



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

April 21, 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 Experimental Methods</b>	<b>2</b>
4.1 Theoretical Basis . . . . .	2
4.2 Methods Description . . . . .	7
<b>5 Results</b>	<b>11</b>
5.1 GROMACS simulations . . . . .	11
5.2 Experiment 1 . . . . .	12
5.3 Experiment 2 . . . . .	13
5.4 Experiment 3 . . . . .	15
<b>6 Discussion</b>	<b>15</b>
<b>7 Conclusion</b>	<b>15</b>
<b>8 Nomenclature</b>	<b>15</b>
<b>9 References</b>	<b>15</b>
<b>10 Appendix</b>	<b>1</b>

## 2 Acknowledgements

### 3 Introduction

Molecular Dynamics (MD) and Monte Carlo (MC) have been powerful tools to simulate molecular interactions of surfactants in solvent systems, allowing a deeper understanding of self-assembly process of this sort of material (Need & Someone, 2025). Therefore, these structural conformations are specially useful to design bio-inspired silica materials (Patwardhan, 2011) since with the addition of silica they remain behaving 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), HR (Jin et al., 2008), MSU-V (Tanev, Liang, & Pinnavaia, 1997) and many others, in which the self-assembly structure depend on the type of surfactant and concentration of the substances involved. Moreover, 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), thus initial self-assembled conformations are predicted as a reflex of the final results and little is known about the mechanistic of this process. However, by using MD simulations it is possible to observe and analyse these initial steps of self-assembly and further predict, with more accuracy, properties and framework provided by surfactant (Tresset, 2009) to generate silica structures.

A major concern is that even though MD uses sophisticated software prepared to simulate enormous 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 hardly can achieve long time horizons and are commonly limited to a few  $\mu s$  depending on the size of the system. For this reason, several techniques have been develop to optimize the performance of the simulations such as coarse-grain models (Need & Someone, 2025). Briefly, the idea behind using this technique is to fit parameters of heavy atoms groups with similar properties in a bigger "bead", which includes atomic masses and electrostatic charges lumped in approximated values. For example, by merging carbon heavy atoms, that means carbon atom and its hydrogen atoms, at a lipid tail in groups of three or four since they have similar hydrophobic properties.

In order to provide topologies to GROMACS simulations several force-fields, for exemple MARTINI (Marrink, Risselada, Yefimov, Tieleman, & de Vries, 2007), use 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, Gomes, & Jorge, 2013) with this method recreated a model of surfactant in 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 focus of many studies (Pronk et al., 2013).

Different concepts has been applied to develop a suitable model for implicit solvents, studies conducted by 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 & Lyubartsev, 2013) is able to recreate a 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 on the approximated potentials. Therefore, changes in parameters as concentration may not affect coarse-grained model performance on recreating self-assembly structures (Mirzoev & Lyubartsev, 2014b).

The work presented on following experiments are an attempt to create a cross-link method to upscale silica crystal-liquid phase 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 MD simulation software, in this case GROMACS (Pronk et al., 2013), will provide a suitable approximation to self-assembly of amphiphilic molecules in presence of silica. Further explanations of the process are described on Experimental Methods section. For the scope of this project, a bolaamphiphilic molecule 1,12-diaminododecane (DMDD) has been chosen as surfactant because that as seen in previous research by Tanev et al. (1997) it self-assembly in multilamellar vesicles that in presence of a silica precursor is capable of generating a mesoporous structure with remarkable properties. In order to validate this structure formation and further framework formation for silica oligomers, a coarse-grain approximation is suitable option since amphiphilic molecules interaction with

solvents can be described efficiently with tabulated Lennard-Jones potentials interactions. As a final objective at the end of this project, a 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 good approximation to interactions with silica oligomers in a mesoscale model.

## 4 Experimental Methods

### 4.1 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 of many systems, in this case thermodynamic equilibrium of molecular systems. Following sections will describe the mechanistic behind each technique and further explain how they are used in simulation software.

#### Molecular Dynamics simulation

Molecular Dynamics is a simulation method that originates from 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 follow differential equation:

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

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 derivatives 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 Eqn.(1) and Eqn.(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 on Verlet method, another approach can be made by accounting 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 preciser step cycle with more stability (Sato, 2010). This set of equations are called "Velocity Verlet method" and they are 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 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 describe Lennard-Jones potentials in 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. Molecular dynamics technique enable 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.

### Monte Carlo simulation

Monte Carlo is a simulation method based on statistical probability of existence of a system at thermodynamic equilibrium. This happens when 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 free Helmholtz energy of the 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 to give a simplified idea assume 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 using a probability distribution function (Sato, 2010):

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

Whether the system configuration is generated with considerable number of microstates that satisfies this probability, the final configuration would have a real physical meaning. But, as soon as it is almost impossible to define an analytical solution to this equation if  $\mathcal{N}$  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:

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

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 such things (Sato, 2010). Finally, now that both molecular simulation methods had been properly 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.

## 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 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 system using supercomputing or clusters, but also it can run small simulations in any ordinary computers with great performance.

In order to interpret interatomic interactions, that means forces between each pair of atoms, GROMACS uses pre-defined parameters called Topologies in which the user can describe almost any molecule using force fields or tabled potentials. Data given in this description includes information regarding atoms' type, mass and charge. Furthermore, it includes detailed description of each 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 interactions potential, hence their applicability is suitable to almost any system from all-atoms to coarse-grain descriptions it is just necessary for the user to describe a realistic model.

When dealing of simulations of complex systems, realism means more computational power expended trying to achieve it (Sato, 2010), therefore a major concern is how to achieve longer time horizons without losing 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-atomic 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:

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

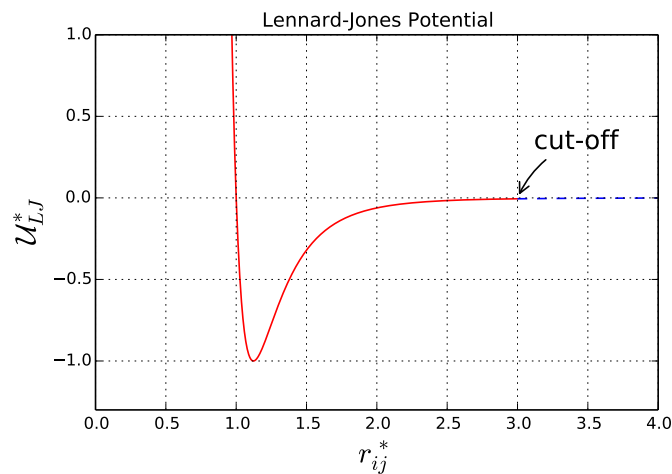


Figure 1: Example of Lennard-Jones Potential(with reduced units)



Where the forces used in Molecular Dynamics equations come from derivative 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, and non-bonded potentials from the summation of Lennard-Jones and short-range Coulomb potentials. As shown in Fig.(1) whether  $r_{ij} \rightarrow \infty \Rightarrow \mathcal{U}_{LJ} \rightarrow 0$  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 recreates a continuous media environment for simulation boxes bigger than two times the cut-off radius, since particles can not interact with itself.

Although speed efficiency is a concern it is not only the unique preoccupation, it is also necessary to control environment variables such as temperature and pressure, in order to maintain the simulation within physical limits and that demands computational resources. GROMACS is capable of using three types of thermostat which are 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 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, and for this reason the aim of this project is to demonstrate the capacity of this software to simulate realistic multi-scale molecular models.

### About MagiC

MagiC is 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 Fig.(2) method consists in 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 tabled potentials and a Topology for a coarse-grained model that uses these potentials to run Molecular Dynamics simulations using GROMACS.

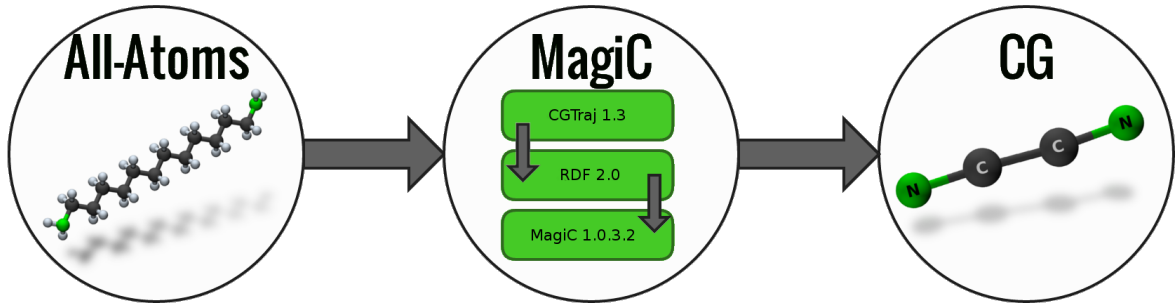


Figure 2: MagiC

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 assign 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 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.

In the kernel is where the actual inversion precesses occurs, as schematised in Fig.(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 on the next inversion step.

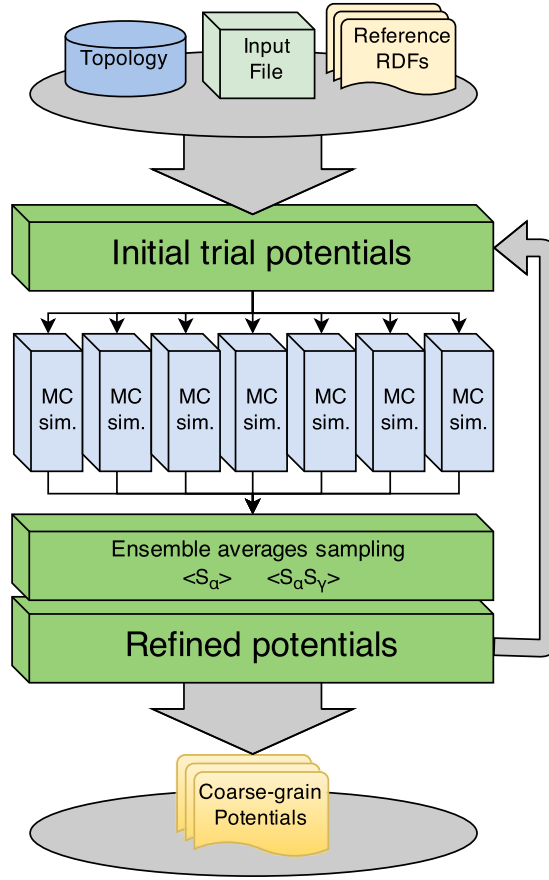


Figure 3: MagiC kernel work-flow scheme

The Iterative Boltzmann Inversion is a reasonable methodology to apply in the initials 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 regulation parameter to avoid excessive variations. For a system in a canonical ensemble, temperature will be constant, hence the 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 deviance 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 Inversion 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  $\Delta r = \frac{r_{\text{cut}}}{\mathcal{M}}$ , obtaining  $\mathcal{M}$  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:

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

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

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 becomes useful for inverse Monte Carlo. Since this method deals with pairwise interactions the value 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:

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

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, MagiC has a built-in capacity of running parallel Monte Carlo simulation as it can be seen on Fig.(3), which enable the user to reach the order of billions of simulation steps in a feasible timespan. Moreover, there is not an unique solution for a given RDF (Moore et al., 2014) and eventually solution can possibly diverge at a certain point. Therefore, a satisfactory result depends on analysis of all outputs from inverse steps in order to guarantee convergence. Further explanations regarding functionalities and parameters for MagiC and GROMACS will be provided the next section.

## 4.2 Methods Description

### GROMACS All-Atom simulations

The experimental process of this project is divided in 3 main parts: Firstly, develop a efficient coarse-grain model and validate its use. Secondly, analyse the relationship between concentration of all-atoms simulation and concentration of coarse-grained simulation. Then finally, create and reproduce a Silica surfactant mesoscale system.

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. The 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 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 system.

Initially, all boxes received a energy minimization step in order to avoid system blow-up deal 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. After, an equilibration step was necessary to reach desired initial temperature and pressure conditions for the MD simulation. During equilibration temperature coupling was kept at 298 K using a v-rescaling thermostat (?, ?) with time constant of 0.01 ps and pressure coupling was kept at 1 bar using a Berendsen barostat (?, ?) with time constant of 0.5 ps. The 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, four all-atom boxes were created with composition and size described in Table (1).

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

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 <sup>a</sup>	240	-	28.2972	3.30
B4	30	670	30	???	???

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

at 298 *K* by a Nosé-Hoover thermostat (?, ?) and pressure kept at 1 *bar* by a Parrinello-Rahman barostat (?, ?). 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 because during this time the system still in equilibration, and therefore that would affect reference RDF values.

### Experiment 1

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

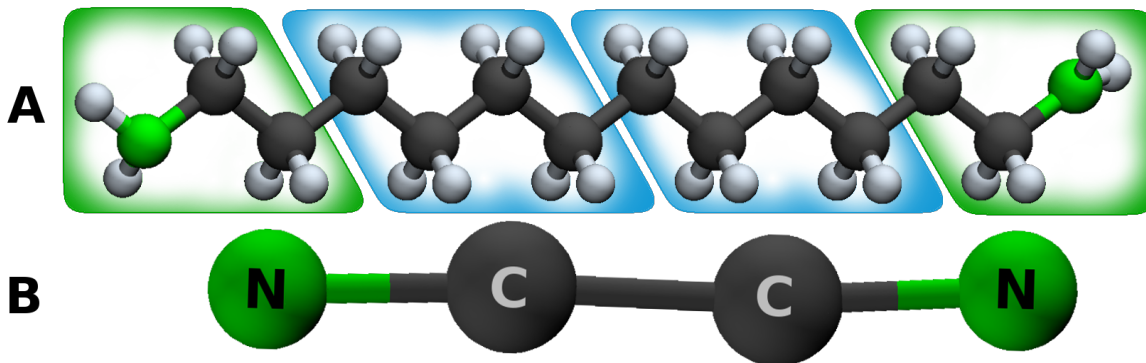


Figure 4: A) Schematic of M1 molecule split plan, Carbon atoms represented in black, Nitrogen in green and Hydrogen in gray. B) Final coarse-grain model for M1. This model aggregated the three most polar heavy atoms on the tips and four apolar heavy atoms at the center.

The differences between M1 and M2 are 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 final charge remain 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 deal to number of beads. The All-Atom simulation box chosen for this experiment was B1, because at this experiment concentration of reference was not a major concern, more information regarding this point is given at Experiment 2 (see Section 5.3)

In order to fully describe the system, MagiC generates potential tables for interaction between each bead. The first type of table are for each bound type, for example on M1, there are two different tables: one for N-C bond and other for C-C bond. However, as soon as the molecule is symmetric the two external N-C bonds uses the same potential table. Secondly, it generates tables for each angle type, following the symmetry rule. And finally, each type interatomic pair interaction has its own table, i.e. on M1 there are three: N-N, N-C and C-C. A more detailed description for both models is exposed in Table (2).

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 each

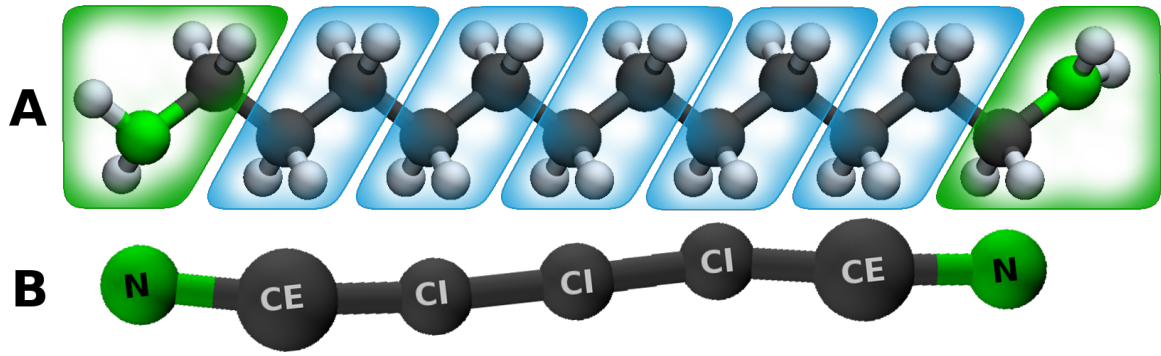


Figure 5: A) Schematic of M2 molecule split plan, Carbon atoms represented in black, Nitrogen in green and Hydrogen in gray. B) Final coarse-grain model for M2. Heavy atoms were grouped pairwise in such a manner that amphiphilic parts were kept as distinct as possible. It is important to notice the necessity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

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]	[N- $\widehat{CE}$ -CI]	[N-N] [N-CE]
	[CE-CI]	[CE- $\widehat{CI}$ -CI]	[N-CI] [CE-CE]
	[CI-CI]	[CI- $\widehat{CI}$ -CI]	[CE-CI] [CI-CI]

bead. These informations must be part of input file of "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 "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 inversion process for each potential table. Additional information regarding model description in input files for M1 and M2 can be found on 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 also analyses the self-assembly of amphiphilic molecules. Nevertheless, considering that DMDD owns a much 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 (3) exposes the specific number of steps per core and whether it uses IBI or IMC method for inversion. Since the water is implicit, the dielectric constant for Monte Carlo simulations was set to experimental dielectric constant of water at 298 K and 1 bar which is  $\epsilon = 78$  (? , ?).

Table 3:  
Inversion  
Process  
description

a	b
c	d

Finally, after each inversion step MagiC generates RDF based on ensemble averages calculated during Monte Carlo simulations and then it calculates the deviance between reference RDFs and inversion RDFs. When the deviance 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 is null, PME long-range electrostatics were

replaced by cut-off scheme which increased simulation speed. An example of reproduction test input parameters can be found on Appendix (see Section 10).

## Experiment 2

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 an NVT ensemble, because deal to the absence of solvent, pressure has no physical meaning in these systems. 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 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 begins 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 ( $M_L$ ).

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, in 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, we used as a basis the final potential obtained for  $M_L$ . Further explanations regarding this technique are shown in the Discussion section (See Section 6).

After, with the three models, a comparison test was made in order to observe wheter the systems 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 (4). All the simulation parameters were the same as the reproduction tests of Experiment 1, however now the simulations were run using a higher timestep of  $20 \text{ fs}$  until the moment that self-assembly structure had appeared to achieve a stable configuration.

Table 4: 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
Time (ns)									



### Experiment 3

Finally, this last part of the project is where silica is introduced to the system. For this propose an All-Atoms Silica oligomer molecule was developed based on OPLS-AA force field. The preferred Silica model have been successfully applied in previous works done by (Pérez-Sánchez et al., 2013), therefore it provided a reliable representation of silica-surfactant interactions.

- MagiC CG silica topology
- gromacs Set-up Self-assembly Model X + Si
- 1000, 10000 (Maybe more?)

## 5 Results

### 5.1 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 received its characterization based on their periodicity and Radial Distribution Function.

As a initial approach, a qualitative evaluation of the simulations is provided in Fig.(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.

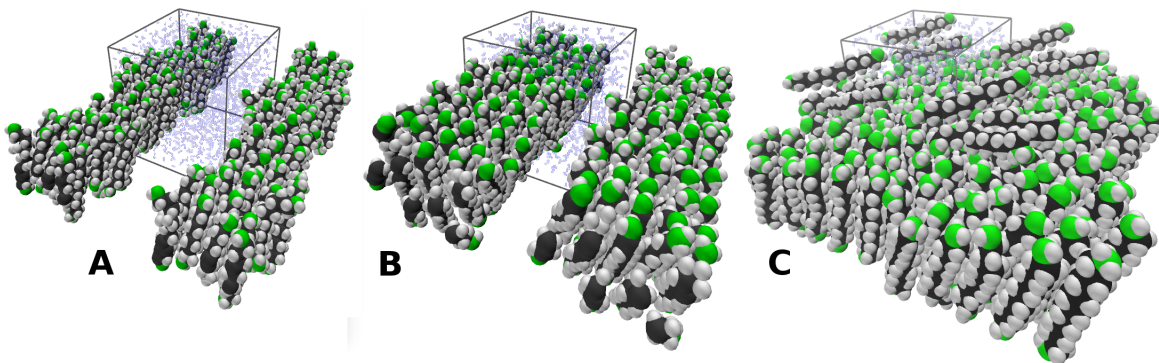


Figure 6: Snapshot of simulation boxes with periodic replicas. a) box at low concentration (B1), b) box intermediate concentration (B2) and c) box at high concentration (B3). The original images are the ones located inside the grid box surrounded by water. Nitrogen atoms are represented in green, Carbon 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 Fig.(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.

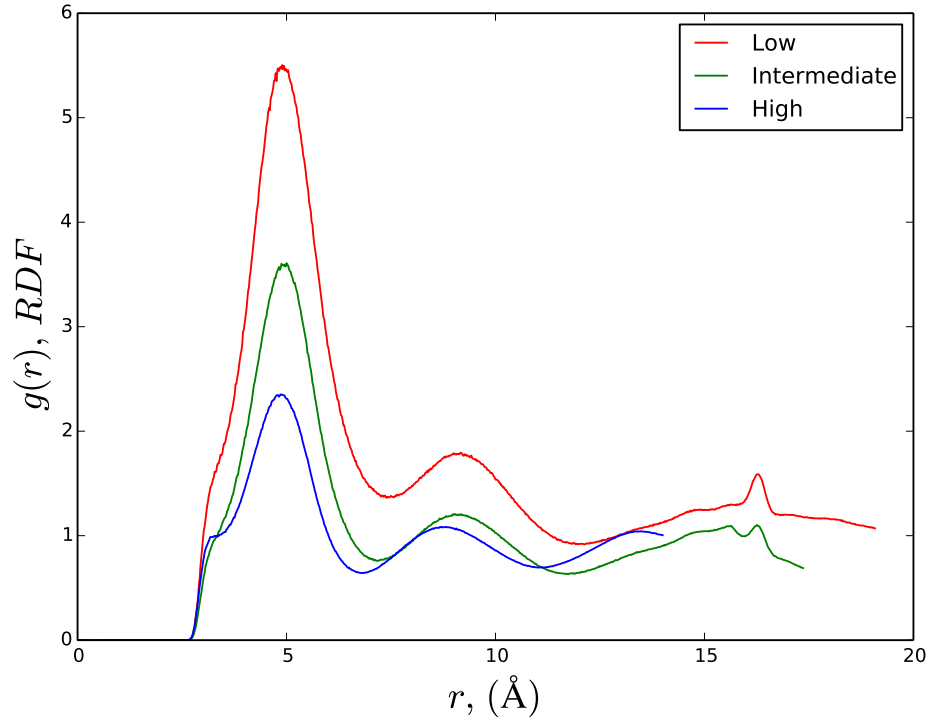


Figure 7: Radial distribution function between N-N atoms.

## 5.2 Experiment 1

- Model 1 vs Model 2
- gromacs Simulation
- IMC process Convergence
- gromacs Reproduction Tests RDFs and Properties



### 5.3 Experiment 2

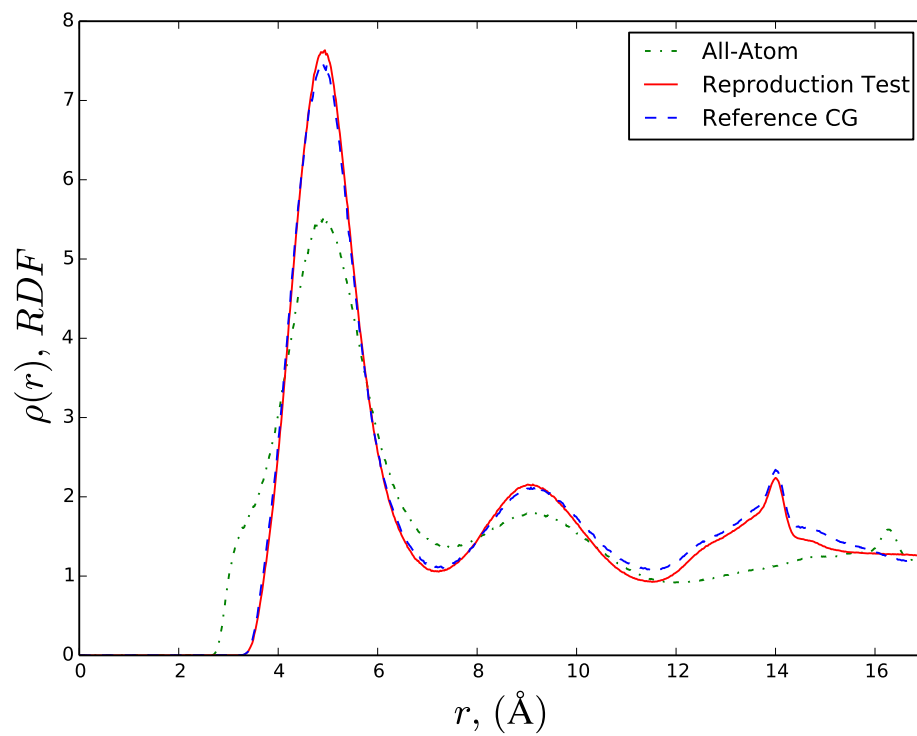


Figure 8: rdfLow

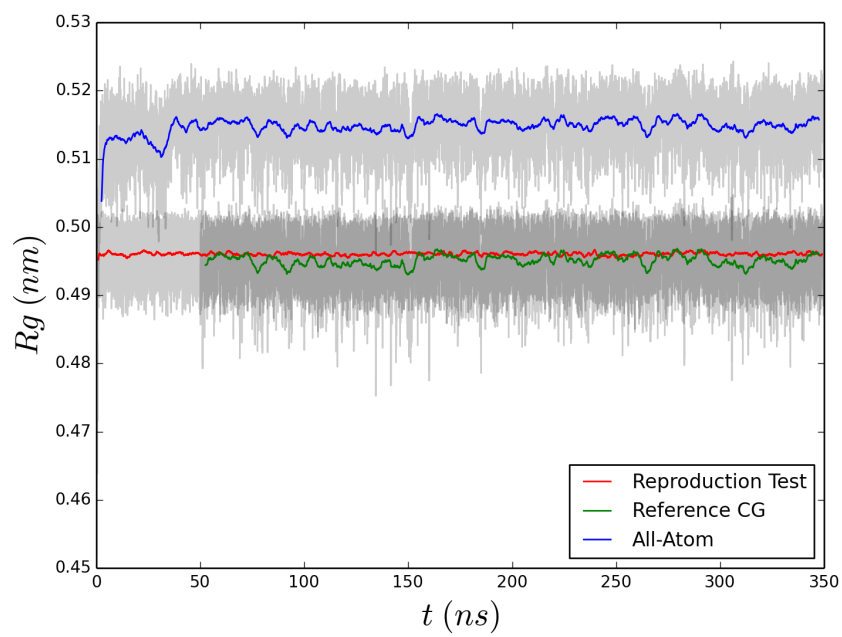


Figure 9: Gyration Low

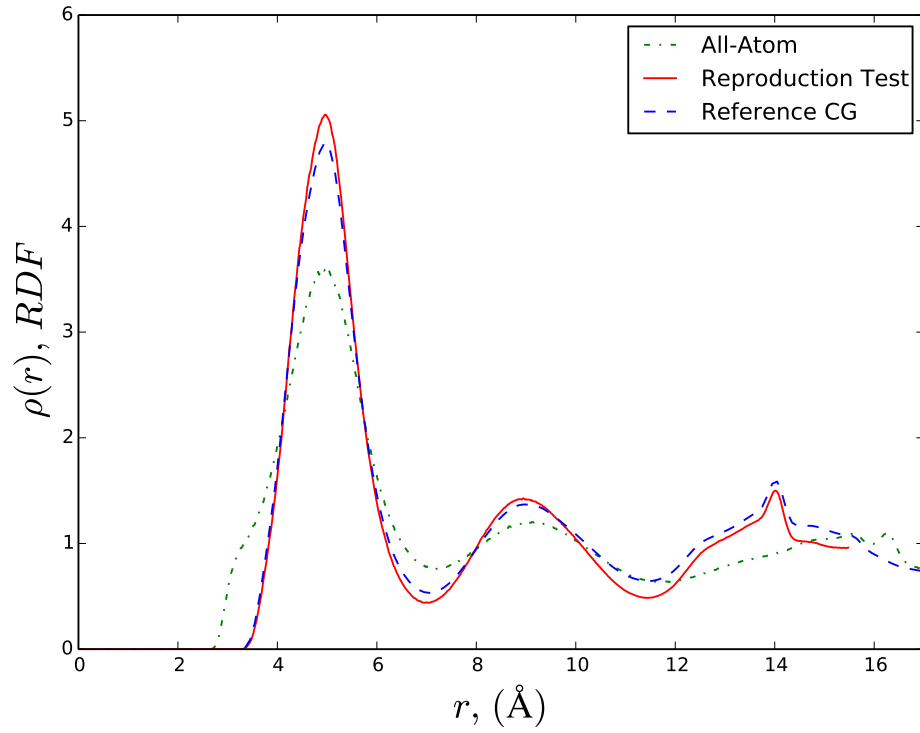


Figure 10: rdfIdeal

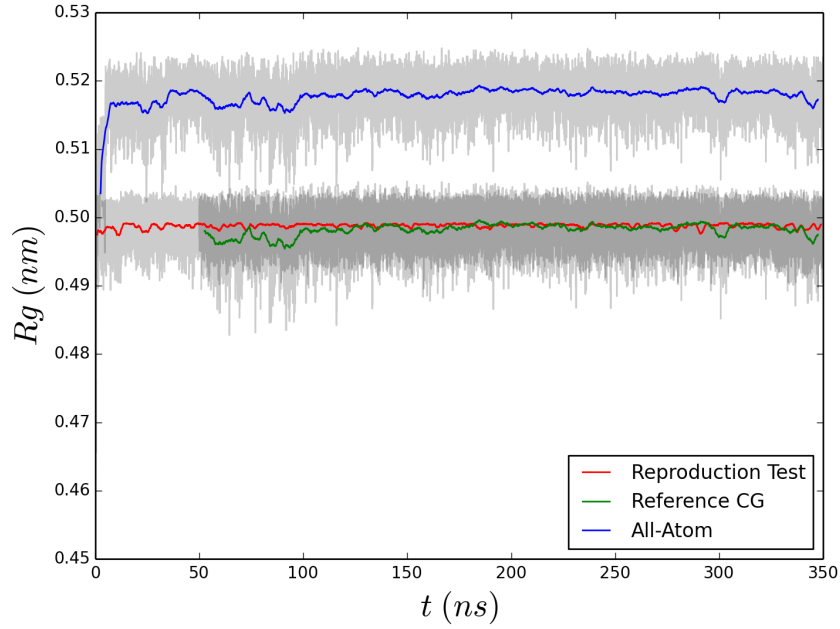


Figure 11: Gyration Ideal

- Self-assembly vs concentration Model X
- Ideal con validation: Low, Ideal, High con comparison
- Ideal @ Low -> validate
- Ideal @ Ideal -> evaluate
- Ideal @ High -> validate

- Extremes evaluation:
- Low @ Low vs High @ Low -> compare
- High @ High vs Low @ High -> compare

## 5.4 Experiment 3

- Silica addition
- CG silica Properties
- Self-assembly Model X + Si
- 1000, 10000 (Maybe more?)

## 6 Discussion

Experiment 1

- Which is the best Model?
- Why Model X is better than Model Y?
- To what extent Model X represent well the AA model?
- What are the limitations and advantages of Model X?

Experiment 2

- What is the relation between the entropy "error" and concentration?
- Is the concentration really an problem?
- Do you need to run an AA for each con or the Ideal can represent any?
- Even extreme cases can be approximated? In this case how much is the "error"?

Experiment 3

- Is The silica CG model suitable?
- Can you the proprieties validate it?
- The self-assembly behaviour change with silica addition?

## 7 Conclusion

What did you really see from the results?  
 There was any bad assumption that is contestable?  
 What will be the next step to your research?

## 8 Nomenclature

$\mathcal{H}$	Hamiltonian
$k_B$	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
- Frenkel, D., & Smit, B. (2001). *Understanding molecular simulation: From algorithms to applications*. Elsevier Science.
- Jin, H., Qiu, H., Sakamoto, Y., Shu, P., Terasaki, O., & Che, S. (2008). Mesoporous silicas by self-assembly of lipid molecules: Ribbon, hollow sphere, and chiral materials. *Chemistry – A European Journal*, 14(21), 6413–6420. Retrieved from <http://dx.doi.org/10.1002/chem.200701988> doi: 10.1002/chem.200701988

- 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>
- Need, A., & Someone, T. C. (2025). Mock article: You need to read more. *Journal of Void Theory*, 1(1), 151-1510.
- 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.
- Tanev, 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].

## 10 Appendix