this second experiment has two main objectives, both related to concentration of DMDD in the system. The former objective is focused mainly on influence of concentration variations over the inverse steps, that means how the complexity of the self-assembly structure affects convergence during IBI and IMC. The latter objective is to understand the relationship between the variation of DMDD concentration in the all-atom reference system and its effects in the self-assembly behaviour of the coarse-grain model chosen. Moreover, it will be an attempt to confirm the assumption that implicit water coarse-grain models can assume multiple concentrations, independently of the concentration chosen in the all-atoms reference.

This assumption derives from the concept that implicit water models can only assume a NVT ensemble because pressure has no physical meaning in these systems due to the absence of solvent. It is supposed that the entropy from free energy (see Eq.(11)) that comes partially from solvent interaction is lumped into the potential obtained using the inversion process. However, differently from temperature it is not yet clear to what extent concentration has influence over entropy. Hence, in this case just by varying the size of simulation box, the implicit water system should approximately behave as if it was on a different concentration.

The experiment begun with the development of two new coarse-grain models based on M1 from the previous experiment. However, at this moment the all-atoms box's concentrations were increased to $C_{B2} = 1.62\text{ mM}$ and $C_{B3} = 3.30\text{ mM}$, thus the Intermediate Concentration Model (M_I) and High Concentration Model (M_H) were created based on those concentrations respectively. Furthermore, for this experiment, M1 which has the lowest concentration will now be denominated as Low Concentration Model (M_L).

A complete description regarding the Inversion process for both latter models is given in Table (4), showing the peculiarities of each one. To be more specific, in order to evaluate the efficiency of inversion method for different concentrations, M_I followed almost the same process as like M_L with the exception of number inverse steps and Monte Carlo simulations. But, on the other hand, M_H had two modifications: the former was the inclusion of a higher number of DMDD molecules in the simulation box and the latter was a new technique adopted to optimize the Inversion process. The first alteration was an inevitable consequence of the increasing concentration, because otherwise the number of molecules in the all-atom box would not be enough to fulfil the minimum requirements of box size to the adopted cut-off radius $R_{cut} = 1.3\text{ nm}$. The second alteration in the process of model creation was on the initial phase of inversion, because instead of using IBI to approximate suitable initial trial potentials, the final potentials obtained for M_L were adapted and used as a initial basis. Further explanations regarding this technique are shown in the Results and Discussion section (See Section 6.3.1).

After, with the three models, a comparison test was made in order to observe whether the systems

Table 4: Inversion Process description for M_I and M_H

M_I	N_{Inv}	10	10	10	5	10	5	5	5	9
	Method	IBI	IMC	IMC	IMC	IMC	IMC	IMC	IMC	IMC
	N_{Cores}	24	24	24	24	24	24	24	24	24
	N_{MCe}	1000000	1250000	1875000	2500000	10000000	20000000	30000000	40000000	50000000
	N_{MC}	4000000	5000000	7500000	10000000	30000000	60000000	80000000	100000000	150000000
	N_{Avg}	72000	90000	135000	180000	480000	960000	1200000	1440000	2400000

M_H	N_{Inv}	(No IBI) Initial trial potentials extracted from M1	10	10	5	10	3
	Method		IMC	IMC	IMC	IMC	IMC
	N_{Cores}		8	8	8	8	8
	N_{MCe}		10000000	20000000	30000000	33000000	40000000
	N_{MC}		30000000	60000000	90000000	100000000	120000000
	N_{Avg}		160000	320000	480000	536000	640000

behave similarly with the same box size, that means concentration. Nevertheless, since the surfactant self-assembly simulations were using coarse-grained models, the system had been up-scaled. Each simulation box contained 1000 molecules of Coarse-grained DMDD and the size of the box's edge was determined by rescaling average 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. Full description of all nine simulation boxes is given in Table (5). All simulation parameters were exactly the same as used in Experiment 1 up-scaling simulations, but now the simulations were run until the moment that self-assembly structure had appeared to achieve a stable configuration.

Table 5: Description of coarse-grain simulation boxes

Box	$B_{L@L}$	$B_{I@L}$	$B_{H@L}$	$B_{L@I}$	$B_{I@I}$	$B_{H@I}$	$B_{L@H}$	$B_{I@H}$	$B_{H@H}$
Model	M_L	M_I	M_H	M_L	M_I	M_H	M_L	M_I	M_H
Size (nm)	12.3147	12.3147	12.3147	10.08	10.08	10.08	7.9556	7.9556	7.9556
Concentration (mM)	0.89	0.89	0.89	1.62	1.62	1.62	3.30	3.30	3.30

6 Results and Discussion

6.1 GROMACS simulations

Regarding the All-Atom simulations, it is interesting to highlight the different types of self-assembled structures obtained in each concentration, all of them having its own observable peculiarities. In order to obtain a cohesive comparison basis between themselves and their coarse-grain models, each one of the simulation boxes was characterized based on their periodicity and Radial Distribution Function.

As a initial approach, a qualitative evaluation of the simulations is provided in Figure (6). With the aid of periodic replicas, this image reveals that only at a high concentration (6c) it became possible to form a complete layer in all periodic directions. Contrastingly, the other two boxes only showed periodicity in one of the directions, perpendicular the normal vector of the formed layer. Concerning the dynamic behaviour of both systems, it is observable that this structure remain stable during all simulation and the layer kept constantly rotating around the periodic axis, although meanwhile the high concentration layer seldom presented any rotation.

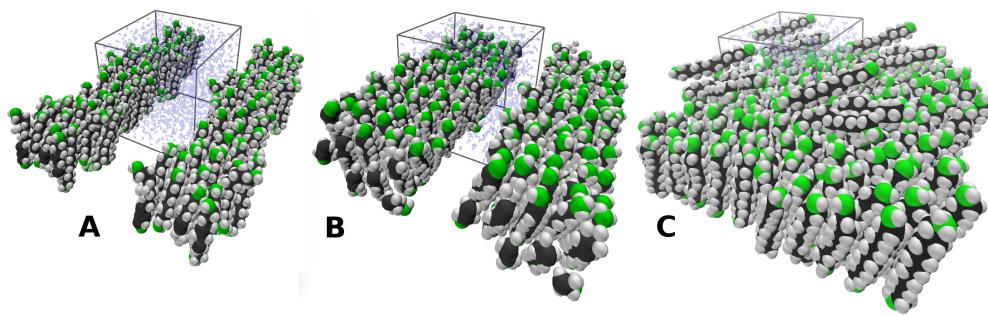


Figure 6: Snapshot of simulation boxes with periodic replicas. A) box at low concentration (B1), B) box intermediate concentration (B2) and C) box at high concentration (B3). The original images are the ones located inside the grid box surrounded by water. Nitrogen atoms are represented in green, Carbon in black, Hydrogen in gray and water molecules in transparent blue.

Next, 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 (7), the RDFs of all 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.

As a final consideration, it is possible to observe that the peaks of the high concentration box are located in a slightly distinct position, and that may be a consequence of the lack of space inside the box. Moreover it seems that there is an ideal concentration between Intermediate and High concentration where one complete periodic layer should be formed without leftover molecules. A single layer is the desirable configuration since in a real system seldom will present any non-aggregated molecule. Therefore, for future experiments including silica this concentration will be adopted.

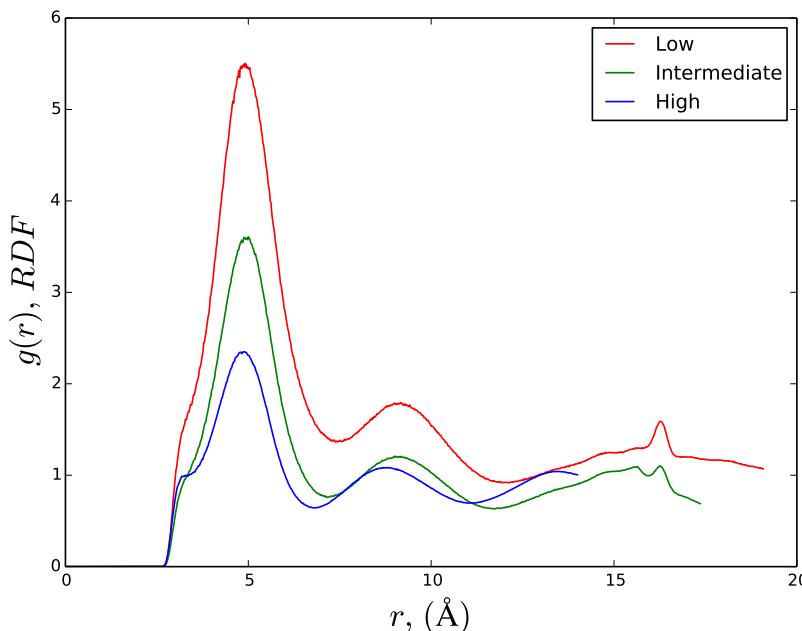


Figure 7: Radial distribution function between N-N atoms.

6.2 Experiment 1: Model Creation

6.2.1 Convergence Analysis

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. A good example of this phenomena in shown in Figure (8) were the deviation between reference RDF and Reproduction test RDF is numerically low, however it is very clear visually that the system is not reflecting the expected result. After extensive trials with all models, it was verified that an acceptable value for deviations was any sum below 0.025. In order to proceed beneath this amount, it would require an excessive number of Monte Carlo simulations which might be unnecessary, since at this levels all models reproduced plausible results. Hence, before focusing on the analysis of the final coarse-grain model it is important to make an overview upon the whole Inversion process for both models.

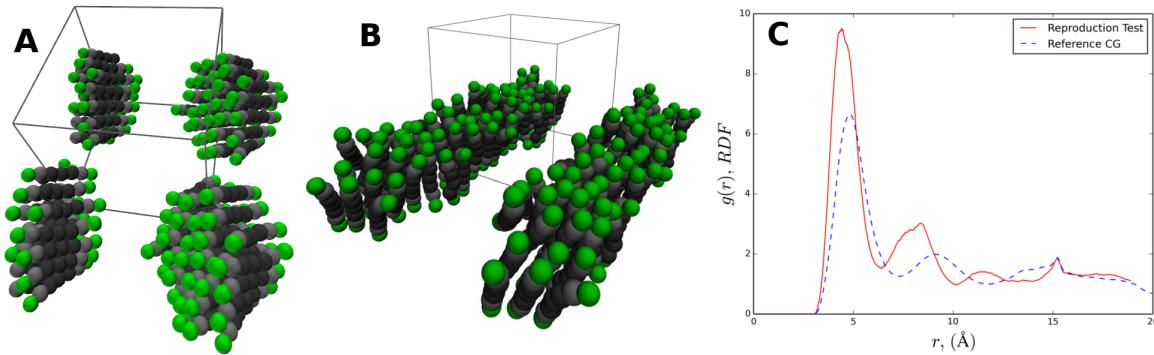
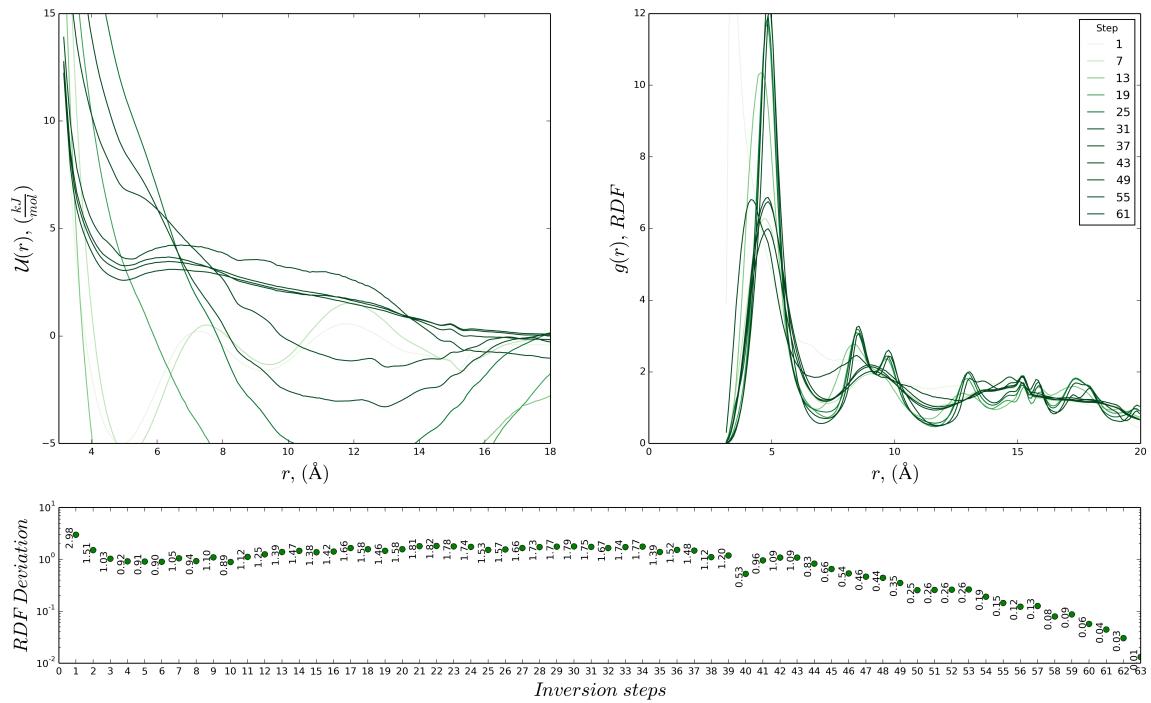
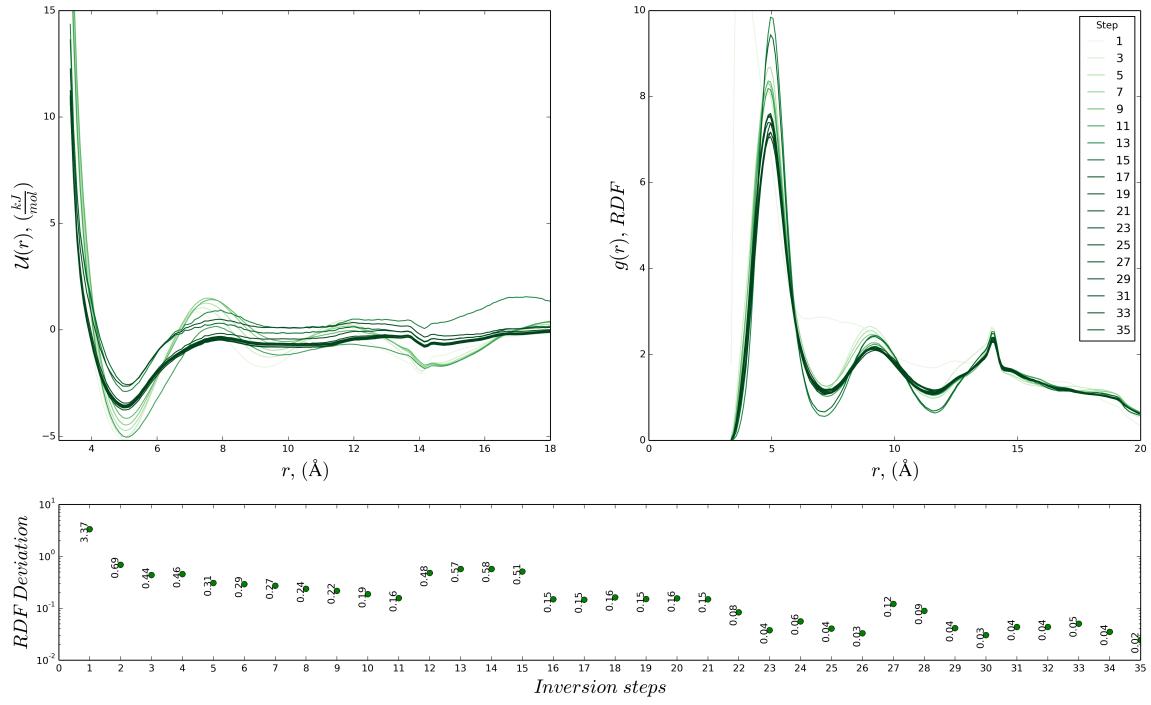


Figure 8: Demonstration of a qualitative evaluation of convergence based on comparison between expected result and trial simulation. A) Snapshot of a equilibrium state after a long simulation time using M2 with poorly converged potentials B) Reference Coarse-grain from CGtraj. Periodic replicas are provided to improve visualization. C) RDFs representing each system (Deviation = 0.52504)

Starting with M1, it is possible to observe in Figure (9) that the process was smooth. Both graphics refers to N-N head interactions which as explained provides a good representation of DMDD behaviour. The trial potentials generated after each iteration went directly to a specific region, therefore meaning that the desired curve should be located somewhere within that area. The deviation graph provided a numerical basis to predict if it was necessary to increase the number of Monte Carlo steps. The sign for that was whenever the error fluctuation started to drift around a certain value, meaning that potential conversion had been affected by statistical noise.

The RDF progression graph had shown a topical behaviour observed in all inversion processes, which is the presence of three distinct convergence phases. In M1's case, during the initial phase (under 13 steps) the RDF rapidly converged to a shape very similar to the reference, but looking at the potential curves it is clear that they were far from the final result. During the second phase (between 13 and 21 steps), the potential curves began to adopt the same aspect of the final curve, however the deviation increased or stagnated at a point whilst the RDF peaks and valleys became more sharp. Then finally the deviation suddenly started to decrease, consequently the RDFs and potentials began to converge to the desired region in a very slow pace. Even though for M1 these signs are not very clear, for other models such as M_1 this behaviour is well defined (See Section 6.3).



Subsequently, Figure (10) presents a similar overview for M2. It is important to keep in mind that even though they have the same name, the N beads are completely different for each model. Therefore there are no similarities between those potential curves. Nevertheless the RDFs should be similar because the center of mass of both beads are relatively near. A key feature of this model is that it was the one who required the largest amount of Monte Carlo simulations for the Inversion steps. Just as shown in Table (5.2), some steps demanded over 6.4 billion steps in order to avoid statistical noise, thus taking a much longer time to reach errors at desired values.

Once again, three phases of convergence were present during the process, and at this time they are clearer to visualize graphically. The shape of the deviation curve reveals the first phase, where potential are far from convergence region, but the error is decreasing. Between steps 7 and 30 approximately, the deviation diverged considerably while potentials assumed a position within a characteristic region, indicating a phase two behaviour. Lastly, both graphics began to slowly converge to the final aspect until desired deviation value was reached.

6.2.2 Reproduction Tests

The efficiency of both models can be evaluated by making a side by side comparison between reproduction test results. Figures (11) and (12) promptly provide not only a quantitative, but also a qualitative validation of both models' capacity of representing the atomistic system. It is important to notice that the RDFs should reproduce exactly the same RDF as the reference; however, at some points they mismatch. This fact is directly related to two factors: level of convergence and cut-off radius adopted.

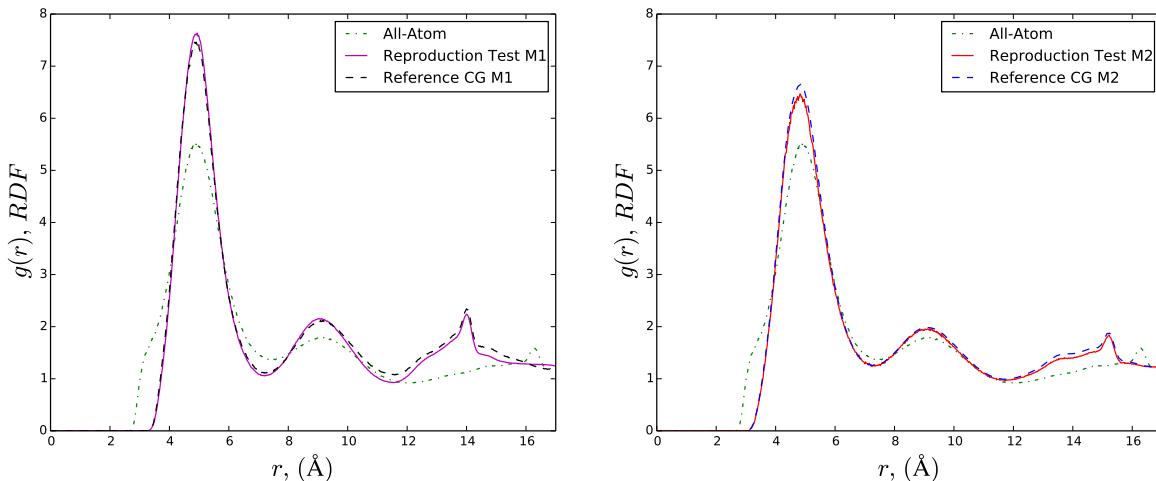


Figure 11: Radial distribution function between N-N beads/atoms. Left graph representing M1 and right graph representing M2

Firstly since M2 is a more detailed model, as expected the RDF aspect is more similar to the atomic RDF. Secondly as soon as the deviation for M1 is higher, the reproduction test RDF showed a greater disparity from reference principally for long-range interactions. Both simulations were run using a $R_{cut} = 1.8 \text{ nm}$. Then based those observations, for the level of convergence factor it is supposable that if M1 had had a lower deviation it would reproduce a finer result. However, it is not necessarily a general rule because coarse-grain models may loose information due to the level of detail of the model. Even though it requires much more computational resources, adopting a finer model such as M2 might enable a lower deviation values, where as a more lumped model such as M1 could drift indefinitely through statistical noise. Nevertheless, the simulation boxes reveals that the layer formed in both cases are qualitatively similar, therefore the loss of detail in M1 is tolerable.

Regarding the influence of the cut-off radius, the simulation analysis using different cut-offs in M2 demonstrated that long range interactions are important for Coarse-grain models. Figure (13) provides a brief example about to what extent R_{cut} can influence the system's behaviour. Hence, differently from All-Atoms simulations long-range interatomic interactions carries a major responsibility regarding self-assembled structures, and the adoption of a short cut-off will cause loss of important information thus reproducing unrealistic results.

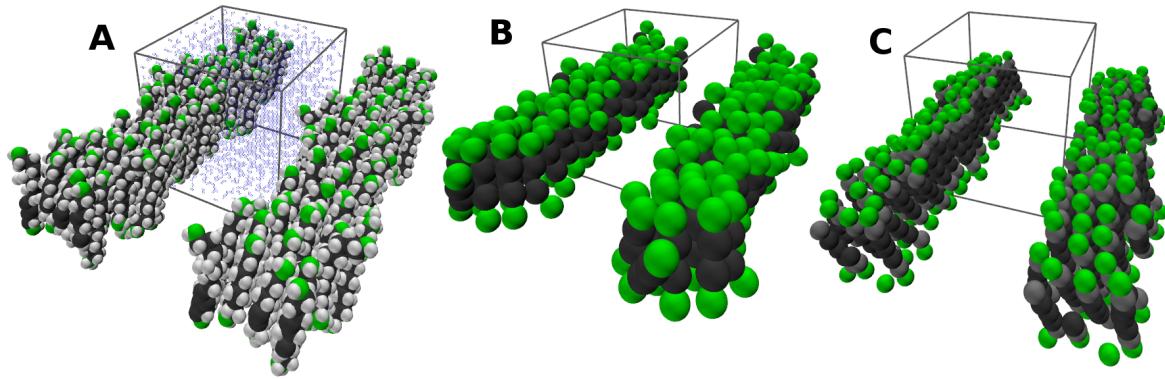


Figure 12: Snapshot of simulation boxes with periodic replicas. A) All-Atom reference at low concentration (B1), B) M1 reproduction test and c) M2 reproduction test.

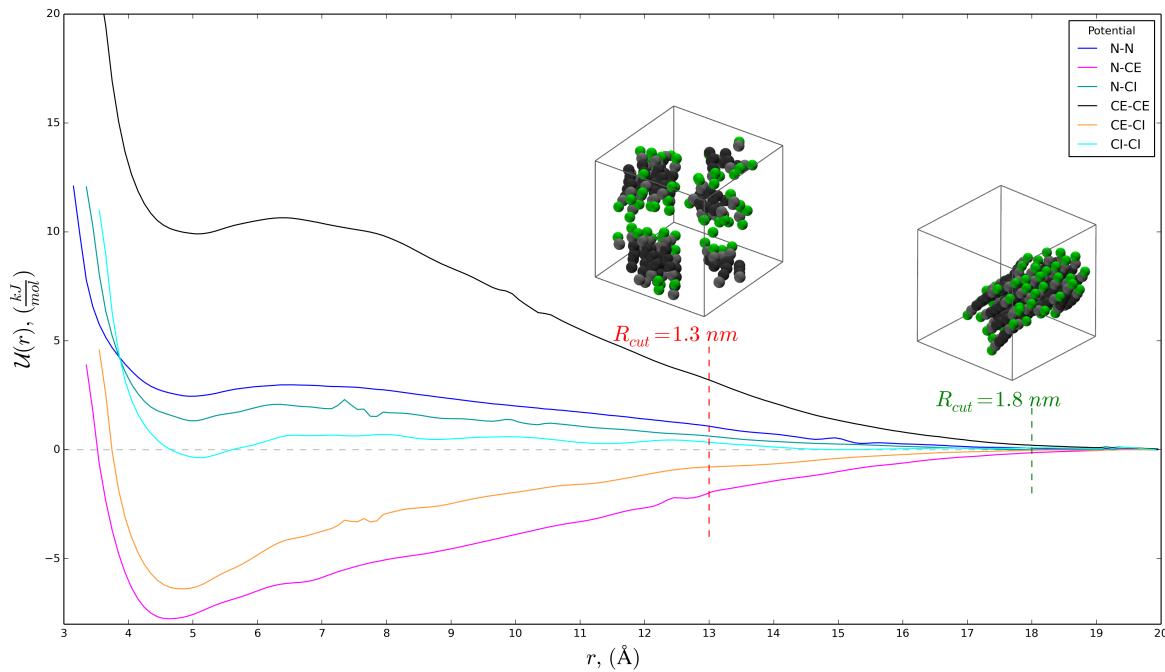


Figure 13: Graphical representation of the potential cut-off radius influence over simulation. Each box is using the respective cut-off. For the simulation any point after the dashed lines will be considered zero.

6.2.3 System up-scaling

A major concern for coarse-grain models is that even though they are made be a good approximation of atomic models, they should be designed in such a manner that the system can be up-scaled. Simulation speed and system size need to be optimized as much as possible without excessive information loss due to lack of level of detail. In order to improve simulation performance, the time-step for each model was readjusted because it is related to the bond size. The movement of one bead in relation to its bonded pair must not exceed maximum bond distance. Therefore, a lower level of detail result in an increased bond size enabling the use of bigger time-steps. Both models proposed for this experiment were able to breed reasonable results from reproduction test, then the last criteria to evaluate which is the most suitable model are the up-scaling capacities. Table (6) shows a comparison of multiple system sizes and respective simulation speed.

Decisively, M1 was the model chosen for further experiments since it seems to be in all cases much

Table 6: System up-scaling for M1 and M2

Number of Molecules	100	1000	5000	10000	20000
Simul. speed for M1	9086 ns/day	985 ns/day	235 ns/day	133 ns/day	67 ns/day
Simul. speed for M2	1458 ns/day	208 ns/day	62 ns/day	33 ns/day	15 ns/day

more efficient when up-scaling than M2. Although M2 is able to reproduce finer results, the error committed by using M1 can be neglected for the benefit of simulation speed. Moreover, the Inversion process for M1 presented a smoother behaviour requiring lesser amounts of Monte Carlo simulations.

6.3 Experiment 2: Concentration Analysis

6.3.1 Systematic Replication

A valuable advantage of using MagiC Coarse-grain method is that once a model is fully described it becomes simple to replicate it using different references. The systematic procedure uses the same structural description to recreate a new model. Nevertheless, minor were changes still necessary in the input files for the tools such as, the number of molecules in the system (solvent and solute), trajectory files from the new All-atom simulation and modifications in simulation box size.

However, the experiment proposed in this section revealed that major differences emerged during the Inversion process. The use of higher concentrations of surfactant in the reference simulation increased drastically the computational effort to obtain the set of potentials. Beginning with M_1 , by observing Figure (14) and crossing with data provided in Table (4), the previous statement is readily confirmed. Moreover, the three phase behaviour is very evident throughout the process. Another important fact is that during later steps even though the number of Monte Carlo simulations was increased greatly, the deviation remained drifting around a certain region. This could indicate that this model probably reached its maximum level of detail, thus hardly precision would not get any higher.

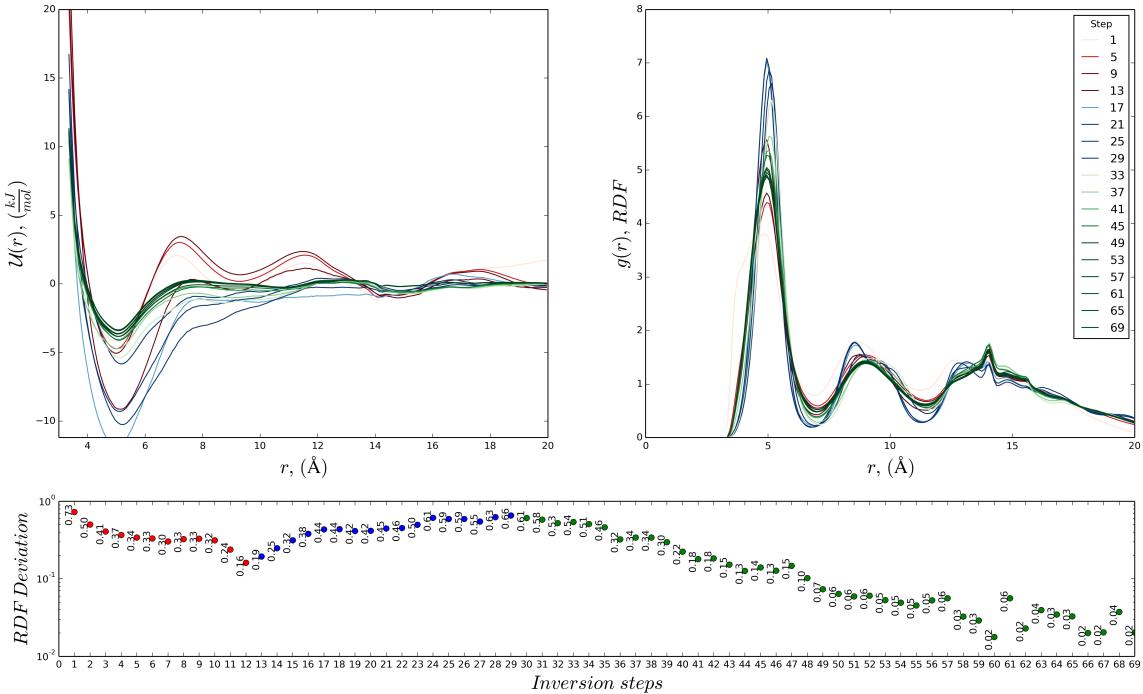


Figure 14: Graphical representation of the inversion process for M_1 . Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing N-N interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. Red lines indicate phase one, blue lines indicate phase two and green lines indicate phase three. 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.

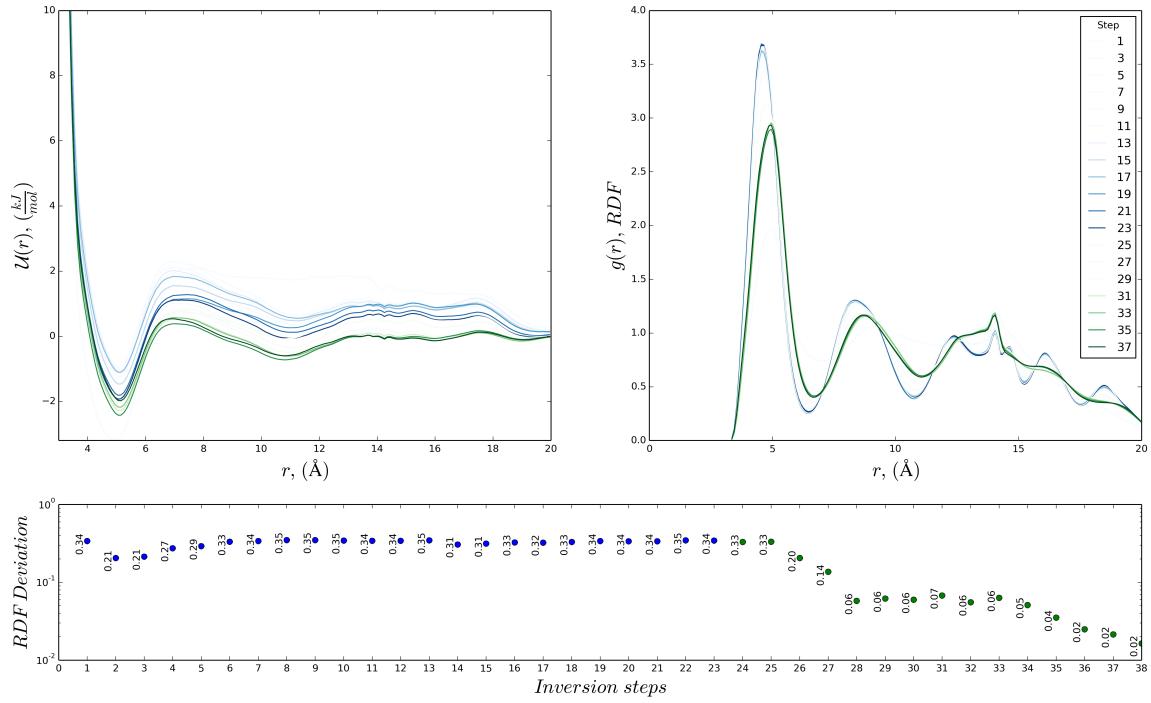


Figure 15: Graphical representation of the inversion process for M_I . Graph 1 (top left) show potential progression at certain steps. Graph 2 (top right) show RDF progression at the same steps as the potential. Both graphs are representing N-N interactions and equal colors represent at the same step furthermore, the darker the line color the higher the step. Blue lines indicate phase two and green lines indicate phase three. 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.

In the case of the high concentration model, the increased complexity demanded modifications to standard procedure because, even with numerous attempts, the Inversion process for M_H model kept diverging indefinitely during phase two. By comparing potential plots of M_L and M_I , it is possible to assume that potential for M_H should be located within the same region. Hence, by setting the initial trial potentials for M_H as being the final potentials of M_L the Inversion process simply skipped phase one and jumped directly to a convergent phase two. As demonstrated in Figure (15), the whole process was considerably smooth and seldom the deviation strayed significantly.

Therefore, this phenomenon indicates that it might be possible to derive potentials for complex system just by adopting a gradual procedure. For further experiments with silica where the number of intermolecular interactions will increase, this technique should be a useful tool to deal with divergent Inversion processes. It is important to keep in mind that this method still need a deeper analysis in order to probe its validity, because it may not be applicable in other cases. However, this analysis is out of the scope of this project and will be evaluated in future studies.

6.3.2 Potential Comparison

In atomic models the Lennard-Jones potentials only accounts for short range intermolecular interactions because surfactant's atoms seldom interact with each other. Therefore, long-range interaction occurs via a solvent which carries the received forces through solvent media until it reaches other molecules. However for an implicit water coarse-grain model, these forces must be included in each bead interaction potentials. With use of different concentrations, the set of potentials obtained for each model revealed interesting features regarding this phenomenon. As showed in Figure (16) the long-range potentials presented more regions of low energy rather than the initial first pit. Hence, these regions will be responsible for the self-assembly structures, which means for example, that M_H will always tend to aggregate in more well structured and rigid structures. That is because the low energy regions are deeper and limited within a shorter region, but in the other hand the other models present wider regions which allows a more flexible structure. This is better illustrated by the C-C potential, however it occurs in all the others.

For instance, the intramolecular interactions are quite similar with the exception of the angular

potential. It seems that for M_H angular positions smaller than 120° the potential energy starts to increase drastically meaning that the molecules will rarely present these angles. This characteristic implies that molecules will occupy less space in the structure formed, therefore the structure will be more compact when compared to the other models.

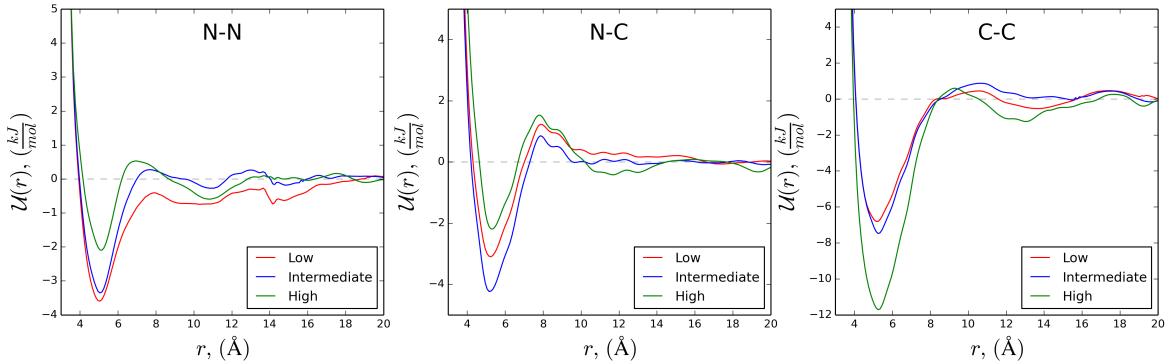


Figure 16: Comparison between Intermolecular potentials for different concentrations. The color shown in the label indicates which model the curve refers. The title in the top indicates which bead types will have these potentials

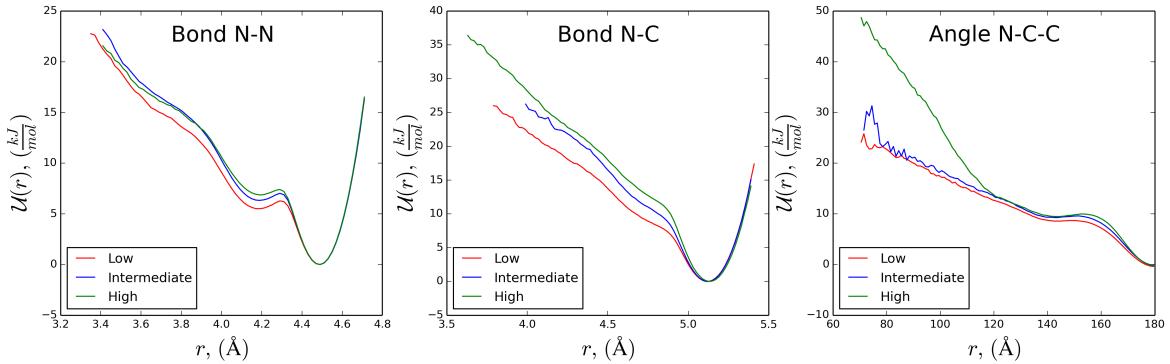


Figure 17: Comparison between Intramolecular potentials for different concentrations. The color shown in the label indicates which model the curve refers. The title in the top indicates which bead types will have these potentials

6.3.3 Interchangeability Test

The final procedure proposed for this project is an attempt to make a cross-comparison to evaluate the capacities of each model at different concentrations. For this, the following nine simulations were run until the system equilibrates in a certain configuration. In case the models were interchangeable, the expected result should be self-assembled structures with similar layout. It is important to keep in mind that the number of molecules is the same in all simulations, therefore only the box size was modified to reproduce varying densities.

The first test used low concentration for the simulations, so Box A in Figure (18) is the reference. Due to the excess of space in the system the final structure was similar to a vesicle, which is an efficient way to reduce contact between solvent and the apolar part. It is noticeable that it is not a complete vesicle, instead the transversal cut of the aggregate reveals that several short crossed planes were the preferred structure. In the other cases, M_I formed a tube like periodic structure and M_H formed a complete "tetrahedral" vesicle with well defined planes.

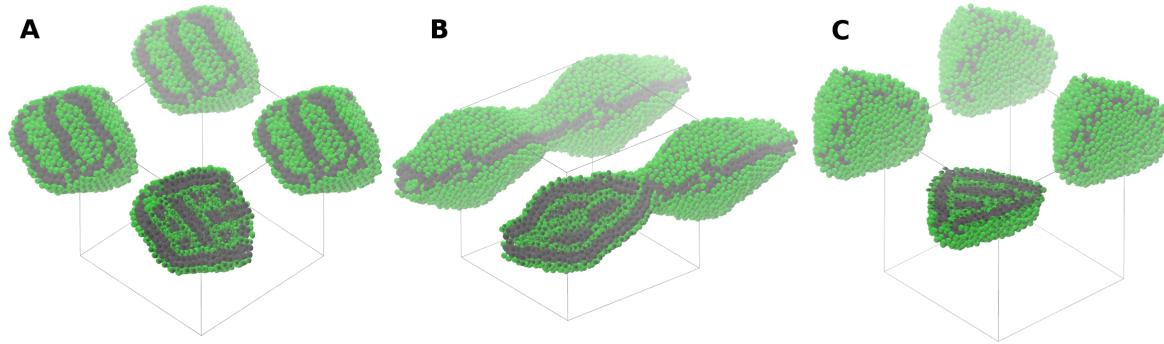


Figure 18: Snapshot of simulation boxes at low concentration. The model used in each box is A) M_L , B) M_I and C) M_H . Periodic replicas are provided to ease mesostructure visualization. The original images are the ones located inside the grid box. The original box was divided by a cutting plane, showing internal conformation of the aggregate. N beads are represented in green, C beads in black.

Secondly, the intermediate concentration test possible is the one that revealed the real nature of each model. In figure 19 it is observable that all three models presented a periodic interconnected cylindrical structure. In the case of M_L and M_I , the stable configuration shown bridge like interconnections which seemed to not only maintain the structure but at the same time leave space to solvent flow through it. Notwithstanding, M_H preferred a complete plane to interconnect the cylinders blocking solvent flow through the membrane. Furthermore, the cutting planes expose that inside the cylindrical structure M_L kept forming randomly distributed short planes while the others attempted to complete the structure with another concentric cylinder like aggregate.

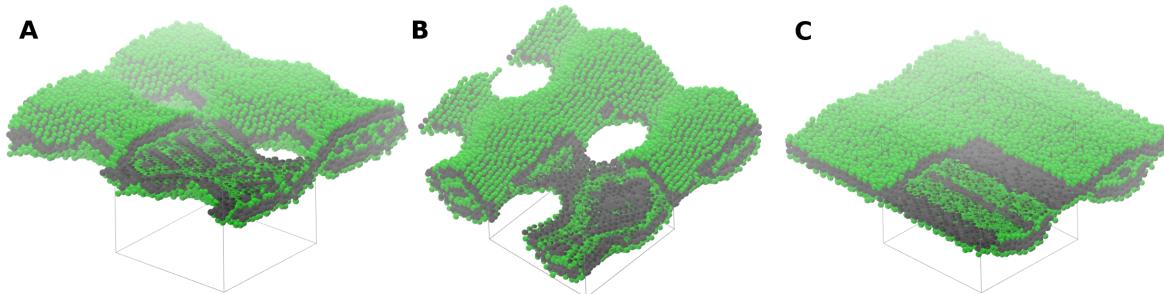


Figure 19: Snapshot of simulation boxes at intermediate concentration. The model used in each box is A) M_L , B) M_I and C) M_H . Periodic replicas are provided to ease mesostructure visualization. The original images are the ones located inside the grid box. The original box was divided by a cutting plane, showing internal conformation of the aggregate. N beads are represented in green, C beads in black.

Lastly, the high concentration test induced complete periodic planes in all cases. Even for M_L which in all other cases preferentially formed short planes, at this concentration stabilized forming long planar lattice structures. Contrastingly, instead of a mesh the other two models favoured parallel planes. Another noticeable feature for M_L and M_H shown in Figure (20) is the existence of a interplanar region with high density of short planes arranged in alternative directions. However, for M_H case there is only one direction for all planes and as expected, for M_L they appear to be randomly distributed. Exceptionally, M_I does not presented this region forming three planes instead.

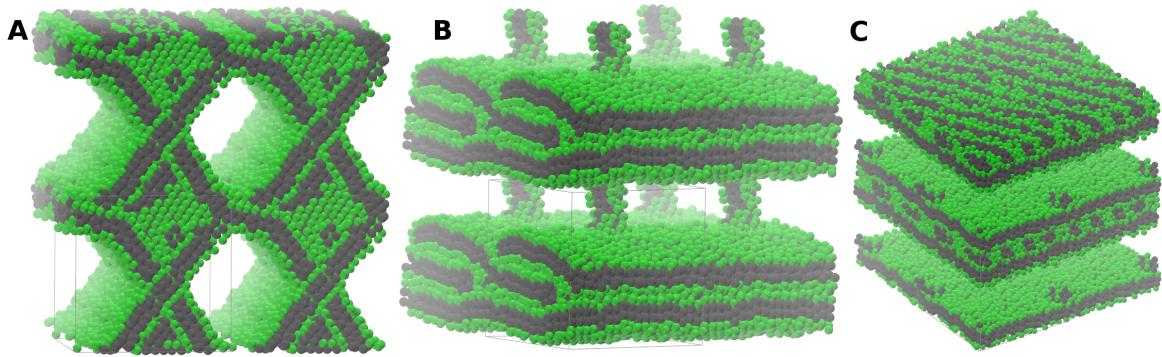


Figure 20: Snapshot of simulation boxes at high concentration. The model used in each box is A) M_L , B) M_I and C) M_H . Periodic replicas are provided to ease mesostructure visualization. The original images are the ones located inside the grid box. N beads are represented in green, C beads in black.

In general, M_L and M_I presented exposure of the apolar part of entire layers to the solvent media. The reason is because the reference simulation presented this characteristic for both models thus this artefact was carried to the coarse-grain model. Differently from the others, M_H presented exposure of apolar parts only in the last case, which was expected since the all-atoms reference itself presented spare molecules over the layer at high concentration. Hence, once again an ideal concentration for the reference simulation seems to be an important factor since apparently undesired characteristics will be carried to the Coarse-grain model.

7 Conclusion

In this project, a Coarse-grain model was created following the systematic method provided by MagiC package. The efficiency and limitations of this model were evaluated using comparisons between several models. Moreover, several techniques regarding the usage of MagiC were developed.

Regarding the models, either the complex and the simpler model presented a satisfactory efficiency to represent self-assembly of amphiphilic molecules, being capable of reach outstanding system sizes and long time horizons. The interchangeability test revealed that the implicit water condition can be rather explored within a certain limit since the models demonstrated common characteristics for different concentrations. However, it still necessary to determine an ideal All-Atom reference in order to avoid unrealistic behaviours. Another important characteristic derives from the angular potential plot in which indicates that an even simpler model can be proposed. The reason is because the DMDD molecules seldom presented significant angular bending in other locations rather than the central region, hence this fact suggests that a three bead model could provide enough approximation to this surfactant.

Due to the enormous discrepancy between number of Monte Carlo steps per Inversion step required for M_2 , M_1 and M_L , it is possible to conclude that there is a direct relationship between model and reference complexity to the effort necessary to obtain a potential set. As long as implicit water model need to carry information concerning long-range interactions, reducing the size of the reference simulation to reduce the complexity during Inversion process is not a suitable solution. Instead diminish model complexity or make the proposed gradual modelling seems to be better options. Nevertheless, both possibilities need to be better evaluated by this means enabling the creation of more complex scenarios, such as with silica addition or ionic environments.

In summary, this project provided reasonable introduction to molecular multi-scale modelling and simulation, analysing vastly the characteristics of the DMDD without silica addition, opening terrain to further studies concerning nanoporous silica materials.

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
- 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. (2014a). Magic: User manual. version 1.0.3 [Computer software manual].
- Mirzoev, A., & Lyubartsev, A. P. (2014b). Systematic implicit solvent coarse graining of dimyristoylphosphatidylcholine lipids. *Journal of Computational Chemistry*, 35(16), 1208–1218. Retrieved from <http://dx.doi.org/10.1002/jcc.23610> doi: 10.1002/jcc.23610
- 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].
- 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

Model 1 description files

CGTraj input:

```
&TRAJ
NFORM='PDBT',
FNAME='IdealAA300.pdb',
PATHDB='./',
NTYPES=2,
NSPEC=30,670,
NAMOL='dmdd', 'tip4p',
NFBEG=1,
NFEND=1,
ISTEP=1,
IPRINT=7,
&END
BeadMapping
CGTrajectoryOutputFile:cgtrajM1.xmol
CGMolecularType:DMDD
  ParentType: dmdd
    N1:9:1,2,3,4,5,6,7,8,9
    C1:12:10,11,12,13,14,15,16,17,18,19,20,21
    C2:12:22,23,24,25,26,27,28,29,30,31,32,33
    N2:9:34,35,36,37,38,39,40,41,42
EndCGMolecularType
CGMolecularType:tip4p
# No beads assigned to this type, so the water will be excluded
EndCGMolecularType
EndBeadMapping
```

rdf-2.0 input:

```
&TRAJ
NFORM='XMOL',
FNAME='cgtrajM1',
PATHDB='.',
NTYPES=1
NAMOL='M1dmdd',
NSPEC=30,
NFBEG=1,
NFEND=1,
ISTEP=1,
IPRINT=7
&END
&RDFIN
FOURDF='M1dmddIdeal',
RDFCUT=20.,
RMII=0.0,
RMAX=10.,
NRDF=3,
NRDFI=2,
NADF=1,
NATOMTYPES=2,
DELTAR=0.1,
DELTARI=0.02,
DELTAPHI=1.0,
&END
# Different CG-type names
N:N1 N2
C:C1 C2
```

```

# List of Intermolecular RDFs
N—N
& 3
 1 1
 1 4
 4 4
N—C
& 4
 1 2
 1 3
 4 2
 4 3
C—C
& 3
 2 2
 2 3
 3 3
# List of Intramolecular RDFs
N—C 1
& 2
 1 2
 4 3
C—C 1
 2 3
# List of Intramolecular ADFs
N—C 1
& 2
 1 3 2
 4 2 3

```

Model 2 description files

CGTraj input:

```

&TRAJ
NFORM='PDBT',
FNAME='IdealAA300.pdb',
PATHDB='./',
NTYPES=2,
NSPEC=30,670,
NAMOL='dmdd', 'tip4p',
NFBEG=1,
NFEND=1,
ISTEP=1,
IPRINT=7,
&END
BeadMapping
CGTrajectoryOutputFile:cgtrajM1.xmol
CGMolecularType:DMDD
  ParentType: dmdd
    N1:6:1,2,3,4,5,6
    CE1:6:7,8,9,10,11,12
    CI1:6:13,14,15,16,17,18
    CI2:6:19,20,21,22,23,24
    CI3:6:25,26,27,28,29,30
    CE2:6:31,32,33,34,35,36
    N2:6:37,38,39,40,41,42
  EndCGMolecularType
  CGMolecularType: tip4p
# No beads assigned to this type, so the water will be excluded
  EndCGMolecularType

```

EndBeadMapping

rdf-2.0 input:

```
&TRAJ
NFORM='XMOL',
FNAME='cgtrajM2',
PATHDB='.'
NTYPES=1
NAMOL='M2dmdd',
NSPEC=30,
NFBEG=1,
NFEND=1,
ISTEP=1,
IPRINT=7
&END
&RDFIN
FOURDF='M1dmddIdeal',
RDFCUT=20.,
RMI=0.0,
RMAX=10.,
NRDF=6,
NRDFI=3,
NADF=3,
NATOMTYPES=3,
DELTAR=0.1,
DELTARI=0.02,
DELTAPHI=1.0,
&END
# Different CG-type names
N:N1 N2
CE:CE1 CE2
CI:CI1 CI2 CI3
# List of Intermolecular RDFs
N—N
& 3
    1   1
    1   7
    7   7
N—CE
& 4
    1   2
    1   6
    7   2
    7   6
N—CI
& 6
    1   3
    1   4
    1   5
    7   3
    7   4
    7   5
CE—CE
& 3
    2   2
    2   6
    6   6
CE—CI
& 6
```

```

2 3
2 4
2 5
6 3
6 4
6 5
CI—CI 1
& 6
3 3
3 4
3 5
4 4
4 5
5 5
# List of Intramolecular RDFs
N—CE 1
& 2
1 2
7 6
CE—CI 1
& 2
2 3
6 5
CI—CI 1
& 2
3 4
4 5
# List of Intramolecular ADFs
N—CI 1
& 2
1 3 2
7 5 6
CE—CI 1
& 2
2 4 3
6 4 5
CI—CI 1
3 5 4

```

Example of GROMACS input parameters for reproduction tests

;cpp	= /usr/bin/cpp
constraints	= none
integrator	= sd
dt	= 0.001 ; ps ! 5 fs
nsteps	= 350000000; 50 ns
tau_t	= 1.0 ;! ps
ref_t	= 298.0 ;! K
tc_grps	= System
comm_mode	= Linear ;! Remove COM move and rotation around
nstcomm	= 1000
nstlist	= 10 ;! Frequency to update neighbor list
nstxout	= 2000 ;! Frequency to write output coord
nstvout	= 10000 ;! Frequency to write velocities
nstfout	= 10000 ;! Frequency to write forces
nstenergy	= 2000 ;! Frequency to write energy
ns_type	= grid ;! Way of creating neighbor list
pbc	= xyz
rlist	= 1.8 ;! Cutoff distance for short-range neighbor list
coulombtype	= cut-off ; rcoulomb=rlist

```
fourierspacing      = 0.35
pme_order          = 5
rcoulomb           = 1.8 ;! Real-space electrostatic cutoff
vdwtype            = User
rvdw                = 1.8 ; Same as R_cutoff in MDynamix
; dispcorr          = Ener
energygrps         = N CE CI
energygrp_table    = N N N CE N CI CE CE CE CI CI CI
epsilon_r           = 78 ; ! relative dielectric constant
table-extension    = 0.5
```