



**UNIVERSITÀ
DEGLI STUDI
DI GENOVA**

FACOLTÀ DI FISICA
DOTTORATO IN FISICA COMPUTAZIONALE DELLA MATERIA

**Multiscale modeling of synthetic and
biological supramolecular systems**

Candidate:
Anna Lucia de Marco

Advisors:
Prof. Giulia Rossi
Prof. Giovanni Maria Pavan

Academic Year 2020-2021



**UNIVERSITÀ
DEGLI STUDI
DI GENOVA**

FACOLTÀ DI FISICA
DOTTORATO IN FISICA COMPUTAZIONALE DELLA MATERIA

**Multiscale modeling of synthetic and
biological supramolecular systems**

Candidate:
Anna Lucia de Marco

Advisors:
Prof. Giulia Rossi
Prof. Giovanni Maria Pavan

Academic Year 2020-2021

UNIVERSITÀ DEGLI STUDI DI GENOVA
FACOLTÀ DI FISICA
DOTTORATO IN FISICA COMPUTAZIONALE DELLA MATERIA
Via Dodecaneso 33 – 60131 Genova (GE), Italy

A Stefania e Marcello

Chapter 1

Introduction

1.1 Molecular modeling and molecular dynamics

In recent years, the development of synthetic objects designed to interact with biological molecules or possess bio-inspired properties has become increasingly common in biomedicine and materials science. For instance, functionalized nanoparticles (NPs) of nanometric size can find applications in many biological systems as they can directly interact with them [1],[2], [3]. At the same time, supramolecular polymers can be designed to mimic the stimuli-responsiveness and self-repairing properties of biological materials [4]. In both cases, a profound understanding of the physical processes involved requires investigating the interaction between the system components at high spatial and temporal resolutions, which is usually challenging. For example, elucidating the behavior of an NP at the interface with a lipid membrane is experimentally prohibitive [5]. The same argument applies to the case of supramolecular polymers (SPs) [6]. The latter are one-dimensional aggregates with fascinating intrinsic dynamics almost inaccessible through experiments. Even if experiments can provide powerful insights, they struggle to reach the resolution needed to highlight the mechanisms involved in the self-assembly of complex materials or the phenomena happening at biological interfaces.

In this regard, computational approaches, especially those based on molecular dynamics (MD) simulations, offer a wide range of opportunities. Classical MD simulations are appropriate when electronic degrees of freedom can be ignored. They can access processes occurring on time and length scales of microseconds and tens of nanometers respectively, while maintaining single-atom resolution. In this way, one can observe the molecular mechanisms of interactions and follow the dynamics of the systems as if he is near the action. However, the reliability of every computational result depends on the quality of the model. Indeed, computational approaches based on MD simulations must utilize suitable molecular models with an appropriate description of their interactions to generate meaningful simulations [7]. In MD simulations, the ensemble of functional forms and parameters describing the intramolecular and intermolecular interactions is called a force field (FF). The FF parameterization heavily impacts the simulation outcome's reliability. The FF functional forms and parameters, together with the topologies of the molecules, are

usually selected to reproduce specific structural and thermodynamic features of the system. However, the main challenge is to find the best compromise between the model's precision and the simulation's computational efficiency.

Force fields used in classical MD simulations can be divided into two main groups: fully atomistic and coarse-grained (CG) [8]. In a fully atomistic FF, atoms are the basic interacting units, providing a high level of detail. However, one must consider that the system's time evolution is computed based on interactions between all atoms. As the size of the system increases beyond a few nanometers, and the processes of interest occur on the microsecond timescale, atomistic models become computationally demanding, and the computing power required could become unattainable. For example, increasing the size of an atomistic bilayer membrane from 30 nm to 300 nm would require a million times more computing work to equilibrate the system [9]. Coarse-grained models aim to overcome this problem by simplifying the system description. This result is achieved by grouping several atoms, typically belonging to well-defined chemical groups, into single interaction centers called beads. Despite the unavoidable loss of precision, CG models offer significantly higher efficiency, providing access to larger spatio-temporal scales and, thus, to processes that are impossible to study with fully atomistic models.

1.1.1 The power of CG models

In recent decades, CG models have gained tremendous popularity in the scientific community, becoming essential tools for biomolecular and nanotechnological research. This popularity is due to their ability to overcome the spatio-temporal limitations of all-atom models discussed above [10], [11]. Compared to fully atomistic models, CG models can be orders of magnitude faster. It is important to note that the speedup achievable depends strongly on the specific CG mapping procedure of the system and the parameterization of the particular CG FF adopted. However, a few general factors contribute to this speedup, the most immediate of which is the reduction of degrees of freedom (DOFs). Reducing the number of particles by a factor of n (easily around 10) leads to fewer forces to compute, resulting in simulations that are faster by a factor of n^2 (and thus quickly two orders of magnitude). The reduction factor n can become much more prominent in the case of implicit solvent models. Reducing DOFs also reduces friction and smooths the potential energy surface, allowing the system to explore more phase space in the same amount of time. Furthermore, the smoother energy surface allows for larger time steps. With CG FF, it is common to use time steps of tens of femtoseconds compared to 1-4 femtoseconds in all-atom MD simulations. Another factor contributing to the CG speedup is that most CG models account only for short-range interactions (typically with a cutoff distance of around 1 nm), avoiding the use of expensive techniques to include long-range electrostatic interactions.

All that glitters is not gold, i.e., CG models also have limitations and subtleties.

The interpretation of the simulation time in terms of physical time when using a CG model could be tricky and requires careful comparison with experimental data or full-atomistic simulations. Another common issue is the need for transferability, known as the representability problem [12]. There are different CG FF tuned on the system in hand or even on the mechanisms of interest, which are difficult to use in different contexts. The challenge is improving compatibility and developing models suitable for more than just a single class of molecules.

The Martini CG FF was created considering this need for transferability and became one of the most popular CG models used in biophysical and material science simulations [13]. The original idea was an extendable CG model based on modular building blocks (beads) which retain chemical specificity and whose interactions are mainly based on the hydrophobicity of the chemical group represented by each bead. At first, the Martini FF was built for lipids and later extended to model a vast class of biological and synthetic molecules, including proteins, polymers, supramolecular polymers, and many others [14], [15], [16]. The Martini FF was a perfect choice for this thesis, representing the best compromise between accuracy and efficiency for our case studies.

1.2 Case studies

In the present work, we exploit the Martini FF to investigate two molecular systems characterized by processes that are too long and complex to be studied with atomistic models. On one hand, the intrinsic dynamics of a one-dimensional self-assembled supramolecular polymer; on the other hand, the effects of the interaction between functionalized NPs and multi-component lipid membranes. The following paragraphs provide a general description of these two systems and the open questions we tackled.

1.2.1 Supramolecular polymers

Synthetic SPs research has become more and more fascinating during the last decade. SPs are one-dimensional assemblies of basic units, i.e., monomers, that can possess bio-inspired properties such as environmental adaptivity, stimuli-responsiveness, and self-healing abilities. These extremely captivating features come from supramolecular fibers continuously exchanging monomers with the solution and with other fibers [17], [18]. Indeed, unlike in polymers, in SPs, monomers connect via non-covalent interactions, and thus their bonds are reversible. The challenge here is to rationally design monomers to control the properties of the resulting self-assembled material. This task requires a profound understanding of the mechanisms underlying the self-assembly and the dynamics of the SP [19].

To experimentally look at these structures at submolecular resolution is particularly difficult; a diffuse technique is the cryo-TEM that freezes the self-assembled structures giving high-resolution images but compromising the understanding of the system

dynamics (Figure 1.1). Trying to capture the exchange of monomers between the fibers and the solution is even more arduous. A handy tool for this purpose is represented by molecular modeling and computer simulations, working together as a virtual microscope. Two main approaches have been adopted to study supramolecular polymers with MD simulations: the top-down approach and the bottom-up approach.

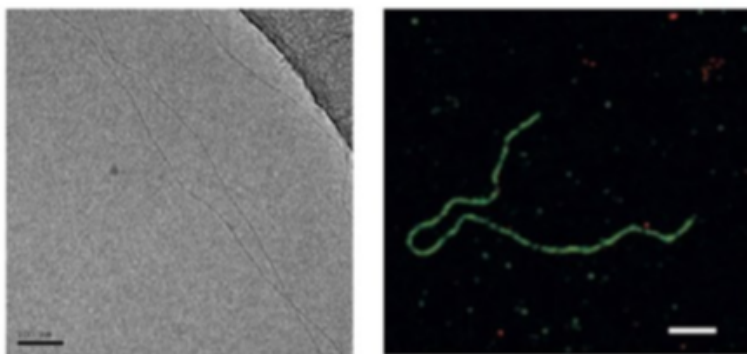


Figure 1.1: Experimental images of BTA supramolecular fibers. On the left the image obtained by cryo-TEM with a 100 nm scale bar (copyright 2013 The Royal society of chemistry) [20]. On the right, a sTORM image with a 1 μm scale bar (copyright 2014 The American Association for the Advancement of science) [21].

The top-down approach consists in building a pre-stacked assembly and then relaxing it through a MD simulation. This technique is appropriate when a reasonable initial configuration has already been suggested by experiments or by high-level simulations or when the assembling time would be too long to simulate. Top-down approaches illuminated the study of BTA (Benzene 1,3,5-TricarboxAmide) supramolecular polymers in water [22]. Experimentalists usually represented them by perfectly stacked and ordered monomers until AA-MD simulations showed a more complex and vibrant structure, while confirming the importance of hydrogen bonding [23]. Even if top-down approaches allowed the study of complex SPs at high resolution, they had several limitations. Indeed, it is known that complex molecular systems can remain trapped in meta-stable states and never sample the equilibrium states. In these cases, enhanced sampling techniques become fundamental, but the complexity of supramolecular systems modeled at the atomistic level makes their use challenging.

The bottom-up approach consists of starting from monomers randomly distributed in a simulation box to follow them while they undergo self-assembling into supramolecular fibers (Figure 1.2). Depending on the monomer complexity, all-atom simulations with this bottom-up approach can be extraordinarily demanding or even wholly unfeasible. In order to overcome this problem, CG models, combined with biased approaches, represent a possible solution. A biased approach made it possible to observe the BTA monomers forming ordered stacks in organic solvent (nonane).

Salvalaglio et al. studied the crystallization of urea both in water and in the presence of additives [24]. However, when the structure of monomers is more complex/flexible, the self-assembly process becomes too slow for the AA-MD simulation accessible timescales. In this case, a coarse-grained model for the monomer description becomes mandatory [25].

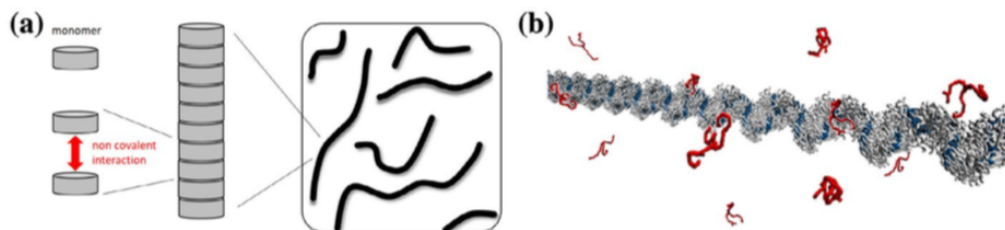


Figure 1.2: a) cartoon of a supramolecular polymer: fibers formed through non-covalent interactions. b) concept representation of the dynamical equilibrium between a supramolecular fibre and the monomers in its surrounding environment (solution).

During the last few years, the popular Martini CG FF has also been used for treating synthetic polymers and BTA supramolecular polymers. The CG model for BTA was opportunely refined to reproduce all the system key features: the strength of monomer-monomer interactions, the behavior of monomers in solution and the dependence of the order in the final structure from fine monomer details. The model provided equilibrium configurations for BTA fibers following both top-down and bottom-up simulations. The finely tuned CG BTA model allowed a direct study of BTA polymerization in water and revealed the presence of defects in the stacking [26]. Such defects are fundamental for the BTA fiber dynamics, and their role is the subject of Chapter 3, in which I present my original results about these systems. In this work, we combine coarse-grained modeling, enhanced sampling, and machine learning to investigate the key factors controlling the monomer exchange pathways in BTA supramolecular polymers. We demonstrate how the interactions between the monomers control the creation/annihilation of defects in the supramolecular fibers, determining the specific exchange pathway. The general nature of our CG models suggest a general approach to manage the key factors that control the dynamic behavior of such complex systems.

1.2.2 Nanoparticle-lipid membrane interaction

Ligand-protected gold nanoparticles (AuNPs) are extensively studied for their potential application in nanomedicine, e.g., drug/gene delivery and imaging [27], [28], [29]. In order to understand how NPs impact biological systems, it is fundamental to clarify the driving forces regulating the interaction between NPs and biologically relevant interfaces, specifically cell membranes (Figure 1.3) [30], [31].

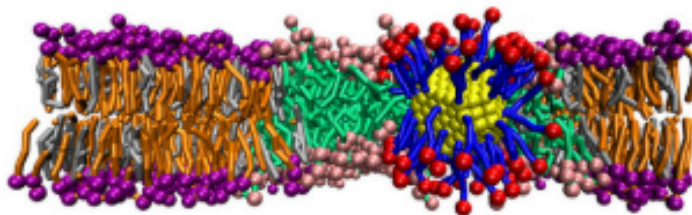


Figure 1.3: Cartoon of a ligand-protected NP penetrating a model lipid membrane. The color scheme is: Au core in yellow, hydrophobic ligands in blue with charged beads in red. DFPC in pink (headgroup) and green (tails), DPPC in purple (headgroup) and orange (tails). Cholesterol in silver [32].

Cell membranes are incredibly complex systems containing many lipid types and transmembrane proteins. In this framework, the challenge is to overcome the intrinsically multiscale nature of the biological environment and the lack of experimental data with sufficient resolution. Indeed, experiments probing the interaction of NPs with biological membranes or model membranes struggle to answer several questions concerning the effect of NP shape, size, surface charge, ligands arrangement, and toxicity [33], [34], [35]. One of the most investigated AuNPs type is the amphiphilic thiol-protected one. In the last decade, the AuNPs functionalized with 11-mercapto-1-undecanesulfonate (MUS) and 1-octanethiol (OT) (hereafter called MUS:OT AuNPs) give precious insights on the mechanism of the penetration into lipid bilayers, in fact they allow spontaneous and non-disruptive absorption into the cell, as well as the possibility of direct drug delivery [36], [37].

Over the last years, MD simulations have clarified many aspects of the NP-membrane interaction mechanism, mostly for single-phase model lipid membranes, i.e., bilayers containing a single or a maximum of two lipid types [38]. The combination of full-atomistic and CG MD simulations elucidates the way MUS:OT AuNPs passively translocate into lipid membrane: a multi-step process with at least three metastable states [39]. The first step is the so called *adsorbed state* when NP adsorbs on the top of the bilayer surface; this condition, as CG simulations predicted, is reversible. The protrusion of a lipid tail to the polar head region of the bilayer starts the second configurational change that is called *hydrophobic contact state*. After the first hydrophobic contact, the NP manages to shift the center of mass towards the membrane core, rearranging its ligands. The third state of the process, called *snorkeling state* is reached when a MUS ligand attaches to the lipid headgroups of the opposite leaflet of the bilayer. One after the other, MUS ligands bind to the opposite leaflet until the NP center of mass reaches the center of the membrane.

Simulations are based on simplified membrane models, such as bilayers containing a single or a maximum of two lipid types [38].

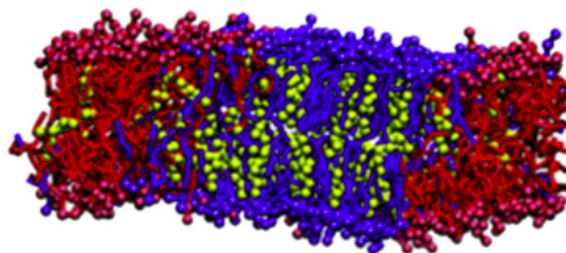


Figure 1.4: Example of phase-separated lipid bilayer [40]. Saturated lipids in blue, unsaturated lipids in red and cholesterol yellow.

The key ingredient for more realistic membrane models is definitely lipid heterogeneity. Real membranes present transient and dynamic nanocompartments called lipid rafts [41]. Lipid rafts are liquid-ordered (Lo) phase nanodomains, typically enriched in cholesterol and saturated lipid species like sphingolipids and gangliosides, embedded in a liquid-disordered phase (Ld) mainly consisting of unsaturated lipids (Figure 1.4). In other words, rafts are phase-separated functional platforms on the cell membrane. There are pieces of evidence that they are involved in signaling and trafficking cellular processes [42]. Both the small size and short lifetime of these domains make direct measurements tremendously difficult. Thanks to CG MD simulations, it was possible to investigate spatio-temporal scales involved while preserving molecular details [43]. For a deeper understanding of AuNPs bioactive role, it is meaningful to investigate their interaction with lipid rafts. Simulation of functionalized AuNPs, which are able to passively penetrate through the membrane, interacting with lipid rafts, can reveal insights that are not accessible with experiments, and this field is still at its beginning.

Chapter 4 presents my actual results on this subject, which concerns the interaction between MUS:OT NPs and model neuronal membranes undergoing phase separation. Atomic force microscopy, quartz crystal microbalance, and molecular dynamics are combined to study the interaction between model neuronal membranes, which spontaneously form lateral phase-separation, and MUS:OT AuNPs. The MUS:OT AuNPs and the membrane molecular models have a coarse-grain resolution compatible with the Martini FF. We found that the NPs interact with the bilayer and form bilayer-embedded ordered aggregates. Another interesting result is that nanoparticles suppress lipid domains, depending on NPs concentration. Finally, through a simple thermodynamic model, we show that the principal driving force for the suppression of phase-separation is the lipid–lipid enthalpy variation.

Chapter 2

Computational methods

2.1 Molecular Dynamics: an overview

To have access to the time and space scales involved in the systems under study, we used Molecular Dynamics (MD) simulations. The matter behaviour at a microscopic level can be understood through classical mechanics; relativistic effects in these systems are not appreciable and quantistic effects can be ignored or, alternatively, included in the equations as semi-classical correction. We made an acceptable error in using classical mechanics equations, if the characteristic energy of the system is big enough to ignore vibrational levels discrete nature. In particular, in bio-molecular systems this condition is always true as the atoms are sufficiently massive. Generally, the matter dynamics can be described as the dynamics of its fundamental components, thus when we study a system via MD we are basically treating an n-bodies problem by numerically solving Newton equations of motion. Clearly, computing power is necessary to describe systems of up to hundreds of thousands particles. The MD basic idea is to reproduce the time evolution of studied objects and their interactions, as well as the underlying mechanisms. From a more technical point of view, MD is the combination of the equations of motion integrator and all the algorithms needed to elaborate the system evolution in specific thermodynamic condition. Statistical mechanics, through partition function, constitutes the theoretical basis for studying thermodynamic quantities as temperature, pressure and so on.

In this chapter, the theoretical basis of MD and some of the more advanced sampling techniques used in our research will be described - i.e. umbrella sampling and metadynamics.

2.1.1 Fundamentals of classical mechanics

In the classical MD description, atoms are considered point-like objects and they interact with each other via an effective potential. The potential energy function (PEF) is defined as $U(\vec{r}_1, \dots, \vec{r}_N)$. From Newton equation, we can say that the force experienced by each atom is

$$\vec{F}_i = -\frac{\partial U}{\partial \vec{r}_i} \quad (2.1)$$

where \vec{F}_i is given by:

$$\vec{F}_i = m_i \ddot{\vec{r}}_i \quad (2.2)$$

Eq. 2.1 is a system of $3N$ second-order differential equations or, equivalently, $6N$ first order differential equations:

$$\begin{cases} m_i \frac{d\vec{v}_i}{dt} = \vec{F}_i \\ \frac{d\vec{r}_i}{dt} = \vec{v}_i \end{cases} \quad (2.3)$$

This is a Cauchy problem - when the system is combined with $2N$ initial conditions $\vec{r}_i(0) = \vec{r}_{0i}$ and $\vec{v}_i(0) = \vec{v}_{0i}$ - with the trajectory $(\vec{r}_i(t), \vec{v}_i(t))$. If we write Eq. 2.2 in terms of the momenta $\vec{p}_i = m_i \vec{v}_i$ we obtain:

$$\frac{d\vec{p}_i}{dt} = \vec{F}_i, \quad (2.4)$$

thus we can define the space phase vector at some instant t as the following set of $6N$ numbers $\vec{x} = (\vec{r}_1, \dots, \vec{r}_N, \vec{p}_1, \dots, \vec{p}_N)$ and solving Eq. 2.4 generates the set of functions

$$(\vec{r}_1(t), \dots, \vec{r}_N(t), \vec{p}_1(t), \dots, \vec{p}_N(t)), \quad (2.5)$$

which describes a trajectory in the phase space.

2.1.2 Fundamentals of statistical mechanics

Classical mechanics provides all the information about the microscopic view of the system, now we are interested to its thermodynamic properties i.e. internal energy, temperature, pressure and so on. The problem of extrapolating macroscopic observable from the complete sets of positions and velocities of all particles has been solved as part of statistical mechanics by Boltzmann and Gibbs. Statistical mechanics constitutes a direct connection between the microscopic and the macroscopic worlds, giving a rigorous derivation of a system thermodynamic properties given only the microscopic information. The basic idea of the statistical mechanics is that *a macroscopic observable of a system does not strongly depend on the complete dynamics of each particle in the system, but rather on an average that cancels out all the details of the microscopic features*. Thus, when iterating an experiment, we will find that for a specific macroscopic state of a system there exists different microscopic states that yield to the same macroscopic properties. This basically lead us to the definition of a statistical ensemble: *a collection of systems subject to a set of common interactions and sharing the same macroscopic properties*. Each N -particle system can be described by its microstate \vec{x} - a point in the phase space - and the total ensemble is a set of points in the phase space with the constraint to be part of the ensemble. The time evolution of the ensemble is described by the evolution of these phase space points according to the classical mechanics. The macroscopic observables can be obtained

simply averaging over all the system of the ensemble thus, at every time, we need to know the *ensemble distribution function* $\tilde{\rho} = \tilde{\rho}(\vec{x}, t)$ which tell us the microstates of the phase space that are part of the ensemble.

Let's consider the infinitesimal phase space volume $dx = dr_1 \dots dr_{3N} dp_1 \dots dp_{3N}$, then

$$\frac{1}{N} \tilde{\rho}(\vec{x}, t) dx = \rho(\vec{x}, t) dx \quad (2.6)$$

where $\rho(\vec{x}, t)$ is the probability to find a system of the ensemble in the microstate \vec{x} at a instant t and N is the total number of microstates in the ensemble.

Clearly, from the probability theory, must be

$$\int_{\Omega} \rho(\vec{x}, t) dx = 1 \quad (2.7)$$

with $\rho(\vec{x}, t) \geq 0$.

Given the most convenient ensemble distribution function, the ensemble average of an observable $A = A(\vec{x})$, a time t , is

$$\langle A \rangle(t) = \int_{\Omega} A(\vec{x}) \rho(\vec{x}, t) dx. \quad (2.8)$$

If the ensemble is in thermodynamic equilibrium, the equilibrium observable A must be time-independent and can be expressed as

$$\langle A \rangle = \frac{1}{\mathcal{Z}} \int_{\Omega} A(\vec{x}) f(\mathcal{H}(\vec{x})) dx \quad (2.9)$$

where $f(\mathcal{H}(\vec{x}))$ is a scalar function of the Hamiltonian \mathcal{H} and \mathcal{Z} is the partition function defined below

$$\mathcal{Z} = \int_{\Omega} f(\mathcal{H}(\vec{x})) dx. \quad (2.10)$$

The thermodynamic constant in time observables for an ensemble in thermodynamic equilibrium are called *control variables* and they must be specified in order to calculate the partition function \mathcal{Z} . An ensemble is characterized by its control variables, below some of the main ensembles used in statistical mechanics with their control variables are listed

- microcanonical ensemble: constant NVE
- canonical ensemble: constant NVT
- isothermal-isobaric ensemble: constant NpT
- grand-canonical ensemble: constant μpT .

In order to complete this brief overview of statistical mechanics, we need a link between statistical averages and data from experiments. When preparing an experiment, we have only *one system* in a specific macroscopic state. The measurement of

a macroscopic observable A requires long times respect to microscopic time scales. A is a function of time and phase space vector and it fluctuates over time due to particles interaction, thus when we measure it we take an average over time.

In principle, we could obtain the observable A "real" mean value by averaging on infinite time:

$$\bar{A} = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_{\tau_0}^{\tau} A(\vec{x}_t) dt \quad (2.11)$$

The connection between time averages and ensemble is provided by the ergodic hypothesis: a system is ergodic if the microstates of the phase space with the same energy are accessible with the same probability over a long time. If a system is ergodic, thus the ergodic theorem says that time averages and ensemble averages are almost equal everywhere in the phase space:

$$\bar{A} = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_{\tau_0}^{\tau} A(\vec{x}(t)) dt = \int_{\Omega} A(\vec{x}) \rho(x, t) dx = \langle A \rangle(t). \quad (2.12)$$

2.2 Molecular Dynamics algorithms

We can say that MD simulations have changed the way to use computers in science. Computers are no longer just a means to make calculations which are too onerous for men, but it becomes a full-fledged virtual laboratory where to conduct experiments. Classical Molecular Dynamics (MD) is a set of computational techniques for numerical resolution of Newton's second law and, consequently, for the *trajectories* (time sequential configurations) of the elementary components of the system under study. The advantage in using MD, compared to a real experiment, is that the system initial conditions can be easily set and the key parameters are relatively few being the system simplified. Furthermore, the experiment reproducibility is easily obtained as well as it is free from interferences of real laboratories. Obviously, the disadvantages are related to the accessible time scales and numerical accuracy which are strictly dependent on the available computational power. In addition, it is crucial to verify the reliability of the results obtained via MD simulations by testing the quality of the models used to reconstruct the potential energy function (PEF).

As we mentioned before, once given the initial configuration (potential, positions, velocities), the problem is to numerically solve the equations of motion. So, in the next paragraphs we will describe how this can be done. The first step consists in dividing the integration time in little steps δt - the MD time steps - in which forces are considered constant. Forces and so the accelerations are computed starting from an initial phase space vector $\vec{x}(0)$ through the PEF. Using the integration algorithm - which combines the accelerations with initial positions and velocities - we obtain a new phase space vector $\vec{x}(t)$ and thus a new set of forces. This procedure is repeated every time step until the end of the simulation time.

Generally, we are interested in time averages and these can be obtained from 2.12,

rewritten by discretizing time:

$$\bar{A} = \frac{1}{N} \sum_{n=1}^N a(\vec{x}(n\Delta t)) \quad (2.13)$$

where $\Delta t = i\delta t$ $i = 1, 2, \dots$ and $N\Delta t$ is the total time of average.

In this work, the versatile package GROMACS has been used to perform molecular dynamics simulation. GROMACS was first designed for biochemical systems like lipids, nucleic acids and proteins. These systems have hundreds to millions of particles and present many of bonded as well as non bonded interaction to calculate. In the last decades, GROMACS has been chosen for research in various systems like standard and supramolecular polymers, molecular machines and so on ([44], [45]).

2.2.1 Configuration setup

As mentioned before, the system setup consists in the assignment of $6N$ initial conditions for the positions and the velocities - that is the initial $\vec{x}(0)$ phase space vector of the particles. Generally, we extract randomly the initial velocities from the Maxwell-Boltzmann distribution function at the equilibrium temperature:

$$\rho(v_x, v_y, v_z)dv_x dv_y dv_z = \left(\frac{m}{2\pi k_B T}\right)^{\frac{3}{2}} e^{-\left(\frac{m \vec{v}^2}{2k_B T}\right)} dv_x dv_y dv_z. \quad (2.14)$$

Since the set of the extracted velocities \vec{v}_i^* are not completely symmetric, the initial momentum of the centre of mass (COM) of the system $\vec{P}_0 = \sum_i m_i \vec{v}_i^*$ will be not zero. In order to avoid a constant drift of the system and to preserve the correct kinetic energy, this can be removed by rescaling velocities \vec{v}_i^* to have $\vec{P}_0 = 0$. COM motion removal is necessary because, in general, the total force acting on the system $\vec{F} = \sum_i^N \vec{F}_i$ is zero; of course this action reduces the system degrees of freedom (DOF) by three.

2.2.2 Periodic boundary conditions

In an MD simulation, we want to introduce a well defined reference system of coordinates. Thus, every sample system evolves in a *simulation box* whose shape can be chosen according to the specific symmetry of the simulated system. Using the simulation box requires a correct treatment of the *boundary conditions*. In fact, we want to avoid surface effects. Furthermore, by imposing periodic boundary condition (PBC) we have the possibility to study bulk features of a system without simulating a too large number of particles.

In Figure 2.1 a two-dimensional box with PBC is represented: only the central red contoured box is the simulation box. The PBC idea is to replicate that box in every direction so that there are no surface effects nor walls in the simulation box.

When a particle moves in the central box, all its images virtually move the same way in the copies of the box; if a particle leaves the virtual boundary of the central

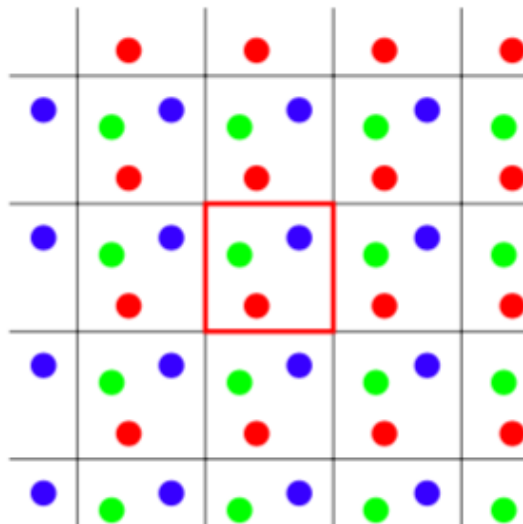


Figure 2.1: Cartoon showing a two-dimensional box with PBC. The red contoured box in the centre is the simulation box and it is replicated along each dimension.

box from one side, it will appear from the other side and the number density of particles in the simulation box remains constant. In a cubic box, for example, if a particle crosses its boundary in the positive/negative x direction, then its coordinate is corrected by subtracting/adding the box side length parallel to x direction. It must be taken into account that imposing PBC to a system may affect its properties. Obviously, we are not able to explore vibrating models with a wavelength greater than the box length. Another issue is the interaction range between particles: we want to be sure that the smallest simulation box is greater than interaction range. This condition is achieved quite easily with short range interactions i.e. the Lennard-Jones potential, but can be more problematic with long range interactions as electrostatic ones.

2.2.3 Integrators

Numerical integrators are algorithms that solves Newton's equation of motion. Liouville theorem says that a system in which energy is conserved shows time reversal invariance and conservation of the volume in the phase space. Thus, an algorithm for MD should have these features. Problems may come from discretization of time, which could introduce fluctuations or drifts in the total energy. In the following paragraphs, the main algorithms to solve equation of motion are shown with a focus on the default GROMACS integrator [44].

2.2.4 Euler algorithm

The simplest way to integrate Newton's equations of motion is the Euler algorithm that approximates the derivatives with finite differences:

$$\frac{d\vec{v}_i}{dt} \approx \frac{\Delta \vec{v}_i}{\Delta t} = \frac{1}{m_i} \vec{F}_i \quad (2.15)$$

$$\frac{d\vec{r}_i}{dt} \approx \frac{\Delta \vec{r}_i}{\Delta t} = \vec{v}_i. \quad (2.16)$$

The first-order Taylor expansion of positions and velocities gives

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i}{dt} \Delta t + o(\Delta t^2) \quad (2.17)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i + \frac{d\vec{v}_i}{dt} \Delta t + o(\Delta t^2). \quad (2.18)$$

Now, we can substitute the finite derivatives expressions obtaining:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t) \Delta t \quad (2.19)$$

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \frac{\vec{F}_i}{m_i} \Delta t. \quad (2.20)$$

In Euler algorithm, positions and velocities are coupled to each other but $t + \Delta t$ and $t - \Delta t$ systems are independent.

Euler algorithm has been abandoned because it does not conserve the volume in the phase space, in other words, it is not symplectic.

2.2.5 Verlet algorithm

The Verlet algorithm is a symplectic integrator and it conserves energy.

It requires the positions and the forces at a time t and the positions a time $t - \Delta t$ to calculate positions at a time $t + \Delta t$:

$$\vec{r}_i(t - \Delta t) = \vec{r}_i(t) - \frac{d\vec{r}_i}{dt} \Delta t + \frac{1}{2} \frac{d^2 \vec{r}_i}{dt^2} \Delta t^2 + o(\Delta t^3) \quad (2.21)$$

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \frac{d\vec{r}_i}{dt} \Delta t + \frac{1}{2} \frac{d^2 \vec{r}_i}{dt^2} \Delta t^2 + o(\Delta t^3) \quad (2.22)$$

By adding the two equations together we get:

$$\vec{r}_i(t + \Delta t) + \vec{r}_i(t - \Delta t) = 2\vec{r}_i(t) + \frac{d^2 \vec{r}_i}{dt^2} \Delta t^2, \quad (2.23)$$

in which the second derivative can be substituted by Newton equation 2.2 obtaining:

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{F}(\vec{r}_i(t))}{m_i} \Delta t^2 \quad (2.24)$$

Equation Eq. 2.24 contains third order time steps terms with $o(\Delta t^4)$ error thus, we expect Verlet algorithm to be more efficient than Euler one.

The velocities do not appear in the equation above but they can be obtained by taking the difference between equations Eq. 2.21 and Eq. 2.22:

$$\vec{v}_i(t) \simeq \frac{\vec{r}_i(t + \Delta t) - \vec{r}_i(t - \Delta t)}{2\Delta t} \quad (2.25)$$

Velocities at time t are an output of the calculation and not part of the integrator itself. Another disadvantage is that Verlet is not a self-starting integrator because the algorithm requires positions at time $t - \Delta t$, so at $t=0$ we need some tricks to calculate the past positions as truncating at the first order equation Eq. 2.22: $\vec{r}_i(-\Delta t) \simeq \vec{r}_i(0) - \vec{v}_i(0)\Delta t$.

2.2.6 Leapfrog algorithm

The Leapfrog algorithm is a variant of the Verlet one and it is the default MD integrator used in this work. It computes the positions at time t and the velocities at time $t + \frac{\Delta t}{2}$ from the forces at time t and the velocities at time $t - \frac{\Delta t}{2}$. In other words, velocities are not calculated at every Δt , but they are updated at intermediate times:

$$\begin{cases} \vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t - \frac{\Delta t}{2})\Delta t \\ \vec{v}_i(t + \frac{\Delta t}{2}) = \vec{v}_i(t - \frac{\Delta t}{2}) + \frac{\vec{F}(\vec{r}_i(t))}{m_i} \Delta t \end{cases} \quad (2.26)$$

Even if it seems Leapfrog algorithm has a first order approximation we can easily demonstrate its equivalence with the Verlet one. We can substitute velocities from the second equation of the system Eq. 2.26 into the first one obtaining:

$$\vec{r}_i(t + \Delta t) = \vec{r}_i(t) + \vec{v}_i(t + \frac{\Delta t}{2})\Delta t - \frac{\vec{F}(\vec{r}_i(t))}{m_i} \Delta t^2, \quad (2.27)$$

which is a second order equation in Δt . Now, by rewriting Eq. 2.26 for r_i and subtracting the two equations, we get:

$$\vec{r}_i(t + \Delta t) - \vec{r}_i(t) = \vec{r}_i(t) - \vec{r}_i(t + \Delta t) + [\vec{v}_i(t + \frac{\Delta t}{2}) + \vec{v}_i(t - \frac{\Delta t}{2})]\Delta t. \quad (2.28)$$

Replacing the term in the square brackets according to Eq. 2.26 we reach the

following expression

$$\vec{r}_i(t + \Delta t) = 2\vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \frac{\vec{F}(\vec{r}_i(t))}{m_i} \Delta t^2 \quad (2.29)$$

which is exactly Eq. 2.24. Instead, velocities at time t - which are necessary for energy computation - can be obtained by $\vec{v}_i(t + \frac{\Delta t}{2})$ and $\vec{v}_i(t - \frac{\Delta t}{2})$ as follow:

$$\vec{v}_i(t + \Delta t) = \vec{v}_i(t) + \left(\frac{\vec{F}(\vec{r}_i(t)) + \vec{F}(\vec{r}_i(t + \Delta t))}{2m_i} \right) \Delta t. \quad (2.30)$$

The main advantage with respect to the Verlet algorithm is that Leapfrog integrator is self-starting, so it does not require the positions at time $t - \Delta t$. Furthermore, velocities are part of the algorithm itself. The disadvantage, here, is that positions and velocities at time t are not synchronized; velocities at time t are necessary for the calculation of the kinetic energy contribution to the total energy and can be computed as follow:

$$\vec{v}_i(t) \simeq \frac{\vec{v}_i(t + \frac{\Delta t}{2}) + \vec{v}_i(t - \frac{\Delta t}{2})}{2}. \quad (2.31)$$

2.2.7 Neighbor list

In order to solve the classical equations of motion, the forces calculation requires most of the computational time; at each step all mutual distances between atoms are calculated in a double loop. The computational cost scales as $\frac{N(N-1)}{2} \approx N^2$ each step, thus for each n -body interaction forces calculation scales with $\approx N^n$. A useful approximation to reduce this computational cost consists in a neighbour list. It is possible to consider, for each particle, a list of its neighbour particles that lie within a sphere of radius r_c . Thus, the interactions are computed between the selected particle and the neighbours in its pair-list. This list can be updated every N steps, where N has been chosen in such a way that particle neighbours do not change in that time interval. An over or under-estimation of the inter-particle energy contribution could be caused by particles crossing the pair-list during the non-updating time.

As suggested by Verlet, we can consider a buffered pair-list or a cut-off Verlet scheme in which the pair list is built considering those particles that are close to the selected one by a distance of $r_l > r_c$, called radius list. A schematic representation of the buffered pair-list is shown in Figure 2.2; the pair-list is updated every N time step, as before, but every time step the pairwise contribution is computed only between those particles that lie within a sphere of radius r_c . The parameters of these approximations need to be chosen wisely, in order to speed up the calculation, but not to affect the conservation of total energy.

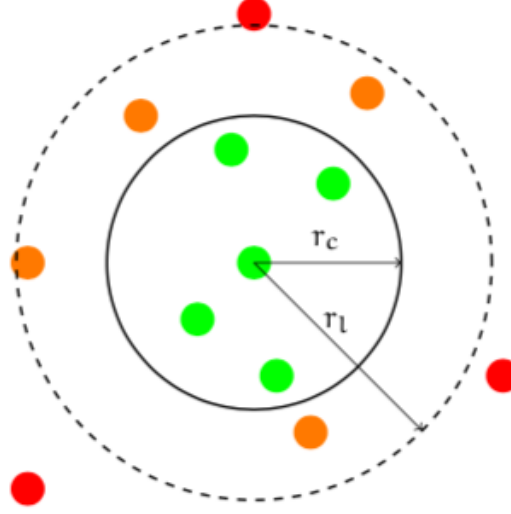


Figure 2.2: Cartoon of the buffered pair-list respect to the central particle. All green particles are included in the interaction calculation, being below the cut-off radius r_c . All orange particles are in the pair-list: at every step it is checked if their distances became smaller than r_c . Red particles outside the pair-list are not in the calculation until the next list update.

2.2.8 Thermostat algorithms

When simulating NVT or NpT ensemble, we need to keep constant temperature by using thermostat algorithms. Some of these algorithms introduce more degrees of freedom (DOF) in the system to mimic a real temperature bath coupling (Nosé-Hoover), others act directly on particle velocities (Anderson, Berendsen, Bussi) like the one used in both the works presented herein. Bussi et al. ([46]) suggest a velocity rescale algorithm in which the velocities of all particles are scaled by the same factor. If we considered the average kinetic energy

$$\langle K \rangle = \frac{3}{2} N k_B T, \quad (2.32)$$

by replacing $3N$ with the total DOF of the system N_f , we obtained the following expression

$$\langle K \rangle = \frac{1}{2} N_f k_B T, \quad (2.33)$$

with T being the target temperature. The scaling factor is then defined as

$$\alpha_T = \sqrt{\frac{\langle K \rangle}{K}} \quad (2.34)$$

where K is the total kinetic energy of the system.

This very simple algorithm is often called weak coupling thermostat and the scaling

operation is usually performed at a fixed rate during the simulation or when the kinetic energy exceeds a fixed interval centered on a target value. This algorithm cannot be safely used when the observables are dependent on the fluctuations rather than on the averages or with small systems. Furthermore, the sample ensemble is not explicit so, in order to have the correct canonical sampling, Bussi et al. ([46]) propose a new scaling factor that is:

$$\alpha_T = \sqrt{\frac{\langle K_T \rangle}{K}} \quad (2.35)$$

where K_T is obtained by a stochastic procedure from the equilibrium canonical distribution of the kinetic energy:

$$P(K_T) dK_T \propto K_T^{\frac{N_f}{2}-1} e^{-\beta K_T} dK_T \sqrt{\frac{\langle K_T \rangle}{K}} \quad (2.36)$$

with $P(K_T)dK_T$ being the probability that the system has a kinetic energy between K_T and $K_T + dK_T$.

Since there is a discontinuity in particle velocities around the scaling step, the K_T value is chosen based on the previous K value. Thus, the total kinetic energy is calculated by using the following auxiliary continuous stochastic dynamics:

$$dK = \frac{\langle K \rangle - K}{\tau_T} dt + 2\sqrt{\frac{K \langle K \rangle}{N_f \tau_T}} dW \quad (2.37)$$

where dW is a Wiener stochastic noise and τ_T is an arbitrary parameter related to the thermostat time response. The expression Eq. 2.37 reduces to the simple weak coupling rescale when $\tau_T \rightarrow 0$. The first term drives the system to equilibrium with a characteristic time τ_T ; after the system reaches the equilibrium, the stochastic contribution samples the canonical distribution.

2.2.9 Barostat algorithm

In order to maintain the system at a constant pressure, a barostat algorithm is used; the idea is to adjust the pressure by changing the simulation box volume. As proposed by Berendsen, when scaling both volume and particle coordinates, the pressure changes with the following rate:

$$\frac{dp}{p} = \frac{p_0 - p}{\tau_p} \quad (2.38)$$

where p_0 is the target pressure, τ_p is a coupling constant related to the time response of the barostat and p is the system pressure at time t . If δt is the simulation time step, the volume scaling factor λ is given by

$$\lambda = 1 - K_T(p_0 - p) \frac{\delta t}{\tau_p} \quad (2.39)$$

being K_T the isothermal compressibility defined as:

$$K_T = -\frac{1}{V} \left(\frac{\partial V}{\partial p} \right)_T \quad (2.40)$$

The factor $\frac{\delta t}{\tau_p}$ takes into account the finite barostat time response. When $\tau_p \rightarrow 0$ the system has an infinity isothermal compressibility which means that a really small change in volume will lead to the correct pressure; the particle coordinates are scaled by the factor $\lambda^{\frac{1}{3}}$. In case of anisotropic system, the Berendsen algorithm can be generalized such that λ becomes a 3×3 matrix, by considering the pressure matrix \mathbf{P} and the volume matrix \mathbf{H} . As for the weak coupling thermostat, the main problem of Berendsen thermostat is that the sampled ensemble is not known. The correct approach to the isobaric ensemble sampling was derived by Parrinello and Rahman and it is described in the next section [47].

2.2.10 Parrinello-Rahman barostat algorithm

In the Parrinello-Rahman barostat ([47], [48]) the system is coupled to an external piston of "mass" M_h and the volume is treated like a lagrangian variable of the system. Then, the volume time evolution can be computed by solving the Euler-Lagrange equations. In this method, both size and shape of the simulation box are allowed to fluctuate and they are described by the volume matrix \mathbf{H} such that $V = |\det \mathbf{H}|$. If an external pressure p_0 is applied to the piston with lagrangian coordinate \mathbf{H} , its potential energy is given by

$$U_p = p_0 |\det \mathbf{H}|. \quad (2.41)$$

The \mathbf{H} variation in time means that a "kinetic energy" is associated to the piston as described in the following expression:

$$k_p = \frac{1}{2} M_h \text{Tr}(\dot{\mathbf{H}}^t \dot{\mathbf{H}}) \quad (2.42)$$

being $()^t$ the transpose operator. By defining the vector \vec{s}_i so that $\vec{r}_i = \mathbf{H} \vec{s}_i$, it is possible to write the lagrangian of the particle coordinates in terms of \mathbf{H} .

Consequently, the square displacement is given by

$$\vec{r}_i \cdot \vec{r}_i = (\mathbf{H} \vec{s}_i)^t \mathbf{H} \vec{s}_i = \vec{s}_i^t \mathbf{H}^t \mathbf{H} \vec{s}_i. \quad (2.43)$$

Thus, the Lagrangian found by Parrinello et al. (CITA) is

$$L = \frac{1}{2} \sum_{i=1}^N m_i \dot{\vec{s}}_i^t \mathbf{G}^t \dot{\vec{s}}_i - \sum_{i=1}^N \sum_{j=i+1}^N U_{ij}(\vec{r}) + \frac{1}{2} M_h \text{Tr}(\dot{\mathbf{H}}^t \dot{\mathbf{H}}) - p_0 |\det \mathbf{H}| \quad (2.44)$$

where $\mathbf{G} = \mathbf{H}^t \mathbf{H}$.

The equation of motion for \mathbf{H} and \vec{s}_i (Eq. 2.45) can be computed by solving the Euler-Lagrange equations; the equations of motion, as derived by Parrinello et al., sample an isobaric ensemble correctly. This can be easily shown by considering the following Hamiltonian:

$$H = \frac{1}{2} \sum_{i=1}^N m_i \dot{\vec{s}}_i^t \mathbf{G}^t \dot{\vec{s}}_i - \sum_{i=1}^N \sum_{j=i+1}^N U_{ij}(\vec{r}) + \frac{1}{2} M_h \text{Tr}(\dot{\mathbf{H}}^t \mathbf{H}) + p_0 |\det \mathbf{H}| \quad (2.45)$$

when the system is at an equilibrium temperature T , the equipartition theorem says that the kinetic term of the system particles contributes to the energy by $\frac{3}{2} N K_B T$, while the kinetic term due to the piston by $\frac{9}{2} N K_B T$. Since $3N \gg 9$, the Hamiltonian can be approximated as

$$H \simeq \langle K \rangle + U + p_0 V = H \quad (2.46)$$

that is the enthalpy. In this way, the correct isobaric ensemble is sampled, being the Hamiltonian a constant of motion. When the system is also coupled to a thermostat, the constant of motion becomes the Gibbs free energy $G = H - TS$, since the $-TS$ term has to be added. Isobaric-isotherm ensemble is correctly sampled. When there is an imbalance between the internal pressure p at a time t and the target pressure p_0 , the Parrinello-Rahman barostat acts in a way that the system recovers this imbalance in a characteristic time related to the parameter M_h . If we are interested only in statistical averages, this parameter can be arbitrarily chosen, in fact the system properties at equilibrium are independent from the masses of its constituents parts. If this is not the case, M_h has to be properly chosen; in this regard the authors suggest to choose M_h of the order of $\frac{L}{c}$ where L is the simulation box size and c is the sound velocity. Parrinello-Rahman method is less stable than Berendsen one, but it guarantees the correct sample of the isobaric ensemble. In order to avoid large volume fluctuations When the system is not well equilibrated, the Berendsen barostat is commonly used to reach equilibrium before switching to the Parrinello-Rahman algorithm.

2.3 MD enhanced sampling

In studying biomolecular systems, the free energy is a very fundamental quantity; free energy variations are useful to understand if a chemical reaction occurs spontaneously, whether some molecules are able to self-assemble in water solution, if a protein configurational change can take place and so on. The Gibbs free energy is specific for the NpT isobaric-isothermal ensemble whether the Helmholtz free energy A for the NVT canonical ensemble. The time evolution of a biomolecular system and its equilibrium properties are then determined by the system's free energy surfaces

(FES). The free energy is related to the partition function of an ensemble thus we can obtain other thermodynamic quantities by differentiation.

2.3.1 Collective variables

The free energy in function of the distance between the centre of mass (COM) of two molecules provides information about their attraction or if they form a bound state. Similarly, when considering the FES as a function of a set of some generalized coordinates, we have access to some mechanical, chemical and thermodynamic processes occurring in the system. These generalized coordinates are called *collective variables* (CVs) of the system; the FES in the CVs space provide a map of the barrier heights that have to be overcome for a transition to take place. The free energy landscape can be rich in minima, even for small molecules, and thus difficult to sample in MD simulations. The advantage of using CVs is the possibility of sampling these regions that are crucial to understand the process under study.

2.3.2 Free energy surface

Let us suppose we are interested in the FES in function of the CV $s(\vec{r})$ in a NpT ensemble for an isotropic system. The following discussion remains valid even for a NVE ensemble with the Helmholtz free energy instead of the Gibbs one and with the correct partition function Z_{NVE} for the ensemble. The Gibbs free energy along the $s(\vec{r})$ CV is given by

$$G(s) = -K_B T \ln Q(s) \quad (2.47)$$

where $Q(s)$ is the partition function that integrates out all the DOF of the system except for $s(\vec{r})$:

$$Q(s) = \frac{1}{Z_{NpT}} \int_0^{+\infty} dV \int_{\Omega} e^{-\beta(H(\vec{x})+pV)} \delta(s(\vec{r}) - s) d\mathbf{x}. \quad (2.48)$$

Since $s(\vec{r})$ does not depend on particle momenta, $Q(s)$ can be rewritten as:

$$Q(s) = \frac{\int_{\Omega} e^{-\beta U(\vec{r})} \delta(s(\vec{r}) - s) d\mathbf{r}}{\int_{\Omega} e^{-\beta U(\vec{r})} d\mathbf{r}} \quad (2.49)$$

where $U(\vec{r})$ is the potential energy function (PEF). We can interpret $Q(s)ds$ as the probability to find the system with $s(\vec{r})$ between s and $s + ds$. The equation contains a direct phase space integration so it can be written using the ergodic theorem (Eq. 2.12), as a time average:

$$Q(s) = \langle \delta(s(\vec{r}) - s) \rangle = \lim_{\tau \rightarrow +\infty} \frac{1}{\tau} \int_0^{\tau} \delta(s(\vec{r}(t)) - s) dt. \quad (2.50)$$

In principle, it is possible to derive the time sampling of Eq. 2.50 by MD simulations. The problem here is that we are not able to simulate infinite time and thus to sample

all the phase space and consequently the ensemble average. Generally, we can obtain a good sample of the phase space regions around a local minimum of the FES. Instead, high energy regions - due to their unfavorable Boltzmann factor, provide only a small contribution to the partition function. Thus, these regions have to be overcome in order to reach other minima, but energy barriers that are higher than the thermal fluctuations ($\sim -K_B T$) have a small probability of being overcome. These transitions are called *rare events*.

2.4 Bias based advanced sampling

Several techniques have been developed to sample rare events. In particular, we want to explore other regions of the phase space by escaping local energy minima and to compute the FES along one or a small set of CVs. The idea is to drive the transition between two metastable configurations by adding a *bias potential*.

In the following paragraph are described the methods used in this work: *umbrella sampling* and *metadynamics* (for a more complete discussion consult the review by Kästen ([49]) for the former and the review by Laio and Gervasio ([50]) for the latter, Tuckerman's book describe further advanced sampling methods ([51])).

2.4.1 Umbrella sampling

Umbrella sampling is one of the most used method for calculating free energy differences between two states. Developed by Torie and Valleau, it proposes to drive the system from a known state A to an other known state B through a path defined by the chosen CV - $s(\vec{r})$ - describing the process. Umbrella sampling computational performances degrades rapidly for more than one CV. The path between A and B is divided into a discrete number of windows N_w , then take a subset s_i , $i = 1, \dots, N_w$ of the continuous CV values along the path.

In Figure 2.3, an example of windows selection is shown.

In order to restrain the system to the s_i target value, we add - for each value s_i - a bias potential w_i to the total potential energy function (PEF). MD simulations are performed for each window and then all the data are combined to obtain the biased FES along the chosen CV and, eventually they are used to also recover the unbiased FES.

The biased PEF of window i is given by

$$U_i^b = U(\vec{r}) + w_i(s) \quad (2.51)$$

where $w_i(s)$ is the bias potential on window i in function of the CV $s(\vec{r})$ and the superscript b indicates the biased quantities. The corresponding biased partition function, integrated over all DOF but s , can be obtained from Eq. 2.49 by the

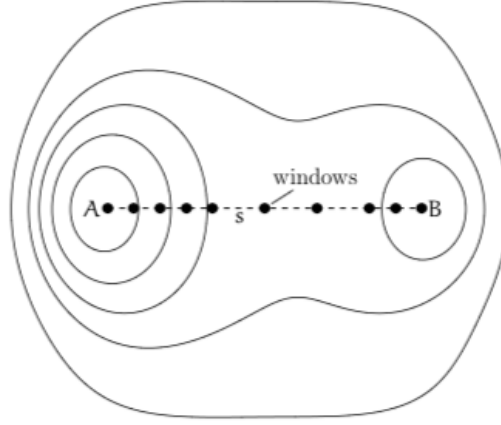


Figure 2.3: Cartoon showing an umbrella sampling windows selection. The contour line represents a 2-minima free energy profile through which the system evolves. The dashed line is a path along the CV $s(\vec{r})$ to connect the state A to the state B. The black continuous line corresponds to the selected windows for different values of s .

substitution $U(\vec{r}) \rightarrow U_i^b(\vec{r})$:

$$Q_i^b(s) = \frac{\int_{\Omega} e^{-\beta U(\vec{r})} \delta(s(\vec{r}) - s) d\vec{r}}{\int_{\Omega} e^{-\beta (U(\vec{r}) + w_i(s)(\vec{r}))} d\vec{r}}. \quad (2.52)$$

The unbiased partition function can be recovered with the following expression:

$$Q_i(s) = \frac{Q_i^b(s) e^{\beta w_i(s)} \int_{\Omega} e^{-\beta U(\vec{r})} e^{-\beta w_i(s)(\vec{r})} d\vec{r}}{\int_{\Omega} e^{-\beta U(\vec{r})} d\vec{r}} = \quad (2.53)$$

$$= Q_i^b(s) e^{\beta w_i(s)} \langle e^{-\beta w_i(s)} \rangle. \quad (2.54)$$

Then, we obtain the FES using Eq. 2.47

$$G_i(s) = -k_B T \ln Q_i^b(s) - w_i(s) + C_i, \quad (2.55)$$

where $Q_i^b(s)$ is given by the ensemble average from biased MD simulation, $w_i(s)$ is a known function and C_i is an additive constant independent from s that connects the free energy curves $G_i(s)$ of different windows. In order to compute the global FES, we need to combine together more windows and all $\{C_i\}$ must be computed.

2.4.2 Weighted histogram analysis method

In order to extract the global FES $G(s)$ we need to compute the global unbiased partition function $Q(s)$ from the set of biased partition function $\{Q_i^n(s)\}$. Several analysis methods have been developed to combine information collected from the N_w windows; here we described the most commonly used Weighted Histograms Analysis

Method (WHAM) [52], [53]. We represent the set $\{Q_i^b(s)\}$ with a set of histograms $\{h_i(s)\}$ with n_i bins for each biased window N_w . The aim is to minimize the statistical error in the calculation of $Q(s)$; the global distribution function is thus computed as follows

$$Q(s) = \sum_{i=0}^{N_w} p_i h_i(s) \quad (2.56)$$

where

$$\sum_{i=0}^{N_w} p_i = 1 \quad (2.57)$$

being p_i the weights chosen to minimize the statistical error on $Q(s)$:

$$p_i = \frac{n_i e^{-\beta(w_i(s) - C_i)}}{\sum_{j=0}^{N_w} n_j e^{-\beta(w_j(s) - C_j)}}. \quad (2.58)$$

where n_i is the total number of independent bins in the i -th histogram. Since $Q(s)$ depends on $\{C_i\}$ and vice-versa, the set of constants $\{C_i\}$ can be computed in an iterative self-consistent way. A first guess of set $\{C_i\}$ is used to compute the weights from Eq. 2.58, then $Q(s)$ is computed from Eq. 2.56 and it is used to obtain a new set of $\{C_i\}$ from the $Q(s)$ global distribution function:

$$e^{-\beta C_i} = \left\langle e^{-\beta w_i(s)} \right\rangle = \int Q(s) e^{-\beta w_i(s)} ds \quad (2.59)$$

The process is repeated until both equations are satisfied. When the iterative process is completed, we can obtain the global FES $G(s)$ from Eq. 2.47. In order to avoid convergence problems of the iterative process, the histograms in adjacent windows must be sufficiently overlapped otherwise the statistical error due to the combining procedure can be increased too much. A simple harmonic bias potential is used together with WHAM to overcome overlap problems. In order to restrain the system to the target value s_i of the CV $s(\vec{r})$ along the path connecting state A and B, each window is biased with the following harmonic potential:

$$w_i(s) = \frac{1}{2} K (s - s_i)^2. \quad (2.60)$$

The strength of the bias potential K has to be chosen properly to sample the corresponding modes; it has to be as small as possible to allow some overlap between windows.

Summarizing, the implementation of the umbrella sampling method needs a CV well describing the transition from state A to state B and a connecting path; then a biased simulation is performed for a subset of value assumed by the CV along the chosen path. Using WHAM, the global unbiased partition function Q_s is computed

by combining the set of biased partition functions $\{Q_i^b(s)\}$. Finally, the global FES $G(s)$ is computed.

2.4.3 Metadynamics

The metadynamics method was developed by Parrinello and Laio ([54]) in order to accelerate rare events and computing free energies in function of a small set of CVs. The basic idea of metadynamics is to enhance the dynamics of a system along the chosen CVs biasing the dynamics by a history-dependent potential. ([55]). Instead of umbrella sampling, metadynamics does not require previous knowledge of the low-free-energy regions of the system conformational space, but "only" the direction along with the minimum free-energy path develops. Another advantage is that we can simultaneously use several CVs without affecting simulation performance. After the CVs choice, metadynamics simply acts by filling the corresponding energy minimum with an history-dependent bias potential:

$$U(\vec{r}) + W(Y, t) \quad (2.61)$$

where $W(Y, t)$ is a Gaussian of width σ and height w :

$$W(Y, t) = \sum_{t' \text{ min } t} w e^{\left[\frac{(y-y(t'))^2}{2\sigma^2}\right]}.$$

If we consider, for simplicity, a two state process from state A to B in the CV space, the Gaussian are initially localized near the initial configuration (A), that is the first free energy minimum (Figure 2.4).

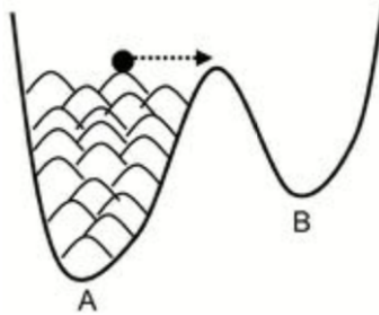


Figure 2.4: Simple cartoon showing the metadynamics concept. A history-dependent bias potential is used to fill the well leading to a transition from the state A to the state B.

When the first free energy minimum is completely filled by Gaussians, the system performs a transition to the second minimum (B). As also the second minimum is filled by Gaussians, the CV starts to freely diffuse between the two minima A and B. By iteratively summing the deposited potential during the biased simulation,

we obtain an estimator of the FES along the chosen CV. The system evolved with metadynamics *tends to escape from any energy minima via the lowest saddle point*. Thus, another advantage of metadynamics compared to umbrella sampling is the possibility to explore new reaction paths by accelerating the observation of rare events. A proper CV leads the system to evolve over the next minimum as it would do in much more MD simulation time. The height w , the width σ and the deposition time τ affect the accuracy and the efficiency of the free energy profile reconstruction.

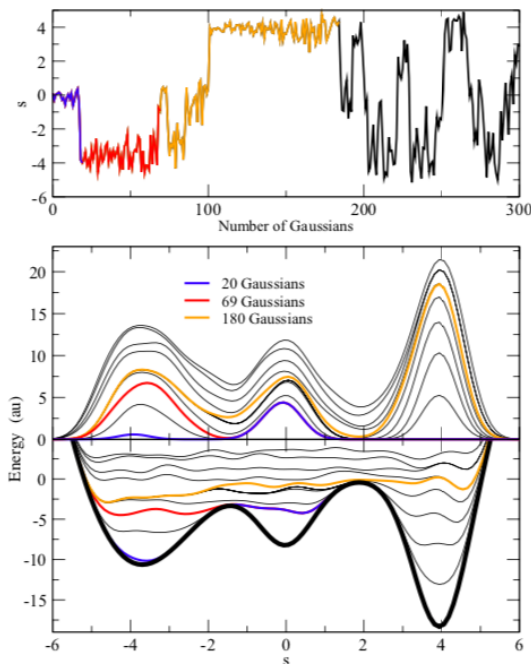


Figure 2.5: Simple cartoon showing the metadynamics concept. A history-dependent bias potential is used to fill the well leading to a transition from the state A to the state B.

Empirical criteria can be used to choose the σ and τ parameters; the former by monitoring the standard deviation of the CVs in an unbiased MD simulation and the latter by observing the relaxation time of the system after a Gaussian deposition. If the Gaussians are big and large or placed too quickly the FES will be explored at a fast pace but the resulting profile will be affected by large errors. Instead, if they are small or placed with too low frequency, the reconstruction will be accurate but it will take more time. In particular, the time required to overcome a local minimum is related to the number of Gaussian necessary to fill the minimum and it is proportional to $(\frac{1}{\sigma})^n$, being n the number of the CVs.

So, we want to maintain the number n of the CVs as small as possible or to increase the Gaussians width σ ; as the number of deposited Gaussians increases, in each MD step, a larger number of exponential terms have to be computed and summed in order to calculate the forces due to metadynamics, i.e. the derivative of the history-dependent potential. The convergence of the metadynamics in a specific region of

the CVs space is reached when the system becomes diffusive, i.e. when the CVs can assume all possible values in the sampled region (Figure 2.5). It is not trivial to understand when the convergence has been reached. A practical method is to monitor the free energy difference between the two reference points: when the difference is approximately flat the convergence has been reached. As regard the error on the FES estimator, it is clear that it depends on the chosen parameters for metadynamics implementation. The error can be estimated by performing statistically independent metadynamics runs. After the convergence has been reached in each run, we can consider as the best estimate of the FES the arithmetic average of all potentials taken at the same time. Instead, in each run, the statistical error of the FES can be considered as the standard deviation between the history-dependent potential $W(\vec{s}, t)$ of that run and the averaged FES $G(\vec{s}) \approx -\overline{W(\vec{s}, t)}$; thus the standard deviation is averaged over the whole CVs space Ω_s as follows

$$\bar{\epsilon}^2 = \frac{1}{\Omega_s} \int_{\Omega_s} \overline{(W(\vec{s}, t) - \overline{W(\vec{s}, t)})^2} d\vec{s} \quad (2.62)$$

Laio et al. ([56]) have derived an approximate expression for the error estimator in function of the system and metadynamics parameters by performing extensive numerical simulations of a Langevin stochastic system:

$$\bar{\epsilon}^2 \propto \frac{LT}{D} \frac{w}{\tau} \sigma \quad (2.63)$$

where L is the size of the simulation box, D is the system diffusion coefficient and T is the system temperature. The ratio $\frac{w}{\tau}$ dominates the error, since the parameter σ is approximately fixed by the fluctuation of the CVs in an unbiased MD run. The use of metadynamics is increasingly widespread in the computational community for its simple implementation and the possibility to control the performance efficiency by tuning the Gaussians and the history-dependent potential parameters. Boichichio et al. ([6]) studied the FES resulting from the transfer of hydrophobic oligomers from the water phase to the hydrophobic cores of a lipid membrane by both umbrella sampling and metadynamics. They found that, if the CVs and the parameters are properly chosen, the resulting free energy profiles are identical between the two methods, but the metadynamics offers the same accuracy in less simulation time.

2.5 Force Fields

Explicit treatment of solvent molecules in biomolecular simulations deals with the time evolution of a very large number of particles i.e. from $10^4 < N < 10^6$. Furthermore, time scales depend on the system under study: from picoseconds for hydrogen bond formation in organic molecules, to seconds for the diffusion of massive colloidal particles. Due to this flexibility in capturing different length and time scales, MD can be used to explore a multitude of objects such as atoms, molecules as well

as macromolecular systems and proteins. The description of such a variety of systems relies on appropriate models of interaction between atoms and molecules, that is - for condensed matter systems - the *Born-Oppenheimer approximation*. This approximation consists in the separation of the atomic nuclei from the motion of electrons. Such a model is necessary if we want to simulate systems bigger than 10^2 atoms; indeed the explicit treatment of the electronic DOF would hugely restrict MD performances.

In the following, we use a *classical model of the inter-atom interactions*, that is we treat nuclei coordinates without considering electrons at all. We call *empirical force fields* (FF) all the functional forms of the interaction potentials between atoms and molecules as well as their parametrization. *Empirical* means that the functional forms derive from a compromise between computational parameters and accuracy, rather than from theoretical principles. The interaction parameters are assigned to reproduce a set of target properties, which are generally derived from measurements or from simulations.

In the context of biomolecular systems, two main FF types have been developed: the *atomistic* FFs in which the building blocks are atoms and *coarse-grained* (CG) FFs in which the basic particles are groupes of atoms. Currently available atomistic and CG FFs present themselves as *packages* suitable for the description of a wide range of compounds both in liquid and in solid phase. The great thing of current FFs is their transferability. This means that even if the FF parametrization has been made to reproduce liquid water features in standard condition, it also able to reproduce non-standard events. In this section, we describe the most widely used functional forms representing inter-particle interactions in MD simulations. The last part of the section focuses on the MARTINI FF [13], the main CG FF used in this research.

2.5.1 The general potential energy function.

In the field of biomolecular applications, the PEF is generally composed by the *bonded* and the *non-bonded* terms:

$$U(\vec{r}) = U_b(\vec{r}) + U_{nb}(\vec{r}) \quad (2.64)$$

The *bonded interactions* contribution involves particles within the same molecules, instead the *non-bonded interactions* contribution represents the Van der Waals and the electrostatic interactions involving all the particles in the system.

In the following, the two contributions are described in detail.

The *bonded* term is usually described as follow:

$$U_b(\vec{r}) = \frac{1}{2} \sum_{\text{bonds}} \frac{1}{2} k_i^b (l_i - l_{io})^2 + \frac{1}{2} \sum_{\text{angles}} k_i^\alpha (\theta_i - \theta_{io})^2 + \frac{1}{2} \sum_{\text{torsion}} V_n (1 + \cos(n\omega - \gamma)). \quad (2.65)$$

The first term models the energy contribution due to the deviation from reference bond length l_{i_0} and the second the deviation from a bond angle θ_{i_0} . The first two terms are both harmonic potentials, the former engages a set of two adjacent particles while the latter a set of three consecutive particles in the same molecule.

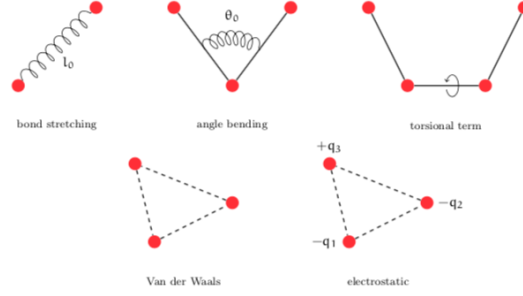


Figure 2.6: Cartoon of the bonded and non-bonded interactions: bond stretching, angle bending, torsional term, Van der Waals and electrostatic interactions.

The last term of Eq. 2.26 represents the energy contribution to the bond torsional change and it involves four consecutive particles in the same molecule mimicking the energy barrier needed to torsion the bond along the bond axis. V_n qualitatively describes the energy barrier for each n -th components with n the number for each components, ω is the torsional angle and γ is a phase factor. The *non-bonded* contribution is described by the following equation:

$$U_{nb}(\vec{r}) = \sum_{i=1}^N \sum_{j>i} (4\epsilon_{ij}((\frac{\sigma_{ij}}{r_{ij}})^{12} - (\frac{\sigma_{ij}}{r_{ij}})^6)) + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}}. \quad (2.66)$$

The first contribution in Eq. 2.66 is the Van der Waals interaction which has the form of a Lennard-Jones 12-6 potential and it depends only on the σ_{ij} and ϵ_{ij} constants assigned to each pair of particles. The last term is the electrostatic energy term which depends on the particle charge q . As specified before, the *non-bonded* contribution engages all the particles of the system; as regard the particles belonging to the same molecule, they are computed only if their interaction are not described by bonded terms thus they are separated by at least three bonds. In Figure 2.6 all the contributions described in the last two equations. On the MD point of view, the calculations of the *non-bonded* interactions energy is the most time consuming part of a simulation. In fact, its calculation scales as $\sim N^2$ even if a simple pairwise additive potential is used. Various methods have been developed to speed-up the simulations. First of all, we can distinguish between short and long range interactions. Depending on the power order d compared to the dimensionality s of the system, we have short range if $d > s$ and long range if $1 \leq d < s$. Particularly, the Lennard-Jones 12-6 potential decays to zero as r^{-6} so it is a short range interaction, while the electrostatic interaction decays to zero as r , being a long range interaction.

In order to treat short range interactions and, sometimes, even long range ones,

the cut-off method was developed. The general idea is to evaluate the non-bonded interactions between one particle and all the others within the cut-off radius r_c . The new potential is thus of the form

$$v^* = \begin{cases} v(r) & r \leq r_c \\ 0 & r > r_c. \end{cases} \quad (2.67)$$

The discontinuity of this potential means discontinuity in its first derivatives i.e. in the forces and the energy conservation is no more satisfied. This problem is fixed by applying a *shift* of the potential value at r_c so that $v^*(r_c) = 0$. The shifted potential is

$$v^* = \begin{cases} v(r) - v(r_c) & r \leq r_c \\ 0 & r > r_c. \end{cases} \quad (2.68)$$

To shift the potential will not affect forces calculation. As regard the forces discontinuity, a simple method is to consider a linear term proportional to the first derivative of the potential as follow

$$v^* = \begin{cases} v(r) - v(r_c) - \left. \frac{dv(r)}{dr} \right|_{r_c} (r - r_c) & r \leq r_c \\ 0 & r > r_c. \end{cases} \quad (2.69)$$

The modifications to the original potential due to the shift methods have to be properly taken into account in order to achieve the correct thermodynamic properties.

An other powerful technique to manage with the discontinuity is the *switch* method. The idea, here, is to consider two cut-off radii r_{c1} and r_{c2} . If $r \leq r_{c1}$, the original form of the potential is used; while if $r > r_{c2}$ the potential is set to zero. If $r_{c1} < r \leq r_{c2}$, a *switching function* is considered to *smoothly* switch the potential to zero.

All the methods used to treat interactions are part of the simulation parameters that are part of the FF. They are interdependent with the model parametrization. It is important to stress out that the simulation parameters should never be changed without retesting the target properties of the parametrization.

2.5.2 Van der Waals interactions

Quantum dynamic effects are described by Van der Waals forces, which include both repulsive and attractive terms. The repulsive term describes the Pauli exclusion principle between electron clouds; the attracting term refers to the dipole-dipole interactions, London dispersion forces, hydrogen bonding and entropy effects. Van der Waals interactions are usually described by a pairwise Lennard-Jones potential which has the following form:

$$v(r) = 4\epsilon \left(\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right) = \frac{C_{12}}{r^{12}} - \frac{C_6}{r^6} \quad (2.70)$$

where $C_{12} = 4\epsilon\sigma^{12}$, $C_6 = 4\epsilon\sigma^6$ and r is the pairwise particles distance. The exponent 6 represents the attractive term of dipole-dipole interactions, thus it has a physical meaning; while the exponent 12 is a computational choice, being easier to calculate r^{-12} as $(r^{-6})^2$. The constants ϵ and σ in Eq. 2.70 are assigned to each particle pair. The former refers to the absolute value of minimum while σ is related to the position of the potential minimum $r_{\min} = 2^{1/6}\sigma$ i.e. the Van der Waals radius. In Figure 2.7, a plot of the function in Eq. 2.70 with $\epsilon = \sigma = 1$ is shown.

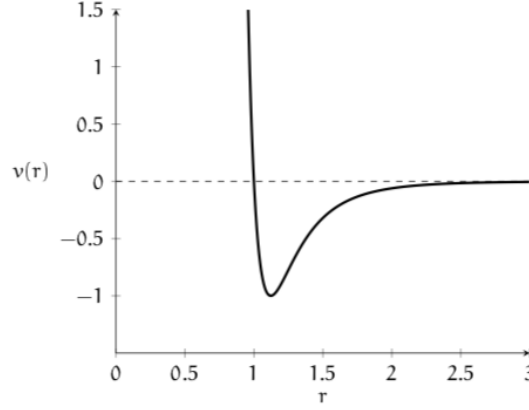


Figure 2.7: Lennard-Jones function with $\epsilon = \sigma = 1$.

As a short range interaction, the Lennard-Jones function can be efficiently treated by a cut-off technique with the shift or switch methods. Since at $r \sim 2$ the function value is less than 1% of the $r \sim \sigma$ value, a good choice for the cut-off radius is $r_c \sim 2\sigma \div 3\sigma$.

2.5.3 Electrostatic interactions

In general, for computational reasons, a common choice is to consider electrostatic i.e. long range interactions as short range interactions by a cut-off method - usually with the same Van der Waals cut-off. Electrostatic energy contribution for the simulation box is

$$U = \frac{1}{2} \sum_{i=1}^N \sum_{j \neq i} \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{r_{ij}} \quad (2.71)$$

where q_i , q_j and r_{ij} are the charges and the distance, respectively, between particles i and j . In order to include also image boxes, we can define a tern of integer numbers (n_x, n_y, n_z) , $n_i = 0, 1, 2, \dots$, so that the position of all other image boxes, with respect to the central simulation box, is $\vec{n} = L(n_x, n_y, n_z)$. The energy contribution becomes

$$U = \frac{1}{2} \sum_{n_x, n_y, n_z}^{+\infty} \sum_{i=1}^N \sum_{j=1}^N \frac{1}{4\pi\epsilon_0} \frac{q_i q_j}{\|\vec{r}_i - \vec{r}_j + \vec{n}\|} \quad (2.72)$$

where the prime indicates that for $\vec{n} = 0$, i.e. the energy contribution of the simulation box, we need to exclude the self interaction term. The main problem for calculation is that the convergence of the summation in Eq. 2.72 is extremely slow, especially for system of the order of $N \gg 10^4$. However, the increasing computing power allows to develop more accurate methods to solve Eq. 2.72.

For a more complete discussion about the advanced methods developed to treat the electrostatic interactions for biological applications, the reader is addressed to the review by [57].

2.6 Coarse Grained Models

Many biological processes regarding lipid membrane and other organic molecules, as well as some mechanisms involving supramolecular polymers, take place on microseconds or milliseconds timescales and tens nanometers length scales. Unfortunately, these time and length scales are not accessible to atomistic FFs. One possible solution is to integrate out some DOF, preserving only those that are relevant for the problem under study: this procedure is called *coarse-graining*. The basic units of CGFFs are called *beads*. Each bead represents a group of atoms or a defined chemical moiety. The coarsening degree of the FF is determined by the bead size. In Figure 2.8 different levels of coarse-graining are showed.

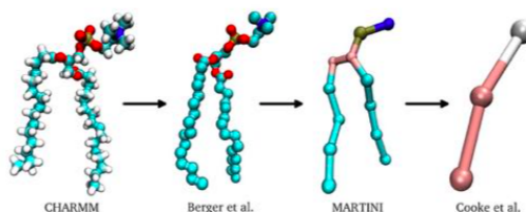


Figure 2.8: From left to right: increasing levels of coarse-graining of a phospholipid.

In order to develop the CG model from the atomistic one, the very first step is the *mapping* procedure. For biological applications, CG FFs are rationally designed to reproduce specific thermodynamic properties; while for material science applications, e.g. polymer science, CG FFs are designed to reconstruct the material structural properties. There is not a unique way to execute the mapping procedure because it depends on various factors as the desired coarse-graining level, the length and time scales of the subject under study and for the material features to be reproduced. Since the DOF of the system are reduced due to the mapping procedure, the CG FF obtained leads to more efficient calculations. Moreover, after removing finer structural details, softer interactions are obtained and this means smoother PEF and faster diffusion. Also the vibrational modes are slower, so in a CG simulation one can use a larger MD time step than in atomistic one.

2.6.1 Martini CG FF

Martini is a CG FF introduced by Marrink et al. [13]. It was initially developed to study lipid membranes and later extended to a variety of other objects as proteins, carbohydrates and a broad class of polymers [15], [58], [59]. Since its creation, Martini has dramatically grown in its transferability becoming one of the most well known CG FF in science community.

The Martini coarse-graining is based on a four-to-one mapping, i.e. on average four heavy atoms are represented by a single interaction centre, with the only exception of ring structures.

There are four main bead types: polar (P), non-polar (N), apolar (C) and charged (Q). Each bead type has several subtypes which allow a more accurate representation of the chemical nature of the underlying atomic structure. These subtypes are distinguished by the hydrogen bonding capabilities: donor (d), acceptor (a), both donor and acceptor (da) and none (o) and/or by their degree of polarity, from lowest polarity (1) to highest polarity (5). In Figure 2.9 examples of mapping procedures are shown, including both atomistic and Martini CG descriptions.

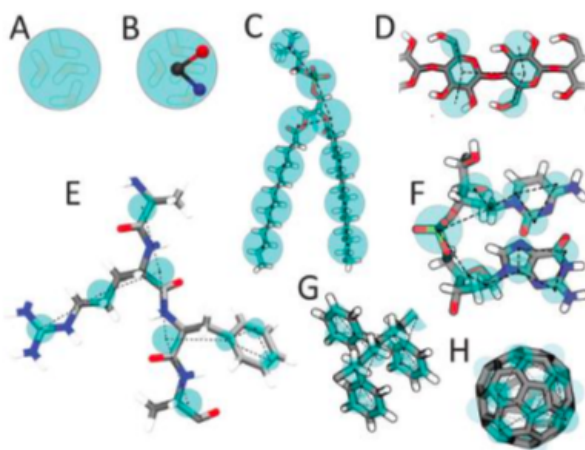


Figure 2.9: Martini mapping procedure for various atomistic structures in which Martini beads are cyan transparent circles underlying the atomistic structure. A) Standard water bead, B) polarizable water, C) DMPC lipid, D) Polysaccharide fragment, E) Peptide, F) DNA fragment, G) Polystyrene fragment and H) Fullerene molecule. Images taken from [60].

The functional form referred to Van der Waals interactions is a Lennard-Jones 12-6 potential as in Eq. 2.70. The interaction has ten levels of strength, reported in Figure 2.10, whereas the strength association matrix is shown in Figure 2.11.

The σ parameter is equal to 0.47 nm except for the Q-C₁ and Q-C₂ interactions for which $\sigma = 0.62$ nm. This value is consistent with hydration shell when a charged bead (a) is dragged into an apolar medium. The charges of the Martini beads are empirically located at the center of the beads and correspond to the net charge of

Level	ϵ [kJ/mol]
O	5.6
I	5.0
II	4.5
III	4.0
IV	3.5
V	3.1
VI	2.7
VII	2.3
VIII	2.0
IX	2.0

Figure 2.10: Interaction strength parameter ϵ for different Martini interaction levels. The last one is for the special case $\sigma = 0.62$.

the chemical compound they represent. Water, for example, is described by a neutral P_4 beads. In order to take into account better water behaviour, e.g. polarizability in lipid membranes, a more sophisticated CG water model, called *polarizable water* (PW) has been developed [61].

	sub	Q				P					N				C					
		da	d	a	0	5	4	3	2	1	da	d	a	0	5	4	3	2	1	
Q	da	O	O	O	II	O	O	O	I	I	I	I	I	IV	V	VI	VII	IX	IX	
	d	O	O	I	O	II	O	O	I	I	I	III	I	IV	V	VI	VII	IX	IX	
	a	O	O	O	I	II	O	O	O	I	I	I	I	III	IV	V	VI	VII	IX	IX
	0	II	II	II	IV	I	O	I	II	III	III	III	III	IV	V	VI	VII	IX	IX	
P	5	O	O	O	O	I	O	O	O	O	I	I	I	IV	V	VI	VI	VII	VIII	
	4	O	O	O	O	O	I	I	II	III	III	III	III	IV	V	VI	VI	VII	VIII	
	3	O	O	O	O	I	O	I	I	II	II	II	II	IV	IV	V	V	VI	VII	
	2	I	I	I	I	II	O	II	II	II	II	II	II	III	IV	IV	V	VI	VII	
N	1	I	I	I	I	III	O	II	II	II	II	II	II	III	IV	IV	IV	V	VI	
	da	I	I	I	I	III	I	III	II	II	II	II	II	IV	IV	V	VI	VI	VI	
	d	I	I	III	I	III	I	III	II	II	II	II	III	IV	IV	V	VI	VI	VI	
	a	I	I	I	III	III	I	III	II	II	II	II	II	III	IV	IV	V	VI	VI	
C	0	IV	IV	IV	IV	IV	IV	IV	III	III	IV	IV	IV	IV	IV	IV	IV	V	VI	
	5	V	V	V	V	V	V	V	IV	IV	IV	IV	IV	IV	IV	IV	IV	V	V	
	4	VI	VI	VI	VI	VI	VI	VI	V	IV	IV	V	V	IV	IV	IV	IV	V	V	
	3	VII	VII	VII	VII	VI	VI	V	V	IV	VI	VI	VI	IV	IV	IV	IV	IV	IV	
	2	IX	IX	IX	IX	IX	VII	VII	VI	VI	V	VI	VI	V	V	V	IV	IV	IV	
	1	IX	IX	IX	IX	IX	VIII	VIII	VII	VII	VI	VI	VI	VI	V	V	IV	IV	IV	

Figure 2.11: Interactions strength association matrix for different Martini types [13]

The bonded contributions include bond length and an angle harmonic term. The former is modeled with a harmonic potential as in Eq. 2.66 in which all bead types have the same bond constant $k^b = 1250 \text{ kJ}/(\text{mol nm}^2)$ and an equilibrium distance $l_0 = 0.47 \text{ nm}$. The latter is a cosine-type harmonic potential

$$U = \frac{1}{2} k^a \left(\cos \theta - \cos \theta_0 \right)^2 \quad (2.73)$$

where $k^a = 25 \text{ kJ/mol}$ and $\theta_0 = 180^\circ$ for aliphatic chains; $k^a = 45 \text{ kJ/mol}$ and $\theta_0 = 120^\circ$ for *cis* double bends and $k^a = 45 \text{ kJ/mol}$ and $\theta_0 = 180^\circ$ for *trans* unsatu-

rated bonds. In some case, e.g. ring systems, a dihedral angle harmonic potential can be used:

$$U = k_{\text{id}} \left(\theta_{\text{ikjl}} - \theta_0 \right)^2 \quad (2.74)$$

where θ_{ikjl} refers to the angle between the planes described by atoms i,j,k and j,k,l; k_{id} and θ_0 are the force constant and equilibrium angle, respectively.

Although CG FFs, so Martini FF as well, are computationally advantageous, there are still some limitation, compared to the atomistic models, in both chemical and spatial resolution. A consequence of the DOF reduction process is the underestimation of the entropy respect to the atomistic case. Another limitation of CG FFs refers to the FES that becomes smoother. This results in more sampling of the energy landscape in a given time period, speeding-up the dynamics of the system. The speed-up is not easily predictable and probably it depends on the type of molecule. Moreover, in order to increase the efficiency, all the Martini beads have the same mass of 72 amu: this makes the time scaling for different beads non-trivial. In the Martini FF, for example, there is an average scaling factor of four, based on lateral diffusion coefficients of lipids in membranes, is usually adopted.

A specific problem related to the Lennard-Jones potential is the water - modelled as P_4 particles - melting point at 290 ± 5 K. A partial solution is the introduction of the BP_4 beads, called *anti-freeze* particles: a mole fraction of $n_{\text{af}} = 0.1$ is sufficient to prevent freezing.

The reader is addressed to the review by Marrink et al. for a more comprehensive discussion about the Martini FF and its limitations [60].

Chapter 3

Controlling exchange pathways in dynamic supramolecular polymers by controlling their defects.

3.1 Supramolecular polymers: a need for finely tuned coarse-grained models.

The reversible nature of supramolecular materials offers numerous advantages such as the ability to respond to physiological stimuli or to mimic structural and functional aspects of cellular signaling. Historically, in medicine, materials such as ceramics, metals and polymers have been used to repair or replace biological elements [17], [62]. These materials have the limit of being mere structural copies of native materials and cannot reproduce their adaptive and regenerative abilities. The next generation materials will be based on the principles of supramolecular chemistry. Instead of covalent bonds, supramolecular structures exploit H-bonding, hydrophobic, π - π and Van der Waals interactions. The peculiarity of these bonds is precisely their reversible and dynamic nature. Taken individually, non-covalent interactions may be weak, however when these interactions add up or are directed, it is possible to generate stable materials with unique properties.

Supramolecular materials are extremely tunable, dynamic and - a key feature - modular in nature [63],[64],[4]. Thanks to these characteristics they are the best candidates to mimic biological objects and thus launch the next generation materials for biomedicine and nanobiotechnology. To reach the controlled design of supramolecular materials, it is necessary to fully understand the nature of the non-covalent bonds underlying these structures, their strength and the complex interplay between all the different interaction. A thorough understanding of the cooperation mechanisms of the constituent entities is fundamental for producing materials with properties similar to those of biological structures. Handling cooperation means intervening on materials at the molecular level. This is where self-assembly comes into play. As already explained in the previous chapters, it is a free-energy driven process that favors the spontaneous organization of ordered and hierarchical assemblies. The final goal therefore becomes to reach a self-assembled structure capable of performing the

function it was designed for.

β -sheet-forming peptides, for example, self-assemble into 1D structures - through intermolecular H-bonding - which, in turn, can form 3D networks. The peptides are also biocompatible and biodegradable; this means that we can induce them to artificially mimic the functions of large proteins through the rational control of their structures [62]. Part of the functionality of proteins derives, in fact, from the folding-unfolding mechanisms, that is from their ability to modify their configuration. Through the structural versatility of self-assembled materials we can hope to produce bio-materials with an unprecedented biological relevance. Another interesting example is that of cellular microtubules (MTs) - composed of self-assembled tubulin units - capable of regulating the motion and differentiation of cells through consecutive polymerizations and depolymerizations. This process consists in losing a unit from one end and recovering one from the other repeatedly and constitutes the origin of the mechanical forces thanks to which the MTs are able to carry out their key functions for the cells.

Artificial supramolecular polymers aim at a dynamic nature similar to that of MTs. Controlling, for example, the polymerization / depolymerization rates of the polymer leads to different degrees of adaptivity, responsiveness to external stimuli and the ability to reconfigure. To differentiate the dynamic properties and exchange pathways, the structure and / or composition of the single monomers from which the general behavior of the polymer derives, must be designed. In this way it is possible to control the exchange modalities of the monomers; these can detach from the tip of the polymer or from its lateral surface [18], [19]. Being able to code the response to a specific stimulus with a specific exchange pathway would generate usable materials for eg. molecular signaling [65], [45]. This feature also implies the ability of the polymer to adapt differently depending on the surrounding environment.

To date, it is still experimentally prohibitive to distinguish the precise dynamics involved in the exchange mechanisms of the supramolecular structures with sufficient space-time resolution. Despite the enormous technical progress of recent years, the information obtainable from experiments alone is not sufficient to identify the kinetic secrets of these assemblies [66]. Once again microtubules represent an interesting case study; in fact they constitute the starting point for the creation of supramolecular machines as they are able to generate mechanical forces on a nanometric scale [67].

Building machines of this type means being able to control the assemblies even in states out of equilibrium and manage the energy dissipation processes. Azobenzene, for example, is a molecule capable of switching its configuration from trans to cis isomers in response to light stimuli. The response of these molecules to energy absorption leads to a progressive increase in the defects present along the structure up to total disassembly [68]. Submolecular resolution is needed to investigate the defects creation mechanism. Defects play a central role in the dynamics of supramolecular structures, in fact they are entrusted with the task of dissipating the energy absorbed from the external environment - in the form of light, temperature, mechanical stimuli

- through specific kinetic pathways [69].

3.2 BTA supramolecular polymers.

One-dimensional fibers in various solvents can be an excellent starting point for understanding the structure-dynamics relationships between the constituent monomers and the resulting supramolecular architecture. The monomers contain all the information that will be transmitted in the polymer, in its intrinsic dynamics and therefore in the exchange pathways. A very well studied type of fiber is the water-soluble 1,3,5-benzenetricarboxamide (BTA) obtained from the directional stacking of the BTA monomers core through threefold H-bonding. While the microtubules, described in the previous paragraph, exchange monomers with the solvent only through the tips, in some cases the exchange can occur continuously along the entire fiber [18], [19], [70].

Albertazzi et al. [71] showed this trend with STORM experiments; when different fibers are immersed in solution it is also possible to use FRET mixing assay to measure the exchange equilibration time between the various polymers [66]. Studies on BTAs are also popular due to the simplicity with which it is possible to change side-arms of monomers in order to obtain soluble fibers in other media [22], [72]. Indeed, adding a stereogenic methyl center to each side arm (C12-PEG4-OH) of the monomers yields fascinating results. The new monomers self-assembles in one-dimensional supramolecular fibers but exhibit chirality, unlike fibers of the first type which do not exhibit any chirality.

The two building block monomers, albeit slightly different, can lead to two resulting fibers with distinctly different dynamic behavior. The two fibers are practically indistinguishable through cryo-TEM and SAXS experiments and both appear as single-columns of BTAs with no branches or superstructures. Using super-resolution STORM imaging instead - whose resolution starts at around 25 nm and goes down to tens of microns - once again the fibers appear very similar, both with high aspect ratios. The only additional info concerns their length: the chiral fiber appears shorter than its achiral counterpart. The cross-sections of the fibers are however out of the reach of the SAXS, TEM and STORM techniques and therefore from these measurements it is not possible to obtain information on the intrinsic dynamics of these structures. Instead, by combining two samples of each BTA-fiber with a Cy3 and Cy5 FRET pair, mixing the two samples and measuring the FRET ratio, we obtain estimates of the monomer exchange times for the two fibers. The results show that the exchange dynamics are much faster in the achiral fiber. Furthermore, the fitting is compatible with a bi-exponential function, suggesting that the exchange mechanism between monomers is not a single-step process.

The difference in timescale between the two exponentials is at least one order of magnitude for both fibers and this data could indicate that, regardless of the speed of the process, the steps associated with the exchange could undergo the

same mechanism [69]. As can be seen from the measures described above, with the experimental techniques available today, it is not possible to obtain information at the molecular level for supramolecular systems and therefore it is still prohibitive to extrapolate the physical mechanisms underlying the exchange processes with experiments alone.

Molecular dynamics simulations have provided valuable insights on the molecular structure of these stacks as well as on the dynamics and thermodynamics of the processes involved.

3.3 Multiscale modeling of BTA-based supramolecular polymers.

In the previous paragraphs it emerged that the structure of the BTA monomers and their interactions are so inaccessible, especially if the polymer is immersed in water. Bejagam et al. [73] studied fibers of BTAs with C6 side chains in nonane solvent using All Atom (AA) MD simulations. Nonane is an apolar organic solvent and the nature of the monomers leads to the formation of ordered stacks since the solvent do not disturb their interaction. On the contrary, BTA-based polymers in water appears more disordered and it could suggest a more complex interaction between monomers as well as with the solvent.

The two BTA-based supramolecular fibers described in the previous section were also analyzed through AA-MD simulations [23]. In this way, the differences that emerged from the experiments - such as the chirality and the different exchange rate of monomers - can be treated at a much higher resolution [74]. However, the time-scales related to the assembly and exchange processes of single monomers are still inaccessible to atomistic models. Coarse grained force fields are a good strategy to overcome these limitations. The simplification of the models and the consequent access to previously unexplored time-scales, have made some CG force fields, such as MARTINI, very used [13].

The spontaneous self-assembly of supramolecular fibers, lipid bilayers, vesicles, micelles and proteins, to name a few, was studied through CG models [13] [15] [11]. The MARTINI coarse-graining strategy is now a widespread practice for studying self-assembly processes and the resulting supramolecular architectures; in fact, a large collection of widely tested chemical groups, lipids and small molecules is available. Ad hoc CG models were also created to investigate the BTA monomers polymerization process in explicit apolar solvent [25]. The ad hoc models can be very precise in the representation of the single structure, but remain limited in the possible implementation of the model, as well as in the possibility of studying the interaction of the model in question with a variety of biomolecules.

The advantages of using MARTINI are its high transferable parameterizations and the numerous components already present in its database and in the literature

([15] [11] [14]). Inspired by the work of Shatz et al. [10] on supramolecular fibers of MARTINI CG amphiphilic peptide monomers, Boicchio et al. developed a CG MARTINI model for BTA-C12-PEG and BTA-C6 monomers generating supramolecular fibers in water and in organic solvent respectively [75].

In these two models the representation of the BTA-core is fundamental as it is responsible for the directionality of the bonds between the monomers by means of the hydrogen bonds they form. To reproduce this behavior, each MARTINI bead amide comprises a dipole, i.e. two charges ($\pm q$) at a fixed distance [75].

The coarse-grained system gives access to longer time scales, in fact it is possible to monitor the polymerization process directly during the simulation. To confirm the results, BTA-C6 monomers form fibers with a slower dynamics, greater persistence of the aggregates than BTA-C12-PEG fibers and consequently, different dynamics of monomer exchange [23]. Thanks to the coarse-grained models combined with advanced sampling techniques, it is possible to state that the exchanges with the solvent occur - in both cases - from the defects present along the polymer.

3.4 The role of defects in supramolecular polymers.

In the previous paragraphs the fascinating dynamics of BTA-based supramolecular fibers has been highlighted: these aggregates continuously exchange monomers with the solvent. The properties that result from this dynamic behavior are the reason why these directional architectures are so interesting and yet so difficult to study. Controlling the ability of these structures to exchange with the external environment and then reconfigure themselves would mean opening the way to new materials capable of self-repairing or readjusting following an external stimulus.

Using coarse-grained models it is possible to reproduce the exchange pathways, compare their time-scales and take a first step towards the rational-design of supramolecular polymers. Although the simulations of MD-CG give access to a high spatial resolution, the exchange of a monomer is a rare event, though inaccessible to the typical times of simulation [76], [77], [78], [79], [80]. A technique developed in recent years to access more extended time-scales is well-tempered metadynamics (WT-MetaD) which can provide the free energy profile of our exchange event [81]. The simplest fiber to use as a case study is BTA-C6 in organic solvent - already introduced in the previous paragraphs - which forms sufficiently long and ordered fibers (Figure 3.1) [22], [23], [73].

The starting point is to investigate in detail the exchange of a single monomer with the solvent. Through the WT-MetaD (see chapter 2) it is possible to activate the exchange of a single monomer. The runs show how the monomer leaves the fiber and is subsequently re-incorporated by the polymer. To allow the monomer to leave the stack, the fiber first creates a local defect, then the monomer only sticks on one side and, in a second step, the monomer detaches completely. The two collective variables used in WT-MetaD to calculate the free energy profile of the monomer that

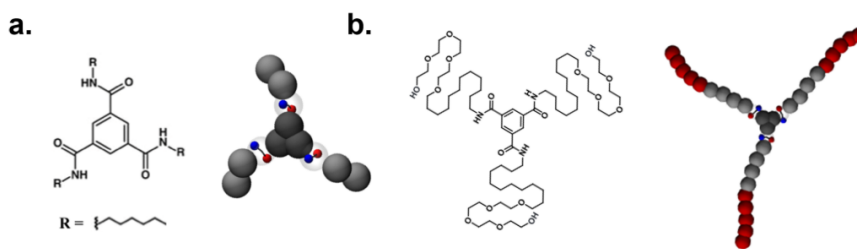


Figure 3.1: Chemical structure and CG model of a) BTA-C6 monomer and b) BTA-C12-PEG monomer generating supramolecular fibers in organic solvent and in water, respectively.

leaves the stack are the stacking distance and the coordination between monomers.

The stacking distance between two monomers for the equilibrated fiber has a value of $c \sim 3.5$ Å; the coordination between monomers is worth about 2 when the monomer belongs to the bulk while it is worth about 1 when the monomer is one of the two tips of the stack, or when a defect has been created and is therefore attached only on one side. The figure Figure 3.2 shows the free-energy surfaces obtained from the exchange process as a function of these two parameters.

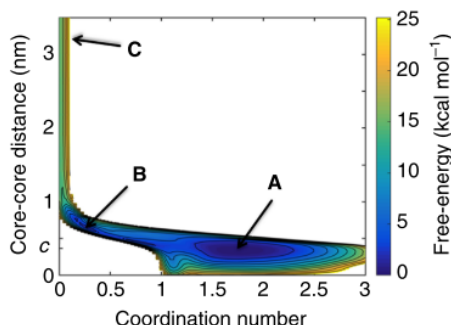


Figure 3.2: Free-energy surfaces (FESs) for monomer exchange as a function of the minimum stacking distance between monomer cores and the of core-core coordination. The darker areas identify the energetically favourable regions for the exchange events; on the other hand, lighter areas corresponds to less probable exchange pathways.

The metastable states of the monomer exchange pathway are identified by the free energy minima that correspond to the darker areas in the FES. The exchange of the monomer with the solvent never occurs directly (AC) but the creation of the defect must first occur (B). The lowest free energy minimum corresponds to the stack ordered with $c \sim 3.5$ Å and monomer coordination about 2. The same two-step exchange process is observed regardless of the solvent [82].

The complexity of the aggregate increases considerably if we consider BTA in water, both for the size and for the flexibility of the water-soluble monomers. Furthermore, as we will see in the next paragraphs, the hydrophobic effect plays a fundamental

role in the behavior of the supramolecular structure in water. The water soluble BTA-C12-PEG supramolecular fibers appear as a fiber as well, but along the entire stack there are monomers attached to the assembly only on one side; in other words, the fiber has defects along its entire length. In the BTA-C6 fiber the only defects present are the two tips.

Albertazzi et al. ([71]), combining STORM microscopy with stochastic modeling they showed how these polymers exchange homogeneously along the entire fiber (Figure 3.3). The FES have shown that the coordination value between monomer cores is a good parameter to describe the state of a monomer within an ordered stack such as BTA-C6. In fact, a monomer immersed in the solvent has coordination 0, a fiber tip or a defect has coordination 1, while a monomer belonging to the bulk has coordination 2. In terms of coordination, the creation of a defect corresponds to the formation of a new tip, therefore the coordination passes from value 2 to value 1.

Figure 3.3 represents the two steps of the exchange: the creation of a defect along the fiber and the exchange of a tip or, equivalently, of a defect with the solvent. The transition time distributions for these two events were obtained through multiple runs of infrequent WT-MetaD, in which monomer exchange was activated. The collective variable (CV) chosen to deposit the bias (see chapter 2) is the average number of contacts between the core of the exchanging monomer and all other monomer cores along the fiber.

Each run provides a value for the transition time of the triggered event. The characteristic transition time can be extrapolated from the profile obtained from all transition times, which is the typical profile of the cumulative Poisson distribution.

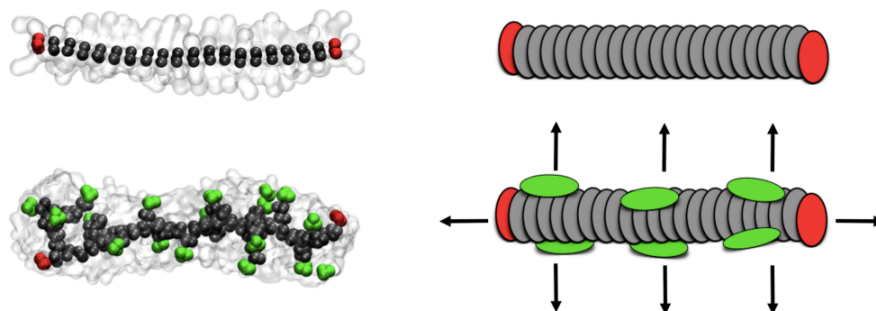


Figure 3.3: Equilibrated coarse-grained models of octane- (top left) and water soluble BTA fibers (bottom left. On the right, cartoons show a stack without defects (top) a stack with bulk defects (green) from which monomers can always exchange (bottom)). Both cartoons show fiber tips (red) considered as defects.

Thus, the probability of having at least one trade event in time t is $P_{n \geq 1} = 1 - e^{-\frac{t}{\tau}}$. The characteristic time τ of the event corresponds approximately to the midpoint of the Poissonian sigmoidal curve.

The cumulative Poisson distributions associated with the BTA-C6 fiber for i) the

creation of a defect, in green and ii) the exchange of a monomer from the tip in red, are shown in the Figure 3.4(top). Creating a bulk defect is about 4 orders of magnitude slower than swapping a monomer from tip. The characteristic times (gray lines) are respectively $\tau_i = 10^4$ and $\tau_{ii} = 10^{-1}$.

While the values of the single characteristic times cannot be realistic - as they are obtained from a coarse-grained model (see chapter 2) - their comparison, as well as the comparison between the two different models, instead constitutes a reliable information. The graph below of Figure 3.4 shows the cumulative Poisson distributions for a BTA C6 fiber in which the value of the dipoles charges in the amide beads has been lowered from $\pm 0.8e$ to $\pm 0.65e$.

In this way the characteristic times of processes i) and ii) are approached and the ratio is reduced almost to half compared to the previous fiber (from ~ 4 to ~ 2.5 orders of magnitude).

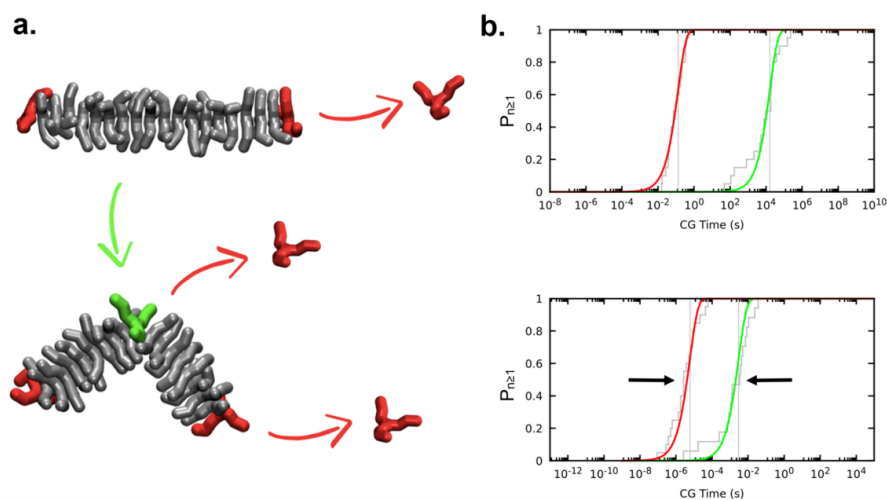


Figure 3.4: a) Cartoon representing the two stepwise process of BTA supramolecular polymers in organic solvent: first, a defect in the bulk (new tip) has to be created and, then, the monomer can exchange with solvent. b) Cumulative Poisson probability distributions ($P_{n \geq 1}$) for the rare events of monomer exchange from the fiber tip (red) and of the generation of a bulk defect along the fiber (green). The distributions refer to a BTA-C6 fiber in octane (b, top) and to the same fiber with artificially weakened directional interactions between the monomer cores (b, bottom). the grey vertical lines represent the characteristic time scales (τ values) for the events. In the latter case (b, bottom), the τ values for the exchange from the tips and the creation of a bulk defect become closer; a similar effect can be obtained by artificially increasing monomers solvophobicity (see Figure 3.5).

The result obtained shows that it is easier for the BTA-C6 fiber to exchange along its entire length following the weakening of the interaction between the BTA hearts. When we lower the charges of the dipoles we are favoring the non-directional

interactions between the monomers to the disadvantage of the directional ones. If we favor non-directional interactions, the fiber will tend to form a more disordered assembly, i.e. with a greater number of defects. On the other hand, favoring directional interactions means increasing the order of the stack and therefore the presence of defects is reduced and the exchange of a bulk monomer becomes slower.

The hypothesis is that by regulating the competition between directional and non-directional interactions we can control the amount of defects and the data reported in Figure 3.4 support this hypothesis. Specifically, the presence of defects can be tuned in two ways. The first - already illustrated above - consists in weakening the directional interactions (charges of the dipoles from $\pm 0.8e$ to $\pm 0.65e$); the second consists in increasing the non-directional interactions by acting on the composition of the side chains making, for example, the beads more solvophobic.

Figure 3.5 shows the Poissonians for a BTA-C6 fiber whose side chains have been made more solvophobic by replacing the beads (SC5 Martini beads instead of SC1). The difference in the characteristic times i) and ii) is quite similar to the case of BTA with reduced charges.

It should be noted that for BTA fibers with a length greater than 70 – 700 nm, the exchange of monomers from the bulk would be statistically the preferred pathway. Thanks to these results it is evident that the defects are directly connected to the exchange pathways accessible to a fiber.

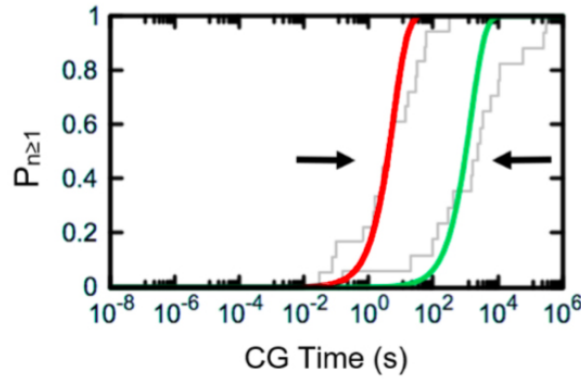


Figure 3.5: Cumulative Poisson probability distributions ($P_{n \geq 1}$) for the rare events of monomer exchange from the fiber tip (red) and of the generation of a bulk defect along the fiber (green). The distributions refer to a BTA-C6 fiber in octane with artificially increased monomers solvophobicity. The τ values (vertical grey lines) for the exchange from the tips and the creation of a bulk defect become closer respect to the case of the BTA-C6 fiber without modifications (see Figure 3.4 b, top). A similar effect can be obtained by artificially weakened directional interactions between the monomer cores (b, bottom) (see Figure 3.5 b, bottom).

Finding a relationship between the defects population and the competition between directional and non-directional interactions allows us to take a step towards the rational design of polymers with specific exchange pathways. Unfortunately, with

BTA-C6 CG model we cannot explore other types of behavior due to the limited possibility of customizing the model and the reduced flexibility of the lateral chains. In fact, if we increase the solvophobicity of the side chains a little more, we do not obtain fibers with a greater number of defects, but nanoparticles, thus 3D aggregates. The fibers that present a greater number of defects are, for example, the BTA-based fiber soluble in water as they have long and flexible side chains (PEG) that act as a solvophilic shield to the internal solvophobic beads, allowing the monomers to form a linear aggregate.

3.5 Minimalistic CG models of supramolecular polymers.

To generalize the obtained results, we developed a minimalistic model that allows to describe supramolecular fibers with three lateral arms. Being a coarser model, there is no chemical accuracy and the monomers are treated in a more abstract way, drastically reducing the number of parameters that describe them. In fact, the advantage of this representation consists in the possibility of accessing different variants of supramolecular polymers simply by intervening on a parameter i.e. by modifying the solvophobicity of the internal beads of the monomers.

Instead of the three dipoles present in the BTA-C6 model, the directional interactions are entrusted, in this model, to a single dipole placed at the center of the monomer structure Figure 3.6 a. The 3 final beads of each arm are always solvophilic (white bonds in figure), while the two internal beads (dark gray bonds) are variable - that is, they can be taken more or less solvophobic - depending on the polymer variant we want to obtain. Specifically, three fiber variants were compared, each made up of forty identical monomers, in which the only difference is the choice of the degree of solvophobicity of the internal beads.

In the first type of monomer (monomer 1), the structure is totally solvophilic: the beads are all identical to those of the solvent (C1 Martini beads); self-assembly takes place by means of the directional interactions carried out by the single central dipole of the monomers. In the second type of monomer (monomer 2), on the other hand, the two internal beads of each side chain are replaced by C5 Martini beads, which are more solvophobic.

In the third type of monomer (monomer 3), the solvophobicity of these beads is further increased with the choice of the N0 Martini beads, which are even less similar to the solvent. As regards the central dipole, the charges have been calibrated in order to reproduce, for monomer 3, the free energy of dimerization for the water-soluble monomers BTA-C12-PEG which is about 10 kcal mol^{-1} and left unchanged in the other two cases [75], [69].

With this choice, the fiber obtained from the monomer 3 (fiber 3) has a behavior in the solvent similar to that of the supramolecular polymers BTA-C12-PEG (Figure 3.1 b) [23], [74], [75], [69]. Instead, fiber 1 looks similar to BTA-C6 fiber in octane solvent (Figure 3.1 a). Finally, fiber 2 shows a behavior that is halfway between fibers 1 and

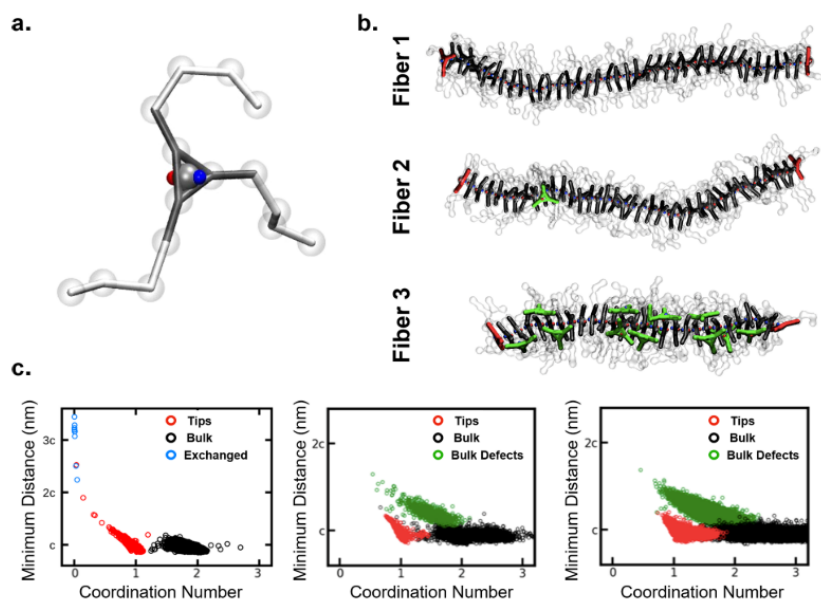


Figure 3.6: a) Minimalistic CG model for the 3-arms self-assembling monomers studied herein. The directional interactions between the cores is provided by a dipole (blue and red) in the central bead. b) Three variants of 40-monomers stack obtained by modifying directional vs non-directional interaction balance: Fiber 1 (completely solvophilic monomers self-assemble only due to directional interactions), Fiber 2 (internal monomers beads slightly more solvophobic than in Fiber 1), Fiber 3 (internal monomers beads more solvophobic than in Fiber 2). c) Supervised clustering analysis of the three fibers 1, 2 e 3 (left to right) showing different structural motifs: bulk monomers (black), fiber tips (red), bulk defects (green) and exchanged monomers (blue).

3.

Figure 3.6 b shows the three fibers obtained from 3 unbiased runs from 30 μ s of the perfect stack of each type of monomer. The equilibrated stacks of the 3 fibers have an increasing number of defects as the two internal beads become more solvophobic; fiber 1 appears linear and without defects, fiber 2 has some defects in the bulk and fiber 3 has numerous persistent defects. At equilibrium, fiber 2 presents a very interesting dynamics of defects since they are not persistent but are formed and then reabsorbed.

We used supervised clustering analysis of the equilibrated trajectories of the three stacks to study the degree of fiber defectiveness. Through a spectral clustering-based algorithm (see chapter 2) it was possible to quantitatively describe the structural behavior of the 3 aggregates by exploring, in particular, two collective variables calculated for each monomer: the coordination number and the minimum distance from the other monomer cores. The coordination number is 2 if the monomer is perfectly aligned with the other monomers of the assembly, so it has precisely 2 neighbors; the minimum distance between two consecutive monomers, in the case of ordered stacking, is c . The clusters were then projected on these two variables and are shown in Figure 3.6 c where the different physical states of the monomers are perfectly distinguishable.

The monomers belonging to the stack perfectly aligned with coordination number 2 and minimum distance 1, are colored black. The monomers belonging to the tips, with coordination number 1 and minimum distance c , are represented in red. All the monomers that during the biased run spontaneously leave the fiber and exchange with the solvent - represented in blue in the clusters - have a coordination number equal to zero and a minimum distance greater than $2c$. Bulk defects have coordination number values between 1 and 2 and minimum distance values between c and $2c$ and are represented in green. Each fiber leads to different clusters for the various monomer populations and therefore different structural motifs.

From the analysis it is evident that the fiber 1 exchanges only from the tips and never forms defects; if it trades from the bulk it is only after a break of the stack. The fibers 2 and 3 instead exchange along the entire length and both have bulk defects (in green). In particular, the defects present for equilibrated fibers 2 and 3 are ~ 0.8 and ~ 12 , respectively. The value lower than one for the number of defects in fiber 2 indicates, as has already been observed, that the defects do not persist over time, unlike fiber 3 which has intrinsic defects distributed along the entire stack. The fact that the blue cluster is not present in fibers 2 and 3 is a consequence of the solvophobicity of the internal beads of the monomers; as the solvophobicity in the interaction between monomers increases, the general dynamics of the fiber slows down, as well as the exchange events of the monomers [69].

3.5.1 High-dimensional analysis with abstract unsupervised machine learning approach.

Another interesting method for investigating the dynamics of the defects of the three fibers is machine learning [83]. In particular, a high-dimensional analysis was used - SOAP: smooth overlap of atomic positions, vectors - combined with an unsupervised density-based clustering technique - PAMM: probabilistic analysis of molecular motif - through which it is possible to highlight the representative clusters of the physical states of the monomers of the system. The advantage of this analysis is that it requires neither collective variables for grouping the various clusters, nor the number of clusters in advance. Through these techniques, we can adopt a bottom-up approach, considering all the cores of the monomers belonging to the three fibers as a single data set.

In this way we can identify the different clusters associated with the states of the monomers in a general and equivalent way for the three fibers: the macro-clusters obtained from these analysis show structural characteristics consistent with those of Figure 3.6 i.e. the monomers exchanged with the solution, fiber tips, defects and stacked monomers (Figure 3.7). Figure 3.7 shows the free energy projected on the two principal components PCA1 and PCA2 of the macro-clusters indicated in the image by dashed ellipsoids.

The SOAP-PAMM technique also provides data on exchange rates, which allow us to compare the fibers with each other and also with the spectral clustering analysis described in the previous paragraph. The black arrows in Figure 3.7 indicate that the monomers are constantly changing their state and the rates of these transitions.

In the central graph, referred to fiber 2, the global minimum is represented by the black cluster, i.e. by the monomers of the bulk, while the green cluster associated with the defects has a free energy value of only $\sim 2 \text{ kcal mol}^{-1}$ higher than the minimum. This explains why fiber 2 has no permanent defects, but these are statistically formed along the bulk; in detail, the green-black pass rate is about $14 \mu\text{s}^{-1}$, while black-green is about $0.3 \mu\text{s}^{-1}$, values that lead to a defects annihilation / creation ratio of about 40.

In fiber 3, on the other hand, it is interesting to note that there is a green minimum and therefore that the defects are a persistent condition and the two rates are of the same order of magnitude (ratio ~ 2). Consistent with the analysis of Figure 3.6, fiber 1 has no green cluster. The analyzes described in the last paragraphs provide a detailed picture of the distribution and dynamics of defects in a supramolecular polymer. The presence of non-directional interactions favors the appearance of defects in the bulk, while their absence leads to ordered and aligned stacks.

Being able to monitor the quantity of defects is interesting because they constitute the hot spots from which the monomer exchanges take place; defects are therefore the way through which the exchange pathways of the supramolecular aggregates are activated.

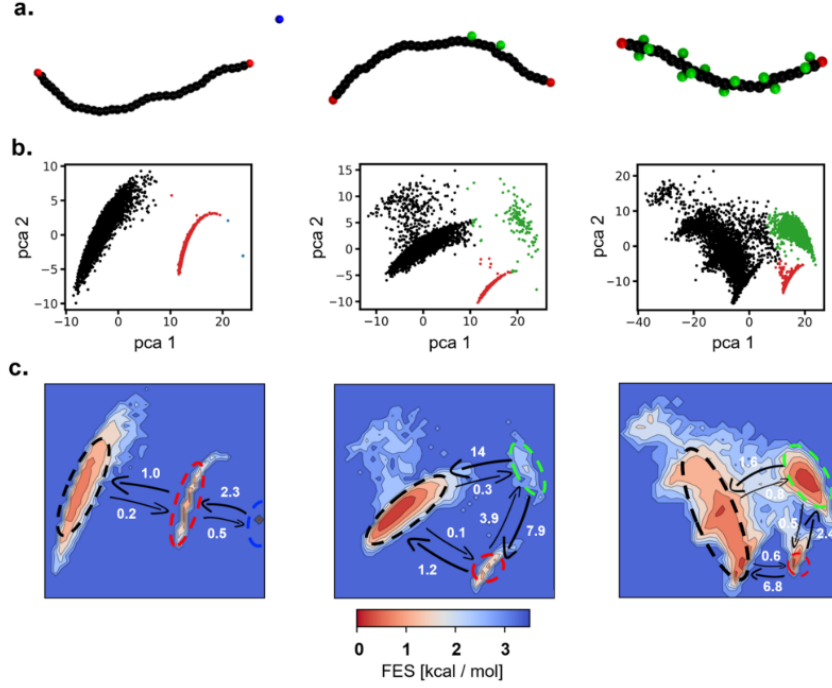


Figure 3.7: Unsupervised machine learning of defects and of defect dynamics. a) Minimalistic representation of the three fiber models, showing only the center for each monomer in the fibers. The monomers centers are colored according to the cluster they belong to (black: ordered bulk of the fiber, red: fiber tips, green: bulk defects, blue: exchanged monomers). b) First two principal components (PCA1 and PCA2) obtained from the dimensionality reduction of the SOAP analysis of the equilibrium CG-MD trajectories of the three fibers. Scatter plots are colored according to the main macro-cluster obtained from the PAMM analysis. c) Low dimensional free energy surfaces of the three fibers, computed from the monomer states distribution of panel (b), showing the kinetic analysis of the relative transition rates between the macro-clusters. Each macro-cluster is reported in the FES as a dashed ellipsoid, roughly corresponding to the area of that cluster. The arrows represent the transition rates for the interconversion of macro-clusters along CG-trajectories; which are left out for clarity (CG transition rates, having a comparative value.)

3.6 Origin and dynamic behavior of defects: determining the most favorable exchange pathway.

The aim of this work is to effectively predict the exchange pathways of a supramolecular polymer. We have seen how the relative intensity of the directional and non-directional forces between the monomers depends on the solvophobicity of the monomers themselves and that an increase in solvophobicity leads to a greater abundance of defects. The interesting part comes when it is understood that the defects regulate the favorite exchange pathways for the polymer.

We define the dimensionless parameter α as follows:

$$\frac{N_{\text{def}} K_{\text{def} \rightarrow \text{sol}}}{N_{\text{tip}} K_{\text{tip} \rightarrow \text{sol}}} \quad (3.1)$$

where N_{tip} is the number of tips in the stack which by definition equals 2. N_{def} is the average number of defects in the bulk; this quantity was estimated through block average with a block size of 5 μs via supervised clustering (Figure 3.6 c) and unsupervised SOAP-PAMM classification (Figure 3.7). Both methods give $N_{\text{def}} < 1$ for fiber 2 and $N_{\text{def}} \sim 12$ for fiber 3.

At equilibrium, $K_{\text{def} \rightarrow \text{sol}}$ and $K_{\text{tip} \rightarrow \text{sol}}$ are the exchange rates between a bulk defect and the solution and between a tip and the solution, respectively. In Figure 3.8 a the possible exchange pathways for a supramolecular polymer are represented: the exchange from the tip (red arrow) and the exchange from a defect along the fiber (green arrow). Through unbiased CG-MD simulations it is difficult to observe these exchanges as they constitute rare events, even more so when we increase the solvophobicity of the monomers as in fibers 2 and 3.

As previously done with BTA fibers and other supramolecular polymers ([84], [69], [85], [86]), the infrequent WT MetaD was used to activate the exchange of monomers both from the tips and from the bulk. The statistics collected for the three fibers are shown in Figure 3.8 b-d, being a good fit of the Poissonian for rare events, from which it is possible to obtain the unbiased characteristic transition times τ , indicated in Figure 3.8 b-d with a vertical grey line ([69], [79], [76]).

The red curves represent the cumulative Poisson distributions for the exchange of a monomer from the tip, while the green distributions represent the exchange of a monomer from the bulk defect. In Figure 3.8 b there is only the curve for the exchange from the tips consistent with the behavior of fiber 1 without defects. Furthermore, the characteristic time τ obtained by the infrequent WT MetaD for fiber 1 is of the order of microsecond, therefore compatible with the result obtained through unbiased CG simulations (Figure 3.4 b-top). Fibers 2 and 3 show both exchange pathway curves.

From Figure 3.8 c-d, it can be seen how the relative distance of the characteristic times for the two exchange pathways is reduced in the case of the fiber 3 with respect to the fiber 2; this means that as the defects increase, the two modes of exchange

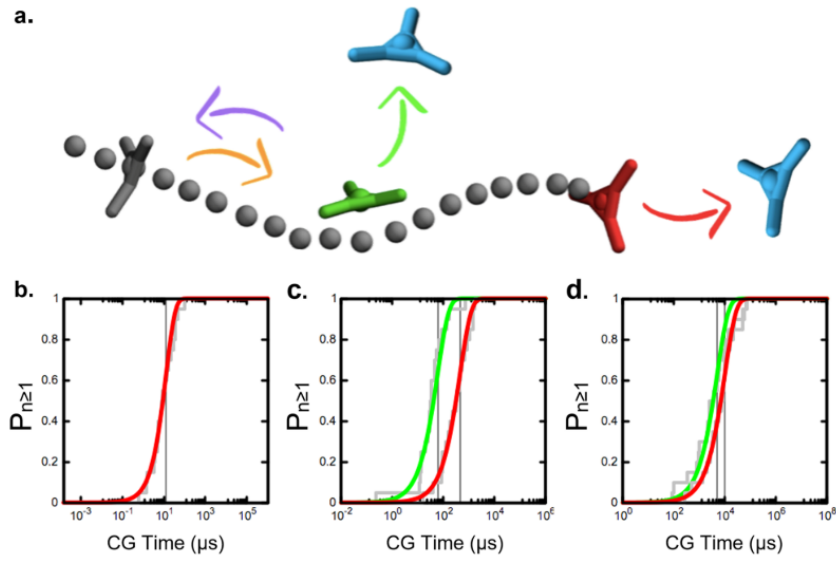


Figure 3.8: Monomer exchange pathways in a supramolecular polymer. a) Scheme of rare events in the exchange pathways: exchange with the solvent from the tip (red arrow) or from a bulk defect (green). The events of creation (orange) and annihilation (purple) of a bulk defect have a statistical nature. (b-d) Cumulative Poisson distribution fits for the events of exchange with the solution, respectively for Fiber 1 (exchange from the tips in red) and for Fibers 2 and 3 (exchange from the tips in red or from a bulk defect in green). Vertical green lines roughly correspond to the characteristic transition time scale (τ) for each distribution.

become almost equally probable. To estimate the values of α from Eq. 4.1 for the two fibers we need the exchange rates K which correspond to the reciprocal (τ^{-1}) of the characteristic times for the transitions obtained from the Poisson distributions.

From Eq. 4.1 we get $\alpha \sim 4$ for fiber 2; this value indicates that the fiber exchanges more from bulk defects, even if the process competes with the exchange from tips. As the length of the fiber increases, an even greater value would be obtained as the probability of swapping from the entire length would be statistically favored. For fiber 3 we obtained the value of $\alpha = 12$ which confirms the exchange from the bulk as the favored pathway.

The combination of biased and unbiased simulations coupled with Eq. 4.1 allowed us to determine which exchange pathway is more likely to occur. It is important to note that Eq. 4.1 depends on the average number of defects in the fiber, so we can only use it when we are able to estimate this quantity with sufficient precision. In general, we have seen that the fibers that have defects such as fiber 2 and fiber 3, provide for the continuous creation and annihilation of the defects inside the stack, as well as the exchange of monomers with the solution; the creation / annihilation process is much faster than the exchange with the outside. Theoretically, for an equilibrated fiber, we can say that the following expression holds:

$$N_{\text{def}}K_{\text{def} \rightarrow \text{bulk}} = N_{\text{bulk}}K_{\text{bulk} \rightarrow \text{def}} \quad (3.2)$$

where

$$N_{\text{bulk}} = N_{\text{tot}} - 2 - N_{\text{def}} \quad (3.3)$$

with N_{tot} the total number of monomers of the model fiber, 2 the number of tips and N_{def} the number of defects in the bulk. $K_{\text{bulk} \rightarrow \text{def}}$ and $K_{\text{def} \rightarrow \text{bulk}}$ are the rates of creation and annihilation of the defects, respectively.

The average number of defects can be obtained from the following equation, by substituting N_{bulk} in Eq. 4.2:

$$\frac{(N_{\text{tot}} - 2)K_{\text{bulk} \rightarrow \text{def}}}{K_{\text{bulk} \rightarrow \text{def}} + K_{\text{def} \rightarrow \text{bulk}}} \quad (3.4)$$

Combining Eq. 4.1 and Eq. 3.4 we obtain the following general equation:

$$\alpha = \frac{(N_{\text{tot}} - 2)K_{\text{bulk} \rightarrow \text{def}}K_{\text{def} \rightarrow \text{sol}}}{2(K_{\text{bulk} \rightarrow \text{def}} + K_{\text{def} \rightarrow \text{bulk}})K_{\text{tip} \rightarrow \text{sol}}} \quad (3.5)$$

This equation depends only on the exchange rates; in our study we obtained these rates by means of the unsupervised SOAP-PAMM analysis (Figure 3.7 c). By inserting the values obtained from this analysis for $K_{\text{bulk} \rightarrow \text{def}}$ and $K_{\text{def} \rightarrow \text{bulk}}$ we obtain from Eq. 3.4 the values $N_{\text{def}} \sim 0.9$ for fiber 2 and $N_{\text{def}} \sim 12.9$ for fiber 3, results consistent with those obtained via clustering analysis. The consistency of the values obtained with the two methods proves that the simulations are at equilibrium.

In other assemblies, if the dynamics of the formation and annihilation of defects is slower, it is possible to use WT-MetaD, as we did to derive the exchange rates of the monomers with the solution from both the tips and the defects. The formulas obtained are general, depend on a limited number of parameters and can be used for supramolecular systems of various types.

3.7 Controlling defects and exchange pathways by customizing monomers.

The balance between directional and non-directional forces is related to the average number of defects in a supramolecular polymer at equilibrium. Through the analysis of fibers 1, 2 and 3 we were able to obtain general indicators to infer the exchange pathways favored by the supramolecular system. If we modulate the balance between directional and non-directional interactions so that one or the other prevails we can influence the abundance of defects. We expect to be able to increase the directional interactions up to the complete disappearance of defects, obtaining a perfectly aligned fiber. Ways to modulate this balance in favor of directional interactions can be, for example, i) increasing the charge of the dipoles and ii) increasing the monomer core area. By making these modifications to the single monomers in the equilibrium configuration of fibers 2 and 3, we should obtain a progressive decrease of the defects until their eventual disappearance, obtaining a final configuration similar to fiber 1.

Figure 3.9 a shows different starting configurations with the customized monomers which, after a 30 μ s CG-MD run, lose all the bulk defects. The dashed line to the right of the red vertical represents the normalized fraction of defects set to 1. The orange and blue curves were obtained by increasing the dipole charges in fiber 3 from $\pm 1.4e$ to $\pm 1.7e$ or $\pm 2e$, respectively. While, the pure and green curves were obtained by increasing the flat area of the monomers of fibers 2 and 3 by a factor of 1.5. Both by increasing the charges of the dipoles and the area of the monomers, the number of defects go to zero in 30 – 40 μ s of CG-MD runs. From the simulations it is clear how all the fibers end up being perfectly aligned and defect free.

An interesting customization to explore is to change the shape of the monomer core, making it square and consequently increasing the number of arms from three to four. As the arms end with solvophilic beads, the effect should be to increase directional interaction through more effective screening against defect formation. Figure 3.9 b shows the two versions created for the four-armed monomers: one with the same area as the original monomer and one with the same area as a porphyrin. Porphyrin self-assembles generally into fibers without defects in organic solvent ([84]). In both cases, the four-side-arm monomers form a perfect fiber (Figure 3.9 c-d).

These results show how the relative difference between directional and non-directional interactions can be modified by acting on molecular factors such as solvophobicity and shape. These are just some examples of how it is possible to

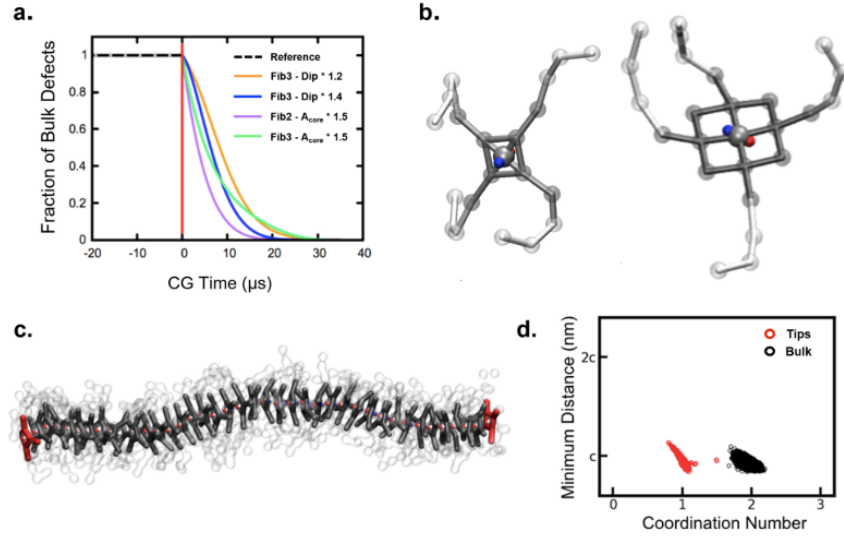


Figure 3.9: Controlling the number of defects by modifying monomer-monomer interactions or monomer shape factors. a) Reducing the number of defects by increasing directional interactions. On the left of the red vertical line: normalized fraction of defects in the fibers at equilibrium. On the right of the red vertical line: change in the normalized number of defects as a function of simulation time obtained by i) increasing the charges in the central dipole (Dip) of Fiber 3 by a factor ~ 1.2 (orange) or ~ 1.4 (blue), or ii) increasing by a factor 1.5 the monomer core area (A_{core}) in both defected fibers (Fiber 2 in purple and Fiber 3 in green). All the curves have been smoothed with Bézier curves. b) Minimalistic CG models for monomer with four-arm cores. Left: a monomer with the same core area as BTA - but having square instead of triangular shape - and with the same amphiphilic arms of Fiber 3. Right: a monomer with the same core area as porphyrin [84]. c) Fiber composed of 40 four-arm monomers as in panel b), right. d) Equilibrium macro-clusters for monomer states of fiber c): fiber tips in red and bulk monomers in black; the fiber shows no defects.

customize a monomer ([84], [87]). The most captivating result is that with simple interventions on the monomer we not only modify the structure of the resulting fiber, but also its dynamic behavior - influencing its number of defects in the bulk and therefore its exchange pathways.

3.8 Conclusions and future outlooks

Multiscale modeling combined with classical and advanced CG molecular dynamics simulations and unsupervised machine learning have provided interesting insight into the intrinsic dynamics of supramolecular polymers.

A minimalistic model representative of the family of three-armed monomers, i.e. BTAs, was created, allowing us to study the structure and behavior of the supramolecular fibers obtained from it.

Thanks to the data collected, a clear and close relationship was found between the polymer structure at equilibrium - abundance of defects - and its exchange pathways.

Defects constitute a key factor of the work as controlling their average number implies controlling the intrinsic dynamics of exchange of the stack.

The quantity of defects derives from the relative difference between directional and non-directional interactions between the building blocks and this difference can be modulated by intervening on the solvophobicity of the monomer.

The transferability of the CG model allows to easily change the monomer structure - for example by inserting more solvophobic beads in the side chains or by changing the surface of the core - to statistically increase / decrease the appearance of defects according to the exchange pathways we want to favor.

The SOAP-PAMM unsupervised analysis also made possible to derive the exchange rates of the monomers and to compare the different dynamics of the three fibers studied. Thanks to these data, with the general formula (Eq. 3.5) we can calculate the dimensionless parameter α which provides a measure of how much the production of defects is favored and consequently the exchange from the bulk compared to that from the tips. This work showed how, in principle, we can uniquely link the dynamic behavior of a supramolecular polymer to the molecular structure of its monomers. The generality of the concepts involved allows us a further step towards the rational design of these fascinating and complex systems.

Chapter 4

Dynamic phase separation in multidomain lipid membranes and the effect of interaction with amphiphilic gold nanoparticles.

4.1 Amphiphilic gold nanoparticles and lipid rafts.

The growing progress in manipulating aggregates on nanometric scale posed the base for investigating various methods to rationally design sophisticated nanomaterials able to mimic/interact with biological systems. Monolayer protected gold nanoparticles (AuNPs) represent a reference tool for diagnostic and therapeutic applications thanks to the biocompatible nature of gold, their simple surface chemistry and their interesting optical and electrical properties. The gold core is commonly covered by a shell of ligands through Au-S covalent bonds. As shown in numerous studies [88], [5], [89], [90], [91], [92], [93], by tuning the nanoparticle size and the monolayer structure, charge and shape, the properties of the AuNPs would change. All these parameters have to be added to the complexity of the biological environments that the AuNP would approach, making the case definitely a tricky one.

The cell membrane is the first physical barrier the nanoparticle needs to overcome when the internalization in a living system occurs. Cell membranes are labyrinthine architectures of lipids and proteins - in a variable composition depending on the cell type - with dynamic and heterogeneous features directly involved in NP attachment process. In order to disentangle, as much as possible, all the relevant physico-chemical factors acting in this interaction, many simplified model membranes with few lipid types and no proteins have been used so far [94]. Different morphological distributions of the ligand shell and its chemical nature and composition affect the internalization pathway of NPs, as shown in both experimental and computational studies.

Verma et al. [5] show that Au NPs with a random hydrophobic octanethiol (OT) and hydrophilic 11- mercapto-1-undecanesulphonate (MUS) ligand shell enters the cell membrane via endocytic pathway, while a striped arrangement of the same ligands favours a spontaneous internalization without damaging the bilayer. Indeed, Gkeka et al. [90] suggested that a homogeneous ligand shell is likely to promote the

translocation within the membrane. Van Lehn and Alexander Katz [88] investigated random, mixed and striped MUS-OT NPs of 2 – 6 nm and they found that ligands fluctuations make very difficult to distinguish various ligand shell patterns in solution. In their subsequent work, by using an implicit bilayer and implicit solvent model they showed that there is no substantial difference in the water-membrane free energy when the interaction with amphiphilic NPs with random or striped ligand patterns occurs [91].

A first description of amphiphilic NPs internalization pathway suggested that the key of the insertion process is the hydrophobic effect. The NP tries to minimize its hydrophobic surface area exposed to water; when it encounters a lipid tail protruded in the solution and the ligand-lipid contact is established, the NP penetration into the membrane core can take place [95]. Simonelli et al. [39] identified a three-step process for the internalization of an anionic Au NP with random and patched ligands morphologies in a POPC lipid bilayer as shown in Figure 4.1; after the adsorption on the membrane surface, a hydrophobic ligand-lipid contact occurs and the NP starts to penetrate into the membrane core until it reaches the final "snorkeled" configuration with ligands uniformly anchored to both membrane leaflets.

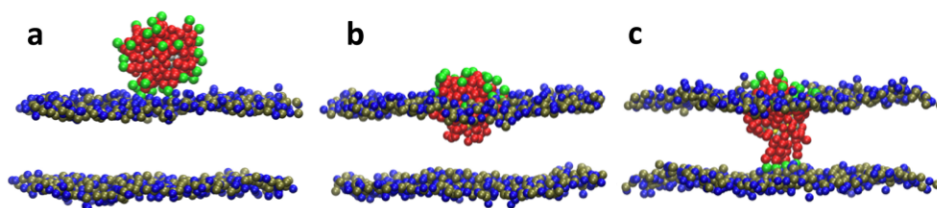


Figure 4.1: Three-steps process of the NP internalization. a) NP adsorption on lipid membrane surface. b) hydrophobic contact and partial NP embedding. c) Snorkeled configuration in which NP has five anchors on the opposite leaflet.

The main interaction mechanism is the same for both random and patched NPs except for the slower kinetics of the latter. Many efforts have been made to comprehend the relationship between NPs and model cell membranes; on the NPs side, the combination of experiments and molecular dynamics simulations have provided relevant findings not only regarding NP ligands arrangement but also about NP surface chemistry and charge [89] [90] [91] [92] [93]. On the contrary, few studies on more realistic lipid membranes are available so far. As mentioned before, biological membranes have very complex features, especially for lipids heterogeneity. Real membranes present distinct relatively ordered subcompartments called lipid rafts or simply rafts.

Rafts are liquid-ordered (Lo) phase nanodomains typically enriched in cholesterol and saturated lipid species like sphingolipids and gangliosides embedded in a liquid disordered phase (Ld), mostly consisting of unsaturated lipids. They form phase-

separated functional platforms on the cell membrane and there are evidence they are involved in many cellular self-organizing functions, in the interaction between membrane-associated components as well as in signaling and trafficking cellular processes [41]. Specialized domains play an important role in regulating the protein traffick; a protein may associate preferentially to the raft domain or alternatively to the Ld phase. Furthermore, being rafts rich in cholesterol and sphingolipids, they may directly act on proteins by modulating bilayer properties. Several works suggest lipid rafts as fundamental key in brain and nervous system functioning [42], but an overall picture of physico-chemical mechanisms regulating their formation and evolution is still lacking.

In the last decades, in order to obtain a more general picture on the molecular evolution of the mechanisms involved, MD simulations have offered crucial insights on functionalized NP kinetics and thermodynamic, model membranes and their interaction. The next step is to investigate, through the combination of simulations and experiments, more realistic membranes, with a lipid composition offering the possibility to observe spontaneous phase separation, such as that expected in raft-containing membranes.

Recent studies show how the embedding of NPs with a diameter of less than 4 nm is favored on the liquid disordered phase and also larger metallic NPs - which are generally just adsorbed on the bilayers - have an effect on the membrane surface i.e. its lipid packing and rigidity [33]. Intrinsic curvature at the boundary of phase-separated lipid bilayers seems to have a key role in driving the interaction with NPs. Cationic NPs seem most attracted to phase-separated bilayers rather than to single-phase ones; the work by Sheavly et al. show that - after being adsorbed on Ld phase - cationic NPs adsorb to the boundary between Ld phase and Lo phase, where the bilayer present an intrinsic negative curvature [96]. Tieleman et al. (2018) showed that the NP adsorption and penetration in phase-separated membranes are generally regulated by ligands density and bending, NPs surface charge as well as by lipid-lipid interaction [32] [97]. In Canepa et al. [98] we show how AuNPs stably interact with model neuronal membranes, leading to the formation of bilayer-embedded, two-dimensional, ordered NP supra-aggregates. Moreover, we find that Au NPs can suppress lipid phase separation, in a concentration-dependent fashion. By combining a simple thermodynamic model with MD simulations we outlined a way to determine if NPs will have the tendency to favor or disrupt the lipid lateral phase separation of the model membrane.

4.2 Experimental amphiphilic NPs and multidomain lipid membranes by our experimental collaborators.

The present computational work has been made in collaboration with the experimental group of Annalisa Relini at University of Genoa. In the next paragraphs, we present

the experimental results that motivated our computational investigation. First, we show the AuNPs and the model neuronal plasma membranes preparation methods; then we discuss the AFM and QCM-D measurements regarding the interaction between the NPs and the bilayer. The experimental amphiphilic Au NPs were synthesized using a one-phase method. A thiol 0.225 mol solution - composed of 1-octanethiol (OT, 0.075 nmol) and 11-mercaptoundecane sulfonate (MUS, 0.15 nmol) - was added to the gold solution and after a short wait the reducing agent was added leading to a reddish-black gold-thiol mixture. The reaction solution was then refrigerated at -21°C , to precipitate the NP overnight.

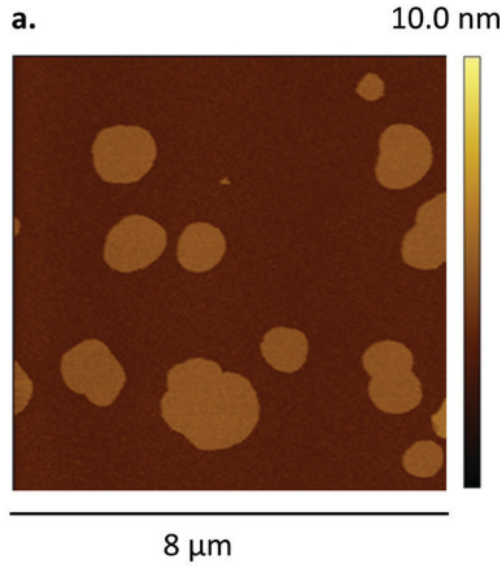


Figure 4.2: AFM image showing the morphology of phase separation; roundish ordered domains of different size are surrounded by the continuous disordered phase.

After repeated washings, the precipitate was dried under vacuum to obtain a shiny dark powder. The NP- core size is 2.7 ± 0.8 (σ) nm and the measure was obtained by transmission electron microscopy (TEM). The MUS:OT ratio after decomposition is 80:20 as found using nuclear magnetic resonance (NMR). The NP- colloidal stability has been ensured by ζ potential measurement of -51 ± 5 mV. The model neuronal plasma membrane (M1) was derived from vesicles composed by 1,2-dioleoyl-sn-glycero-3-phosphocoline (DOPC), sphingomyelin (SM), cholesterol (chol) and ganglioside GM1 in the molar ratio DOPC:SM:chol:GM1 63:31:1:5. This composition guarantees the lateral phase segregation of SM, cholesterol and GM1 [99] [100] [41] [101].

The presence of the negatively charged GM1 leads to a ζ potential of -64 ± 3 mV. The M1 supported lipid bilayer (SLB) is obtained by vesicles rupture on a solid substrate. Figure 4.2 shows the AFM image of the M1 composition with well visible ordered domains of variable sizes surrounded by the disordered phase. Ordered

domains have regular height profiles with a height difference respect to disordered phase is $\Delta z = 2.4 \pm 0.3(\sigma)\text{nm}$; precedent works on rafts containing gangliosides confirm this value [102] [103].

In order to study the interaction of M1 model membrane with NPs, two different sample preparation methods have been developed: the incubation method and the pre-incubation method. QCM-D and AFM were used to investigate the interaction features. In the first setup NPs were incubated on a phase-separated preformed multidomain SLB deposited onto a freshly cleaved mica foil and imaging has been taken at different incubation times. The solid support acts like a motional constraint on the membrane by reducing its fluctuations and lipids diffusion in both in-plane and out-of-plane directions. Furthermore, these constraints suppress the formation of transient defects at bilayer edges like lipid tail protrusions able to stabilize NP-membrane interaction [104]. In our M1 SLB, obtained by the incubation method, there are no defects on a scale of tens of μm (Figure 4.2).

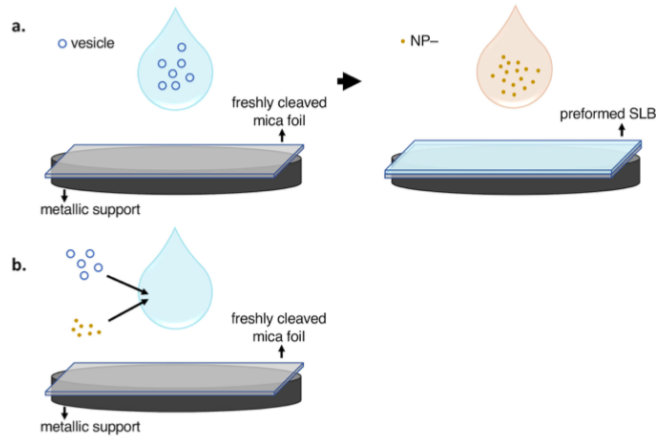


Figure 4.3: Schematic representation of the two sample preparation protocols for AFM imaging and QCM-D. a) the incubation method in which NPs are incubated on a preformed phase-separated SLB and b) the pre-incubation method in which NPs pre-incubated with multidomain lipid vesicles before their deposition on mica substrate.

This scenario will lead to a weak adsorption of NPs on the liquid-disordered phase; the large free energy barrier the NP has to overcome to be partially or totally embedded can be lowered by the presence of defects [105] [8]. The pre-incubation method has been used to ensure a fast kinetics and the formation of defects, though a more stable NP-membrane interaction. In this setup, NPs are pre-incubated in multidomain vesicles before deposition on the substrate. In this way, one has the chance to analyze the behaviour of weakly adsorbed NPs or completely embedded NPs in M1 membrane as well as NPs effects on phase separation. The two setup preparation, shown in the scheme Figure 4.3 is the same for both AFM and QCM-D samples.

4.3 Computational models of the amphiphilic NPs and of the multidomain lipid bilayers.

In *silico* models for the NP and the multidomain lipid membranes are based on the well-known Martini coarse grained (CG) force field [13]. Here, NPs have an Au core with a diameter of 2 nm or 4 nm. The 2 nm core consists of 144 Au atoms and 60 S atoms, with an atomistic representation while the 4 nm core is a hollow sphere made of 346 Au atoms and 240 S atoms equidistant from the center of the sphere.

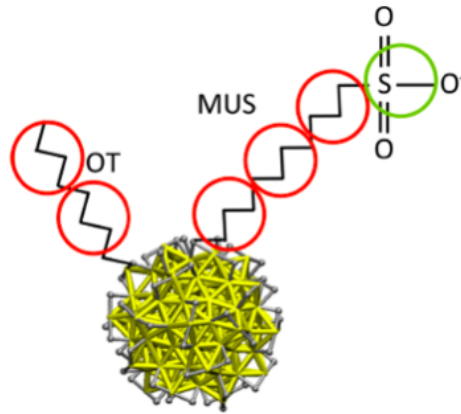


Figure 4.4: Coarse grained model for AuNP. The core is represented by an elastic network between Au-Au and Au-S. OT and MUS ligands are shown through Martini beads schematization.

Au-Au and Au-S interactions are described by an elastic network. In the 4 nm core version, NP surface is functionalized with the same ligand density of the 2 nm NP ($\rho_L = 4.78 \text{ ligands nm}^{-2}$). Ligands are a mixture with a ratio 70:30 of an anionic 11-mercaptoundecanesulfonate (MUS) e hydrophobic octanethiol ligand (OT) - see Figure 4.4. More specifically, the OT ligand is a chain of two hydrophobic Martini beads and The MUS ligand is made of three hydrophobic beads and one negatively charged terminal bead [39] [8]. The *in silico* model neuronal plasma membrane is a mixture of 1,2-dilinoleoyl-sn-glycero-3-phosphocoline (DLiPC), SM, cholesterol and GM1 with the composition DLiPC : SM : chol : GM1 56 : 18 : 17 : 9 Figure 4.5.

M1 membrane consists of 2850 lipids solvated by 35 CG water per lipid and salt at the physiological concentration of 150 mM; in order to obtain a stable multidomain membrane with well separated Lo - Ld phases over long simulation time scale, some adjustments in the composition were needed. The Martini modeled M1 membrane presents two differences compared to the experimental M1. It contains DLiPC - a lipid with a double unsaturation - instead of DOPC - single unsaturation - and a larger amount of cholesterol.

For all the lipids but GM1 standard Martini topology is used [13]. Some previous GM1 topologies result in a strong GM1 clustering attitude. To prevent this behavior,

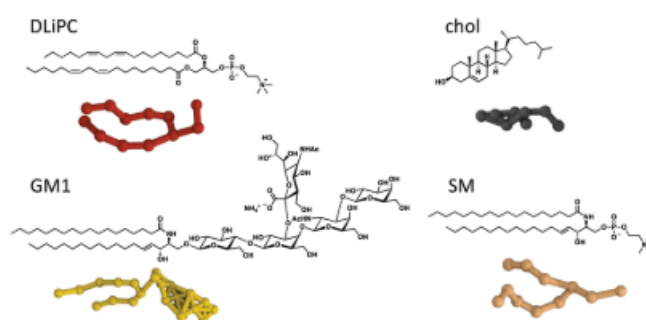


Figure 4.5: Coarse grained models and their chemical formula for the neuronal model membrane lipids.

the GM1 parametrization by Dasgupta et al. turned out to be the best choice according to experimental data; this topology merge the non-bonded parameters by Gu et al. and the bonded parameters by López et al. [106] [107] [16]. The equilibrated phase separated model neuronal membrane is shown below (Figure 4.6).

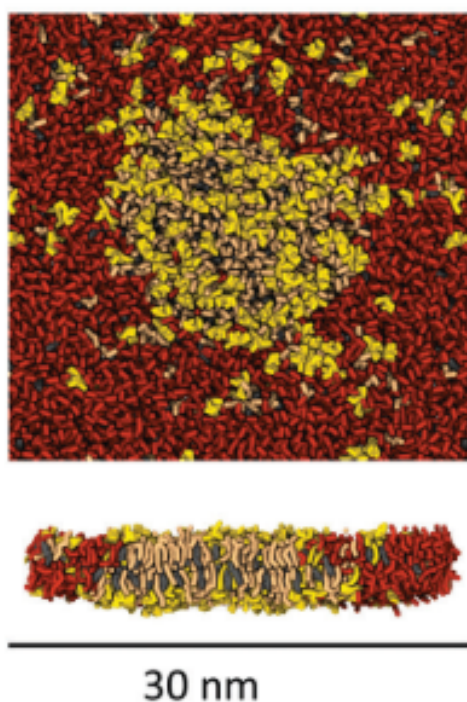


Figure 4.6: Top and side view of the phase separated model membrane at equilibrium (DLiPC in red, SM in light pink, GM1 in yellow, chol in grey)

4.4 Results

4.4.1 Incubation method: Amphiphilic NPs weakly adsorb on the SLB disordered phase

In this section we will describe the effects observed, through QCM-D and AFM, when we deposit NP in various quantities on the preformed SLB for incubation method. The NPs injected into the QCM chamber interact with the SLB on time scales of the order of a few hours. The QCM-D curves in the picture show how the interaction is substantially weak: the NPs are absorbed at the top of the bilayer but they are not embedded, in fact they can be easily removed or spontaneously desorb (Figure 4.7). The decrease of the normalized Δf frequency obtained in the first two hours leads us to calculate the maximum adsorbed mass of 54 ng cm^{-2} . It is possible to return to the starting frequency with the rinsing of the sample, which results completely intact from the data.

The ΔD dissipation variation does not show important perturbations, therefore the stiffness of the membrane is not affected, indicating that the NPs are never integrated into the bilayer - i.e. their penetration process does not reach the embedded state. The same conclusion is reached by AFM imaging of different quantities of NP injected on preformed SLB again with incubation method setup.

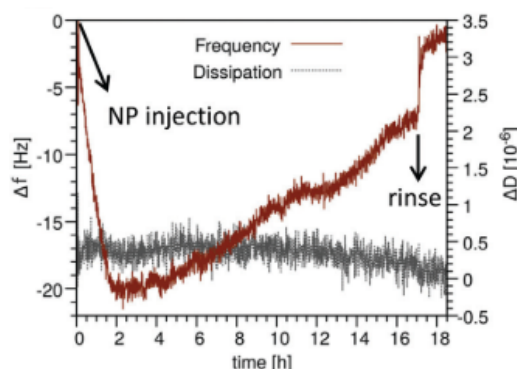


Figure 4.7: Frequency variation Δf and dissipation variation ΔD from QCM-D curves obtained after NPs injection onto a preformed M1 SLB.

Imaging was achieved after several incubation times as shown in the picture. After only forty minutes there are no NPs on the bilayer, while after about 15 h of waiting, numerous clusters of NP adsorbed on the bilayer surface at higher heights than those of the ordered domains are clearly visible, as shown in pictures b e c Figure 4.8. Also in this case the type of interaction between the NPs and the SLB appears weak as it is possible to move the clusters with the tip of the AFM or to remove most of them by rinsing. Cluster formation is a time-dependent phenomenon probably due to the diffusion of NPs on the disordered liquid phase of the model membrane. Both experimental apparatuses confirm a weak interaction of the NPs with the SLB surface; it is also in line with several works with MUS: OT NP and phosphatidylcholine lipid

bilayer.

Van Lehn et al. ([108]), starting from studies on vesicle fusion, have shown how the increase in the stochastic fluctuation of the amphiphilic tails of the neutral bilayer lipids of -PC favors a more stable interaction of the NP with the membrane and the subsequent partial or total embedding. The fluctuation of the tails and therefore the presence of defects in the bilayer, turns out to be one of the factors that lower the free energy barrier by mediating the fusion between the NP and the membrane. Another factor, in addition to the lipid composition of the bilayer, appears to be the presence of rigid supports in the experiments which slow down the dynamics of the lipid. The free energy barrier, instead, is considerably reduced in the case of suspended PC lipid bilayer for which we observe the penetration of the NPs.

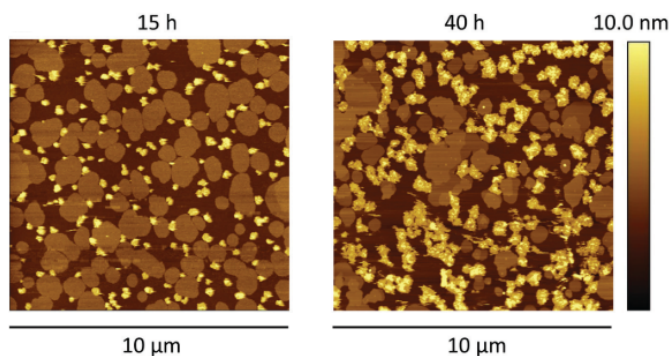


Figure 4.8: Preformed multidomain M1 SLBs after the incubation of NPs. On the left, AFM image acquired 15 h after the addition of NPs ($10 \mu\text{L}$, 0.12 mg mL^{-1}). On the right, AFM image acquired 40 h after the addition of NPs ($40 \mu\text{L}$, 0.12 mg mL^{-1}). Lighter areas correspond to NPs clusters which are higher than the ordered domains.

Van Lehn et al. [108] show how the penetration of NP inside the membrane depends on the overcoming of two barriers: the first is linked to the fluctuation of the lipid tails, while the second is linked to the reconstitution of the leaflet after the internalization of the NP, so that it reaches the minimum energy configuration. It is possible to observe from the picture how our SLB - in addition of being on a mica support - has no defects on a scale of tens of micrometers and this indicates that the first barrier is probably too high to allow the tails to establish contact with the NP (Figure 4.2). Consistent with these studies, in fact, the fusion of the NP with the model membrane is evidently hindered; however, we expect the interaction dynamics to be more consistent in the pre-incubation method, i.e. the second setup used in our experiments described in the next paragraph.

4.4.2 Pre-incubation method: membrane fluctuations influence NP uptake

In this set-up the NPs are injected into the phase separation vesicles for a certain time before they are deposited on the mica substrate for AFM imaging or on the QCM-D sensor, which constitutes a substrate in itself. Once deposited, the vesicle ruptures and forms an SLB with the NPs already inside. In this set up, when we insert the NPs, membrane fluctuations are not subjected to substrate constraints and transient defects such as lipid tails protrusion can form.

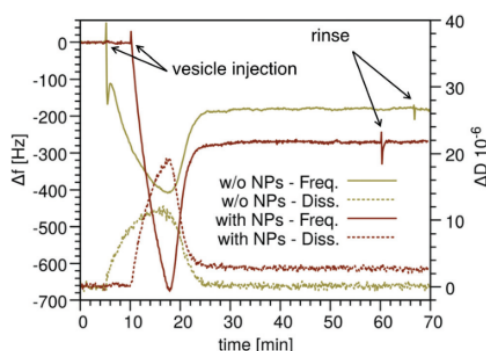


Figure 4.9: M1 SLB formation through deposition of lipid vesicles with and without NPs in the QCM-D chamber. Frequency variation (continuous lines), Δf , and dissipation variation (dotted lines), ΔD , obtained after injection in the QCM-D chamber of M1 lipid vesicles without NPs and M1 lipid vesicles pre-incubated with NPs for 4 h. The adsorbed mass is higher when vesicles are pre-incubated with NP-. When the vesicles attached to the substrate, a positive ΔD was recorded, indicating the presence of a viscoelastic film (i.e., the vesicles layer). During vesicle fusion ΔD decreased gradually to zero.

When the NP meets the tail, contact is established and, depending on the size of the NP, absorption or complete uptake of the NP occurs. Figure 4.9 shows the QCM-D curves during SLB formation after the vesicles have been deposited on the sensor. The measurements were made with and without NP. Compared to the case without NP, the vesicles containing the NPs result in greater mass adhesion throughout the process.

From the final frequency shift it is possible to obtain the value $\Delta m = 267 \pm 18 \text{ ng cm}^{-2}$ which corresponds to the total mass of the absorbed NPs. The NPs number is about 0.017 NP nm^{-2} and the number of lipids per nm^2 is about 1.34 mol nm^{-2} , from which we can derive the number of lipids per NP of about 80 lipids per NP.

Previous studies show that MUS: OT Au NPs interact in a stable manner with vesicles containing a fair percentage of DOPC and SM that show phase separation however, it seems that the same particles interact weakly with SLBs that present always phase separation and containing GM1 ([88], [33], [109], [100]). The last result could be a consequence of the slower dynamics due to the set-up and therefore to

the almost zero stochastic fluctuations of the lipid fluctuations. The pre-incubation method presented in this work seems to be the suitable protocol to study the effects of NPs embedded in phase separation membranes.

4.4.3 Amphiphilic NPs alter the membrane phase separation and NPs disrupt phase separation by altering lipid–lipid interactions

The two samples for AFM imaging were prepared with a pre-incubation of the NPs in the vesicles with two different waiting times, respectively 10 min and 4 hours. As already noted from the analysis of the QCM-D curves, with this setup the NPs interact stably with the membrane.

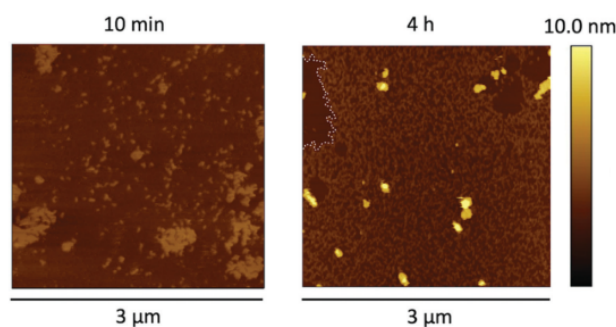


Figure 4.10: Alteration of phase separation after the interaction with NPs. a On the left, AFM image showing the ordered domain fragmentation induced by NPs ($20\ \mu\text{L}$, $0.12\ \text{mg mL}^{-1}$) after 10 min of pre-incubation. On the right, AFM image showing the vanishing of phase separation induced by NPs ($20\ \mu\text{L}$, $0.12\ \text{mg mL}^{-1}$) after 4 h of pre-incubation. The dashed white contour shows one of the sparse disordered patches within the new mixed phase.

The AFM images, however, offer further details on the phenomenon; both NP incubation times lead to an alteration of the phase separation. Figure 4.10 indicates that already after 10 min the NPs make the phase separation less clear, the domains have fragmented edges and an irregular shape compared to the situation without NP (Figure 4.2). After 4 hours of pre-incubation the domains have completely disappeared and the general morphology is totally different, presenting itself as a disordered mix of NP and lipids (FIG.5b). Also visible are some clusters of NPs that we will analyze more closely in the next paragraphs.

Figure 4.11 represents the histogram with the distribution of Deltaz height shifts between the ordered and disordered phase in the case without particles and in the one with the particles pre-incubated for 10 min. Without NP the value is $\Delta z = 2.4 \pm 0.3\ (\sigma)\ \text{nm}$, while with NP the value drops to $\Delta z = 1.5 \pm 0.5\ (\sigma)\ \text{nm}$.

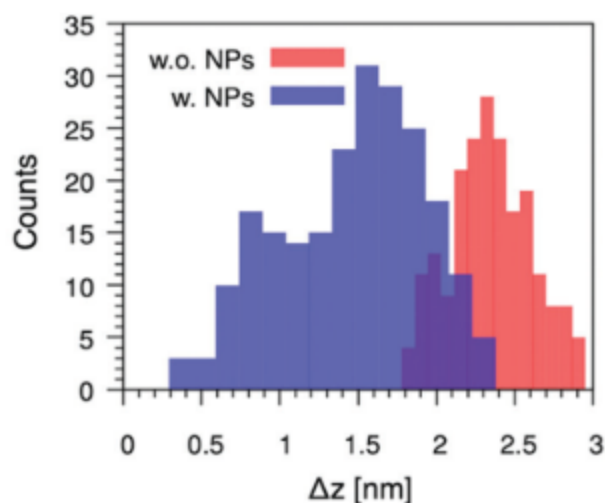


Figure 4.11: Distributions of height difference (Δz) between ordered and disordered domains in M1 SLBs without NPs (red bars) and in M1 SLBs deposited after 10 min of pre-incubation with NPs (blue bars).

4.4.4 MD Simulations

The maximum resolution that can be obtained through AFM imaging is not enough to understand the molecular mechanisms involved in the interaction between the NP and the model neuronal membrane, both in the case of adsorption and in that of internalization of the NP.

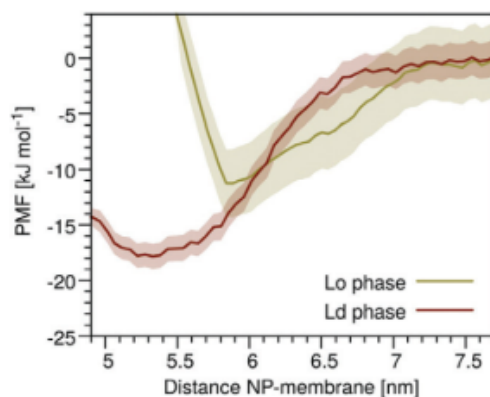


Figure 4.12: Potential of mean force (PMF) profiles for the adsorption of a single NP onto Lo and Ld phase of the model membrane.

As shown in the Figure 4.8, the NP clusters are located in the disordered liquid phase of the multidomain membrane. We used MD simulations to observe "closely" the adsorption of NP by calculating the free energy of absorption of a single NP on the phases Lo and Ld, respectively. In Figure 4.12 the potential mean force profiles are compared; for the adsorption on phase Lo a value of about 18 kJ mol^{-1} (9 kbt) of binding free energy is obtained, while for the absorption on phase Lo the value

obtained is about 11 kJ mol^{-1} (5 kbt).

These results are consistent with the experiments as they show that the interaction on the Ld phase is favored over that on the ordered domains.

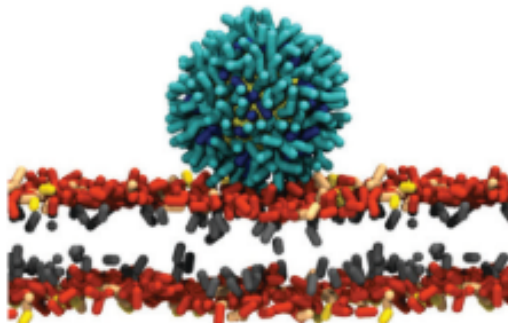


Figure 4.13: Snapshot of a NP adsorbed onto Ld phase.

The thermodynamic advantage is also clear from the behavior of the NP during the simulations; both the 2 nm and the 4 nm NP, initially positioned in the aqueous phase, spontaneously adsorb on the Ld phase. Even in the case of a brief adsorption on Lo phase, the NP migrates to Ld phase after only 100 ns and remains there for the rest of the simulation ($20 \mu\text{s}$).

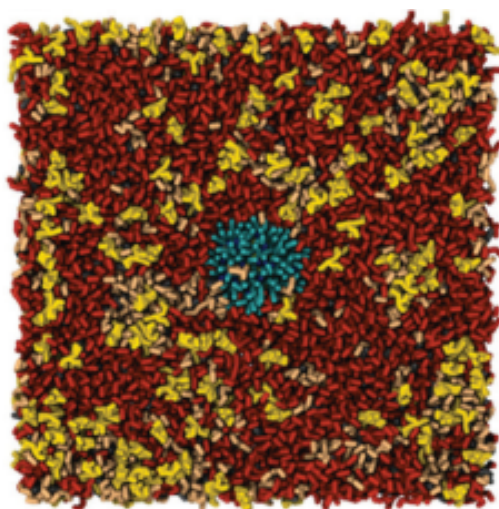


Figure 4.14: Simulation snapshot of the end configuration ($20 \mu\text{s}$), in which the lateral phase separation has disappeared.

The liquid disordered phase allows the NP to interact with the lipid heads of the membrane until it reaches a stable contact Figure 4.13. After the adsorption on the Ld phase, the simulations show that the phase separation is altered and after $20 \mu\text{s}$ there is no longer any trace of domains. Regardless of the degree of NP incorporation and size, the phase separation disappears (Figure 4.14). When the domains fragment, the lipids previously confined to the Lo phase diffuse into the Ld part coming into

contact with the unsaturated DLiPC lipids.

We can quantify this mixing between lipid species by studying the variation of contacts between all lipid pairs. Figure 4.15 shows how after the incorporation of NP, at equilibrium, the intradomain contacts decrease. In this new state of mixed equilibrium, the adsorption of an NP is thermodynamically cheaper than the adsorption on the Ld and Lo phases.

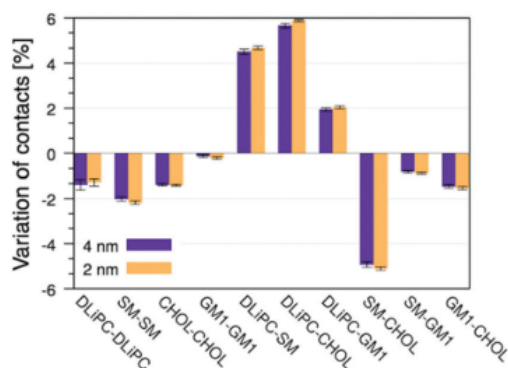


Figure 4.15: Percentual variation of lipid - lipid contacts in the bilayer with NPs (altered phase-separation) and without (stable phase-separation).

4.4.5 MD simulations as a tool to predict the tendency of NPs to alter phase separation in multicomponent lipid bilayers.

The experiments and simulations described in the previous paragraphs show how it is energetically convenient for NP to interact stably with the membrane and how this interaction leads to the fragmentation of the phase separation of the model membrane used. Studying the phenomenon from a thermodynamic point of view requires considerations regarding the Gibbs free energy and its components. In a multidomain membrane in equilibrium in its S phase separated state, the Gibbs free energy is minimal and is described by the following expression:

$$G_S = H_S - TS_S \quad (4.1)$$

if we want to describe a spontaneous transition from the mixed state M to the phase separated state S, then expression Eq. 4.1 can be rewritten in the following

$$\Delta G_{S \rightarrow M} = \Delta H_{S \rightarrow M} - T\Delta S_{S \rightarrow M} > 0 \quad (4.2)$$

The entropy term includes to a greater extent the entropy of mixing and to a lesser extent the configurational entropy of lipids. The enthalpy term instead is the mixing enthalpy. To destabilize the phase separation it is necessary that there is an entropic gain and an enthalpy loss [110]. We are going to omit the $S \rightarrow M$ subscript for the sake of simplicity.

We want to analyze this Gibbs free energy ΔG^{NP} in case the membrane spontaneously interacts with a certain concentration of NPs and to study the sign of ΔG^{NP} to determine the tendency of NPs to alter phase separation. We then rewrite the expression Eq. 4.2 in the case of interaction with NPs:

$$\Delta G^{\text{NP}} = \Delta H^{\text{NP}} - T\Delta S^{\text{NP}} \quad (4.3)$$

The entropic term ΔS^{NP} is substantially equal to the corresponding ΔS term without NP, since c_{NP} is much smaller than the c_{lip} , i.e. the concentration of NP is negligible compared to the lipid one. The enthalpy term therefore decides the sign of ΔG^{NP} ; we then compare ΔH with ΔH^{NP} .

1. $\Delta H^{\text{NP}} < \Delta H$. The decrease in enthalpy with the inclusion of NPs can reduce the stability of the phase separation if the NP concentration is high enough.

we can define c_{tr} as a threshold value for which ΔH^{NP} is equal to the entropic term. Ours is the case of the concentration of NP higher than the threshold value c_{tr} which leads to an enthalpy loss and therefore to a mixing state M favored with respect to the state S. With short simulations it is possible to calculate all the enthalpy terms involved in both the cases.

2. $\Delta H^{\text{NP}} > \Delta H$. Then the state S is favored and therefore the NPs do not alter the phase separation [111], [112]. The values obtained for ΔH and ΔH^{NP} in the case of the model neuronal membrane confirms the condition of the second case: the NPs alter the phase separation, in particular they reduce the enthalpy factor connected to the lipids in the ordered liquid phase, making their confinement less favorable.

Molecular dynamics simulations are a powerful tool for making predictions about the type of effect that functionalized NPs have on multidomain membranes.

4.4.6 Amphiphilic NPs form ordered supramolecular aggregates within the membrane

In the previous paragraphs it has been described how the insertion of NP, through pre-incubation method, in vesicles that spontaneously form phase separation, leads to the fragmentation of the ordered domains until their complete disappearance.

The NPs are therefore in a new mixed phase in which they are free to diffuse. In Figure 4.9, obtained by AFM imaging of vesicles with NP pre-incubated for 4h, supramolecular aggregates are clearly visible. By increasing the pre-incubation time by a few hours, the lattices become even more evident. In Figure 4.16, a digital zoom shows a lattice of NPs embedded in M1 bilayer.

Histogram in Figure 4.17 shows the distribution of height differences between the supramolecular aggregates and the disordered phase; the obtained value $\Delta z = 1.2 \pm 0.3$ shows that the NPs are partially incorporated in the bilayer being smaller than the average diameter of the NPs used for the experiment. Δz data have been obtained from the contour of the rough patches around the NP lattices (e.g., the one outlined by the white dashed line in Figure 4.10). The histogram in Figure 4.18 is instead the

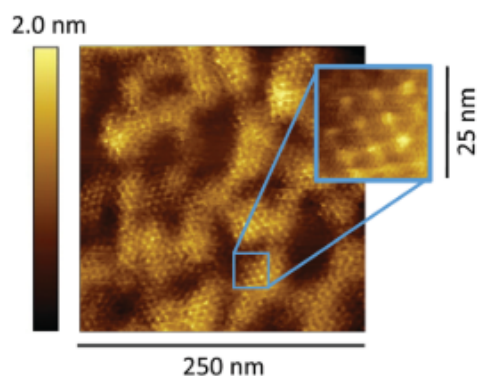


Figure 4.16: Widespread NPs lattice, imaged by AFM, after 4 h of pre-incubation with NPs ($20 \mu\text{L}$, 0.12 mg mL^{-1}). Digital zoom of the area with blue contour showing the lattice order at higher magnification.

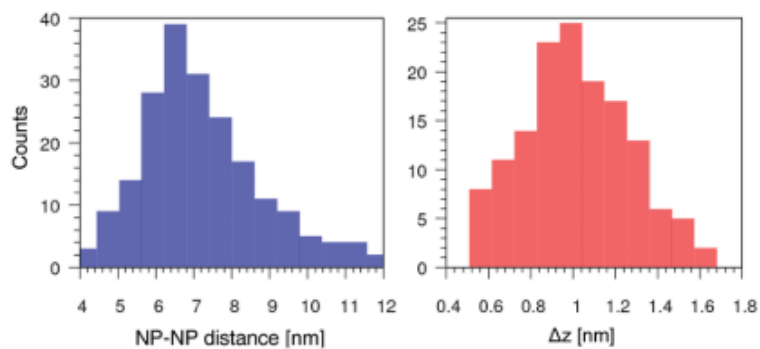


Figure 4.17: On the left: distribution of distances between NPs. On the right: distribution of the height of the supramolecular lattice, Δz , measured with respect to the disordered phase.

distribution of the NP-NP distances, the average distance is 7.5 ± 0.1 nm with a standard deviation of 1.6 nm in line with the NP size distribution.

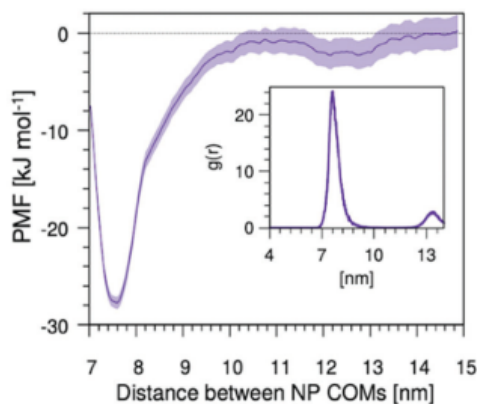


Figure 4.18: Potential of mean force (PMF) for the dimerization of 2 NPs of 4 nm size embedded in the membrane. In the inset, the NP–NP radial distribution function, $g(r)$, obtained during the unbiased run with 7 NP- embedded in the bilayer.

As part of the study of the interaction between amphiphilic gold nanoparticles and lipid membranes, Atukorale et al. highlighted how, in a system like the one described here, there is a certain structure-function relationship, in particular concerning the NP core size [109]. Generally, NP aggregation is observed when the size of the NPs is comparable to or greater than the thickness of the membrane.

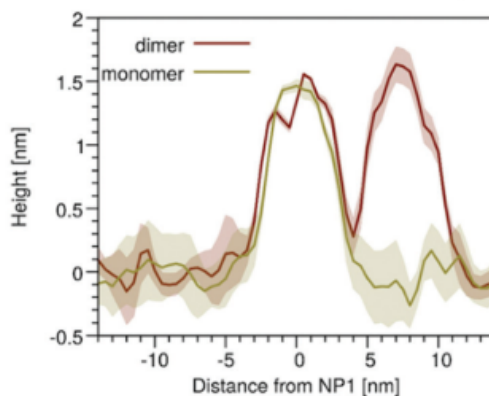


Figure 4.19: Height profile of a single NP and of a dimer of NPs (4 nm), averaged along the simulation time.

In the experiments described, the size range of the NPs is 2 – 4 nm in diameter. Let's try to investigate the aggregation of NP more thoroughly through dimerization simulations, to understand to what extent the size affects the formation process of the observed ordered lattices. Unbiased simulations of $10 \mu\text{s}$ with NP of 2 nm in diameter lead only to transient dimers. On the other hand, using NP of 4 nm in

diameter - comparable in size to the thickness of the membrane - stable dimers are obtained. The dimerization PMF in Figure 4.18 shows how the dimerization for NP of this size is particularly favorable.

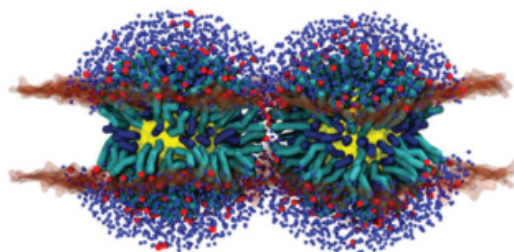


Figure 4.20: Side view of the NPs dimer, in which the formation of a nanopore is shown. Lipid head groups are shown as semi-transparent surface, lipid tails are not shown for clarity, Na⁺ ions are shown in red and water in blue.

To have a comparison with the experimental data, we calculate the height difference of the NP with respect to the bilayer mediated along the simulation time, obtaining an average protrusion value of 1.5 nm compatible with that obtained by AFM imaging of 1.2 nm and consistent with the degree of incorporation of NPs (Figure 4.19).

The snapshot of Figure 4.20 shows the dimerization of two NPs of 4 nm in diameter and the relative formation of a small nanopore which allows the surrounding ligands to reconfigure around the NP [93]. However, the nanopore is the only defect caused by dimerization and remains almost unchanged for the entire duration of the simulation. In addition to dimers, configurations with higher NP numbers were also tested.

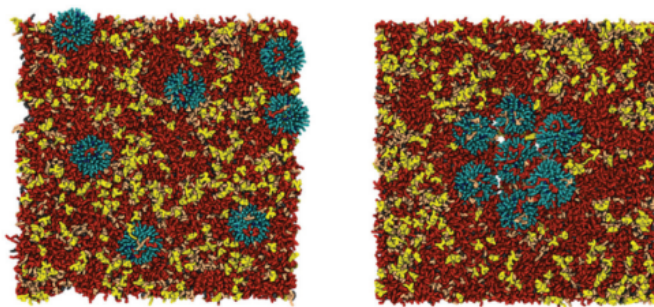


Figure 4.21: Top view of the initial configuration of an MD simulation in which 7 NPs (4 nm), embedded in the bilayer, are let to freely diffuse in the bilayer. e Hexagonal aggregate spontaneously formed by the 7 NPs after 23 μ s of simulations, and stable until the end of the run (30 μ s)

First it was verified that a 7 NP preformed hexagonal lattice was stable with an unbiased run of 10 μ s. Thus, using a configuration with 7 embedded NPs with random positions, free to diffuse in the bilayer and interact with each other, it was observed after 23 μ s of unbiased run that the NPs spontaneously form a hexagonal

aggregate - see Figure 4.21. The lattice distance, shown in the inset of Figure 4.18, is consistent with the experimental data. In both simulations it is possible to observe the spontaneous formation of pores that open and close many times during the simulation (microseconds time scale).

4.5 Conclusions and future outlooks

The work just described shows how negatively charged amphiphilic gold nanoparticles, ranging in size from 2 nm to 4 nm, are capable of altering the phase separation of a model neuronal membrane. The samples used for AFM and QCM-D were prepared by pre-incubating the NPs in multidomain vesicles. The NPs interact spontaneously with the bilayer i.e. via passive plasma membrane penetration and completely suppress the ordered domains already after 4h of incubation. The new mixed phase allows the NP to diffuse in the bilayer to form stable and ordered two-dimensional aggregates (Figure 4.21).

The interpretation of the effect of NP on the ordered phase was given through a simple thermodynamic model: from the comparison of the enthalpy between the lipids in the phase separation state and in the mixed state, with and without NP, it is possible to deduce the energetically more favorable following the interaction with the NP. The model was tested on membranes with different compositions; it can also be used for any embedded NP with a size comparable to the thickness of the bilayer. In this regard, MD simulations prove to be valuable for extrapolating the enthalpy values of the system and therefore constitute an effective tool for making predictions on the effect of inclusion in the membrane of objects of various kinds.

For example, one could proceed in similar ways to understand if certain types of proteins or peptides can influence the lateral phase separation of a lipid bilayer. Or investigate the effect of the same protein in different membrane compositions. Zeno et al. investigated the dissolution of phase separation following the inclusion of large molecular assemblies and nanolipidprotein particles on a small protein [113].

Schieve et al. showed that when several membrane proteins interact simultaneously with the ordered liquid phase, they can destabilize it until the complete rupture of the rafts [114]. The cause of the fragmentation in this case is the increase - due to the collisions between proteins - of the steric pressure until the enthalpy of membrane mixing is overcome. This result supports the thesis according to which the stability of phase separation is directly linked to the balance of free energies. Cases of transmembrane proteins capable of stabilizing lipid rafts ([115], [116]) or even particular peptides capable of inducing or suppressing phase separation in a concentration-dependent fashion have also been documented. This is the case of gramicidin-A, a hydrophobic helical peptide, embedded in DOPC: DSPC: chol vesicles.

By studying the interaction of NPs with the membrane in more detail, we note that there are other similarities with transmembrane proteins. In fact, when a ligand-

protected NP is embedded in the membrane it assumes a cylindrical configuration due to the hydrophobic effect; the flexibility of the ligands allows the NP to modify the hydrophobic surface exposed to the surrounding environment. The NP exposes its lateral surface, i.e. the hydrophobic carbon chains, to the membrane lipid tails, and its hydrophilic sulfonate ligand terminals, i.e. the bases of the cylinder, to the aqueous phase. As in the case of transmembrane proteins, this configuration leads to a hydrophobic mismatch between the NP and the lipid domain, favoring the aggregation of the NPs. The tendency of the membrane to minimize the curvature induced by the presence of NPs also contributes to their aggregation. In fact, the ligands adjust and reorient around the NP stabilizing the aggregate.

The interplay of the physical mechanisms underlying NP membrane aggregation needs further investigation. Nevertheless, the ordered aggregates of functionalized NPs constitute an excellent prospect for bio-sensors or in general for two-dimensional hybrid materials capable of responding to electromagnetic stimuli. This study highlights the need to study in detail the interactions between ligand protected-NP and realistic membranes in order to master the mechanisms involved. Furthermore, it suggests that the way to do this is a synergistic combination of experiments with molecular dynamics, an indispensable tool for extrapolating both qualitative and quantitative data on the systems involved.

Chapter 5

Conclusions

In this thesis, we exploited the synergistic combination of multiscale molecular modeling, molecular dynamics (MD), and enhanced sampling to tackle two complex systems. In the first case study, we investigated the intrinsic dynamic behavior of a Benzene 1,3,5-TricarboxAmide (BTA) supramolecular polymer in water. In the second case study, we inquired about the effect of functionalized amphiphilic gold nanoparticles (Au NPs) on the phase behavior of a multi-component lipid membrane. In the following two paragraphs, I briefly summarize the principal results and perspectives of these two case studies, and in the last paragraph, I draw global conclusions.

5.0.1 Controlling exchange pathways in dynamic supramolecular polymers by controlling their defects.

Here the primary goal of our research was to investigate the relationship between the structure, including the presence of defects, and the dynamic behavior of supramolecular fibers. We developed transferable and minimalistic CG models representative of BTA monomers and, in general, three-armed monomers. Thanks to supervised and unsupervised clustering analysis, we highlighted the presence of defects within the fibers, which is paramount to understanding how the fibers exchange monomers with the solvent. Indeed, defects are monomers that are weakly incorporated into the supramolecular stack and they are more likely to exchange with the solvent than well-incorporated monomers. We used enhanced sampling (infrequent metadynamics) to activate these rare events, i.e., the exchange of monomers with the solvent, recovering their timescale. Combining the information on the abundance of defects and the exchange timescales, we showed how the average number of defects modulates the exchange with the solvent. In other words, we demonstrated how controlling the number of defects implies controlling the preferred exchange pathway of the polymer.

However, how can we control the number of defects in a supramolecular fiber? We showed that the key factor regulating the abundance of defects is the relative difference between directional and non-directional interactions between the monomers, which in turn depends on the solvophobicity of the monomer arms and the geometry of the monomer core. Thanks to the customizability of our CG models, we could modify the monomer structure by placing more solvophobic beads in the side chains

or changing the shape/size of the core, to enhance or reduce the appearance of defects. By doing so, we established an unambiguous link between a supramolecular polymer's dynamic behavior and its monomers' molecular structure. Our results clarified the factors controlling the dynamic behavior of supramolecular polymers, opening up new perspectives for developing other supramolecular systems with controllable properties.

5.0.2 Dynamic phase separation in multidomain lipid membranes and the effect of interaction with amphiphilic gold nanoparticles.

Here the primary goal was to establish how negatively charged amphiphilic gold nanoparticles, ranging in size from 2 nm to 4 nm, can perturb the lipid phase separation of a model neuronal membrane. AFM and QCM-D experiments, combined with MD simulations, demonstrated that these nanoparticles interact spontaneously with the model neuronal bilayers via passive penetration, suppressing the ordered domains (rafts) in a concentration-dependent manner. Based on data extrapolated from our unbiased MD simulations, we developed a simple thermodynamic model highlighting the lipid-lipid enthalpy difference between the phase-separated and the mixed bilayer as the key indicator of the NPs effect on phase separation. As long as the size of the NPs is comparable with the thickness of the bilayer, this thermodynamic model can be used for any inclusion embedded in the membrane.

Furthermore, we also found that NPs diffuse into the bilayer to form stable and ordered two-dimensional lattices. Regarding NP aggregation, we observed that the physical factors determining amphiphilic Au NPs are similar to those occurring in protein-protein aggregation in membranes. In the case of NP-NP aggregation, the flexibility of the ligands, which is greater than the protein conformational flexibility, plays a vital role in stabilizing the NP lattices. Furthermore, our simulations revealed ion-stabilized transient water nanopores forming in the interstices between adjacent NPs in the aggregates. Ordered aggregates of functionalized NPs could lead to the developing of two-dimensional hybrid materials designed to respond rationally to electromagnetic stimuli. Our study opened this perspective while calling for a more detailed investigation of the interactions of ligand-protected NPs with more realistic membranes and the mechanisms underlying NP-NP aggregation.

5.0.3 Concluding remarks

In conclusion, our results show the power of computational techniques in complementing the experimental investigation of complex molecular systems. Through our simulations, we gained a deeper understanding of the structure and dynamics of a class of supramolecular polymers. Additionally, we identified the factors that control the exchange of monomers between the different fibers, which can be used to inform the design of novel supramolecular materials in the future. Our simulations provided insights into the mechanisms underlying the interaction between function-

alized nanoparticles and lipid membranes, extrapolating the factors that influence the stability of the membrane phase separation. The acquired knowledge can be applied in drug delivery systems or to create new hybrid materials containing ordered two-dimensional NP lattices.

In particular, it is worth noting that in both studies, using coarse-grained models with the proper (sub-molecular) resolution was crucial to overcoming the limitations of classic all-atom force fields while maintaining the needed chemical specificity. Overall, the results of these studies have broad implications for materials science and biophysics and demonstrate the potential of computational modeling to inform the design of novel materials and systems.

Bibliography

- [1] Heidi Mohamed Abdel-Mageed, Nermeen Zakaria AbuelEzz, Rasha Ali Radwan, and Saleh Ahmed Mohamed. Nanoparticles in nanomedicine: a comprehensive updated review on current status, challenges and emerging opportunities. *Journal of Microencapsulation*, 38(6):414–436, aug 2021.
- [2] Jinhao Gao, Hongwei Gu, and Bing Xu. Multifunctional magnetic nanoparticles: Design, synthesis, and biomedical applications. *Accounts of Chemical Research*, 42(8):1097–1107, 2009. PMID: 19476332.
- [3] Giulia Rossi and Luca Monticelli. Gold nanoparticles in model biological membranes: A computational perspective. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(10):2380–2389, oct 2016.
- [4] T. Aida, E. W. Meijer, and S. I. Stupp. Functional supramolecular polymers. *Science*, 335(6070):813–817, 2012.
- [5] Ayush Verma, Oktay Uzun, Yuhua Hu, Ying Hu, Hee Sun Han, Nicki Watson, Suelin Chen, Darrell J. Irvine, and Francesco Stellacci. Surface-structure-regulated cell-membrane penetration by monolayer-protected nanoparticles. *Nature Materials*, 7(7):588–595, 2008.
- [6] Davide Bochicchio, Emanuele Panizon, Riccardo Ferrando, Luca Monticelli, and Giulia Rossi. Calculating the free energy of transfer of small solutes into a model lipid membrane: Comparison between metadynamics and umbrella sampling. *The Journal of chemical physics*, 143(14):144108, oct 2015.
- [7] Tommaso Casalini, Vittorio Limongelli, Mélanie Schmutz, Claudia Som, Olivier Jordan, Peter Wick, Gerrit Borchard, and Giuseppe Perale. Molecular Modeling for Nanomaterial–Biology Interactions: Opportunities, Challenges, and Perspectives. *Frontiers in Bioengineering and Biotechnology*, 7(OCT):1–14, oct 2019.
- [8] Sebastian Salassi, Federica Simonelli, Davide Bochicchio, Riccardo Ferrando, and Giulia Rossi. Au Nanoparticles in Lipid Bilayers: A Comparison between Atomistic and Coarse-Grained Models. *Journal of Physical Chemistry C*, 121(20):10927–10935, 2017.
- [9] Helgi I. Ingólfsson, Cesar A. Lopez, Jaakko J. Uusitalo, Djurre H. de Jong, Srinivasa M. Gopal, Xavier Periole, and Siewert J. Marrink. The power of

- coarse graining in biomolecular simulations. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 4(3):225–248, may 2014.
- [10] One Sun Lee, Vince Cho, and George C. Schatz. Modeling the self-assembly of peptide amphiphiles into fibers using coarse-grained molecular dynamics. *Nano Letters*, 12(9):4907–4913, 2012.
 - [11] G. Rossi, P. F.J. Fuchs, J. Barnoud, and L. Monticelli. A coarse-grained MARTINI model of polyethylene glycol and of polyoxyethylene alkyl ether surfactants. *Journal of Physical Chemistry B*, 116(49):14353–14362, 2012.
 - [12] Margaret E. Johnson, Teresa Head-Gordon, and Ard A. Louis. Representability problems for coarse-grained water potentials. *The Journal of Chemical Physics*, 126(14):144509, apr 2007.
 - [13] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. De Vries. The MARTINI force field: Coarse grained model for biomolecular simulations. *Journal of Physical Chemistry B*, 111(27):7812–7824, 2007.
 - [14] Emanuele Panizon, Davide Bochicchio, Luca Monticelli, and Giulia Rossi. MARTINI Coarse-Grained Models of Polyethylene and Polypropylene. *Journal of Physical Chemistry B*, 119(25):8209–8216, 2015.
 - [15] Luca Monticelli, Senthil K. Kandasamy, Xavier Periole, Ronald G. Larson, D. Peter Tieleman, and Siewert Jan Marrink. The MARTINI coarse-grained force field: Extension to proteins. *Journal of Chemical Theory and Computation*, 4(5):819–834, 2008.
 - [16] César A. López, Zofie Sovova, Floris J. van Eerden, Alex H. de Vries, and Siewert J. Marrink. Martini Force Field Parameters for Glycolipids. *Journal of Chemical Theory and Computation*, 9(3):1694–1708, mar 2013.
 - [17] Matthew J. Webber, Eric A. Appel, E. W. Meijer, and Robert Langer. Supramolecular biomaterials. *Nature Materials*, 15(1):13–26, 2015.
 - [18] Jean Marie Lehn. Dynamers: Dynamic molecular and supramolecular polymers. *Progress in Polymer Science (Oxford)*, 30(8-9):814–831, 2005.
 - [19] Anna V. Davis, Robert M. Yeh, and Kenneth N. Raymond. Supramolecular assembly dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, 99(8):4793–4796, 2002.
 - [20] Christianus M. A. Leenders, Lorenzo Albertazzi, Tristan Mes, Marcel M. E. Koenigs, Anja R. A. Palmans, and E. W. Meijer. Supramolecular polymerization in water harnessing both hydrophobic effects and hydrogen bond formation. *Chem. Commun.*, 49:1963–1965, 2013.

- [21] Lorenzo Albertazzi, Nick van der Veen, Matthew B. Baker, Anja R. A. Palmans, and E. W. Meijer. Supramolecular copolymers with stimuli-responsive sequence control. *Chem. Commun.*, 51:16166–16168, 2015.
- [22] Seda Cantekin, Tom F.A. de Greef, and Anja R.A. Palmans. Benzene-1,3,5-tricarboxamide: A versatile ordering moiety for supramolecular chemistry. *Chemical Society Reviews*, 41(18):6125–6137, 2012.
- [23] Matthew B. Baker, Lorenzo Albertazzi, Ilja K. Voets, Christianus M.A. Leenders, Anja R.A. Palmans, Giovanni M. Pavan, and E.W. Meijer. Consequences of chirality on the dynamics of a water-soluble supramolecular polymer. *Nature Communications*, 6(1):6234, dec 2015.
- [24] Matteo Salvalaglio, Thomas Vetter, Federico Giberti, Marco Mazzotti, and Michele Parrinello. Uncovering molecular details of urea crystal growth in the presence of additives. *Journal of the American Chemical Society*, 134(41):17221–17233, 2012. PMID: 22992035.
- [25] Karteek K. Bejagam and Sundaram Balasubramanian. Supramolecular Polymerization: A Coarse Grained Molecular Dynamics Study. *Journal of Physical Chemistry B*, 119(17):5738–5746, 2015.
- [26] Davide Bochicchio and Giovanni M. Pavan. Molecular modelling of supramolecular polymers. *Advances in Physics: X*, 3(1):1436408, jan 2018.
- [27] Lev Dykman and Nikolai Khlebtsov. Gold nanoparticles in biomedical applications: recent advances and perspectives. *Chem. Soc. Rev.*, 41:2256–2282, 2012.
- [28] Ranhua Xiong, Koen Raemdonck, Karen Peynshaert, Ine Lentacker, Ine De Cock, Jo Demeester, Stefaan C. De Smedt, Andre G. Skirtach, and Kevin Braeckmans. Comparison of gold nanoparticle mediated photoporation: Vapor nanobubbles outperform direct heating for delivering macromolecules in live cells. *ACS Nano*, 8(6):6288–6296, 2014. PMID: 24870061.
- [29] Kirsi Salorinne, Tanja Lahtinen, Sami Malola, Jaakko Koivisto, and Hannu Häkkinen. Solvation chemistry of water-soluble thiol-protected gold nanocluster au₁₀₂ from dosy nmr spectroscopy and dft calculations. *Nanoscale*, 6:7823–7826, 2014.
- [30] Reid C. Van Lehn, Prabhani U. Atukorale, Randy P. Carney, Yu-Sang Yang, Francesco Stellacci, Darrell J. Irvine, and Alfredo Alexander-Katz. Effect of particle diameter and surface composition on the spontaneous fusion of monolayer-protected gold nanoparticles with lipid bilayers. *Nano Letters*, 13(9):4060–4067, 2013. PMID: 23915118.

- [31] Reid C Van Lehn, Maria Ricci, Paulo HJ Silva, Patrizia Andreozzi, Javier Reguera, Kislun Voitchovsky, Francesco Stellacci, and Alfredo Alexander-Katz. Lipid tail protrusions mediate the insertion of nanoparticles into model cell membranes. *Nature communications*, 5(1):4482, 2014.
- [32] Xiaojie Chen, D. Peter Tieleman, and Qing Liang. Modulating interactions between ligand-coated nanoparticles and phase-separated lipid bilayers by varying the ligand density and the surface charge. *Nanoscale*, 10(5):2481–2491, 2018.
- [33] Prabhani U. Atukorale, Yu Sang Yang, Ahmet Bekdemir, Randy P. Carney, Paulo J. Silva, Nicki Watson, Francesco Stellacci, and Darrell J. Irvine. Influence of the glycocalyx and plasma membrane composition on amphiphilic gold nanoparticle association with erythrocytes. *Nanoscale*, 7(26):11420–11432, 2015.
- [34] Reid C Van Lehn and Alfredo Alexander-Katz. Fusion of ligand-coated nanoparticles with lipid bilayers: Effect of ligand flexibility. *The Journal of Physical Chemistry A*, 118(31):5848–5856, 2014.
- [35] Pradip Kr. Ghorai and Sharon C. Glotzer. Atomistic simulation study of striped phase separation in mixed-ligand self-assembled monolayer coated nanoparticles. *The Journal of Physical Chemistry C*, 114(45):19182–19187, 2010.
- [36] Yu-Sang Sabrina Yang, Kelly D Moynihan, Ahmet Bekdemir, Tanmay M Dichwalkar, Michelle M Noh, Nicki Watson, Mariane Melo, Jessica Ingram, Heikyung Suh, Hidde Ploegh, et al. Targeting small molecule drugs to t cells with antibody-directed cell-penetrating gold nanoparticles. *Biomaterials science*, 7(1):113–124, 2019.
- [37] Ester Canepa, Davide Bochicchio, Matteo Gasbarri, Davide Odino, Claudio Canale, Riccardo Ferrando, Fabio Canepa, Francesco Stellacci, Giulia Rossi, Silvia Dante, et al. Cholesterol hinders the passive uptake of amphiphilic nanoparticles into fluid lipid membranes. *The Journal of Physical Chemistry Letters*, 12(35):8583–8590, 2021.
- [38] Reid C Van Lehn, Prabhani U Atukorale, Randy P Carney, Yu-Sang Yang, Francesco Stellacci, Darrell J Irvine, and Alfredo Alexander-Katz. Effect of particle diameter and surface composition on the spontaneous fusion of monolayer-protected gold nanoparticles with lipid bilayers. *Nano letters*, 13(9):4060–4067, 2013.
- [39] Federica Simonelli, Davide Bochicchio, Riccardo Ferrando, and Giulia Rossi. Monolayer-Protected Anionic Au Nanoparticles Walk into Lipid Membranes Step by Step. *Journal of Physical Chemistry Letters*, 6(16):3175–3179, aug 2015.

- [40] Jason D. Perlmutter and Jonathan N. Sachs. Interleaflet interaction and asymmetry in phase separated lipid bilayers: Molecular dynamics simulations. *Journal of the American Chemical Society*, 133(17):6563–6577, 2011. PMID: 21473645.
- [41] Erdinc Sezgin, Ilya Levental, Satyajit Mayor, and Christian Eggeling. The mystery of membrane organization: composition, regulation and roles of lipid rafts. *Nature Reviews Molecular Cell Biology*, 18(6):361–374, jun 2017.
- [42] Zeljka Korade and Anne K. Kenworthy. Lipid rafts, cholesterol, and the brain. *Neuropharmacology*, 55(8):1265–1273, 2008.
- [43] Perttu S Niemelä, Samuli Ollila, Marja T Hyvönen, Mikko Karttunen, and Ilpo Vattulainen. Assessing the nature of lipid raft membranes. *PLOS Computational Biology*, 3(2):1–9, 02 2007.
- [44] Lindahl, Abraham, Hess, and van der Spoel. GROMACS Documentation. *GROMACS 2021.3 Manual*, pages 1 – 623, 2021.
- [45] Andrea Torchi, Davide Bochicchio, and Giovanni M. Pavan. How the Dynamics of a Supramolecular Polymer Determines Its Dynamic Adaptivity and Stimuli-Responsiveness: Structure-Dynamics-Property Relationships from Coarse-Grained Simulations. *Journal of Physical Chemistry B*, 122(14):4169–4178, 2018.
- [46] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of chemical physics*, 126(1):014101, jan 2007.
- [47] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, Oct 1980.
- [48] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.
- [49] Johannes Kästner. Umbrella sampling. *WIREs Computational Molecular Science*, 1(6):932–942, 2011.
- [50] Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, nov 2008.
- [51] Mark Tuckerman. *Statistical mechanics: theory and molecular simulation*. Oxford university press, 2010.
- [52] Benoît Roux. The calculation of the potential of mean force using computer simulations. *Computer Physics Communications*, 91(1):275–282, 1995.

- [53] Jochen S. Hub, Bert L. de Groot, and David van der Spoel. g_wham—a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*, 6(12):3713–3720, 2010.
- [54] Alessandro Laio and Michele Parrinello. Escaping free-energy minima. *Proceedings of the National Academy of Sciences*, 99(20):12562–12566, 2002.
- [55] Wei-Tse Hsu, Pascal Merz, Giovanni Bussi, and Michael Shirts. Adding alchemical variables to metadynamics to enhance sampling in free energy calculations. 06 2022.
- [56] Alessandro Laio, Antonio Rodriguez-Forte, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *The Journal of Physical Chemistry B*, 109(14):6714–6721, 2005. PMID: 16851755.
- [57] G. Andrés Cisneros, Mikko Karttunen, Pengyu Ren, and Celeste Sagui. Classical Electrostatics for Biomolecular Simulations. *Chemical Reviews*, 114(1):779–814, jan 2014.
- [58] Cesar A. López, Andrzej J. Rzepiela, Alex H. de Vries, Lubbert Dijkhuizen, Philippe H. Hünenberger, and Siewert J. Marrink. Martini coarse-grained force field: Extension to carbohydrates. *Journal of Chemical Theory and Computation*, 5(12):3195–3210, 2009. PMID: 26602504.
- [59] Giulia Rossi, Luca Monticelli, Sakari R. Puisto, Ilpo Vattulainen, and Tapio Ala-Nissila. Coarse-graining polymers with the martini force-field: polystyrene as a benchmark case. *Soft Matter*, 7:698–708, 2011.
- [60] Siewert J. Marrink and D. Peter Tieleman. Perspective on the martini model. *Chem. Soc. Rev.*, 42:6801–6822, 2013.
- [61] Semen O. Yesylevskyy, Lars V. Schäfer, Durba Sengupta, and Siewert J. Marrink. Polarizable water model for the coarse-grained martini force field. *PLOS Computational Biology*, 6(6):1–17, 06 2010.
- [62] Toshiro Kobori, Satoshi Iwamoto, Kunio Takeyasu, and Toshio Ohtani. Biopolymers Volume 85 / Number 4 295. *Biopolymers*, 85(4):392–406, 2007.
- [63] L. Brunsveld, B. J. B. Folmer, E. W. Meijer, and R. P. Sijbesma. Supramolecular Polymers. *Chemical Reviews*, 101(12):4071–4098, dec 2001.
- [64] Liulin Yang, Xinxin Tan, Zhiqiang Wang, and Xi Zhang. Supramolecular Polymers: Historical Development, Preparation, Characterization, and Functions. *Chemical Reviews*, 115(15):7196–7239, 2015.

- [65] Xuzhou Yan, Feng Wang, Bo Zheng, and Feihe Huang. Stimuli-responsive supramolecular polymeric materials. *Chemical Society Reviews*, 41(18):6042–6065, 2012.
- [66] Lorenzo Albertazzi, Francisco J. Martinez-Veracoechea, Christianus M.A. Leenders, Ilja K. Voets, Daan Frenkel, and E. W. Meijer. Spatiotemporal control and superselectivity in supramolecular polymers using multivalency. *Proceedings of the National Academy of Sciences of the United States of America*, 110(30):12203–12208, 2013.
- [67] Jean W. Freedy, Alejandro Méndez-Ardoy, Supaporn Kwangmettatam, Davide Bochicchio, Benjamin Matt, Marc C.A. Stuart, Jurriaan Huskens, Nathalie Katsonis, Giovanni M. Pavan, and Tibor Kudernac. Molecular photoswitches mediating the strain-driven disassembly of supramolecular tubules. *Proceedings of the National Academy of Sciences of the United States of America*, 114(45):11850–11855, 2017.
- [68] Davide Bochicchio, Supaporn Kwangmettatam, Tibor Kudernac, and Giovanni M. Pavan. How Defects Control the Out-of-Equilibrium Dissipative Evolution of a Supramolecular Tubule. *ACS Nano*, 13(4):4322–4334, 2019.
- [69] Davide Bochicchio, Matteo Salvalaglio, and Giovanni M. Pavan. Into the dynamics of a supramolecular polymer at submolecular resolution. *Nature Communications*, 8(1):1–11, 2017.
- [70] Ricardo M.P. Da Silva, Daan Van Der Zwaag, Lorenzo Albertazzi, Sungsoo S. Lee, E. W. Meijer, and Samuel I. Stupp. Super-resolution microscopy reveals structural diversity in molecular exchange among peptide amphiphile nanofibres. *Nature Communications*, 7(May):1–10, 2016.
- [71] Lorenzo Albertazzi, Daan Van Der Zwaag, Christianus M.A. Leenders, Robert Fitzner, Remco W. Van Der Hofstad, and E. W. Meijer. Probing exchange pathways in one-dimensional aggregates with super-resolution microscopy. *Science*, 344(6183):491–495, 2014.
- [72] Francis Paquin, Jonathan Rivnay, Alberto Salleo, Natalie Stingelin, and Carlos Silva. Multi-phase semicrystalline microstructures drive exciton dissociation in neat plastic semiconductors. *J. Mater. Chem. C*, 3(207890):10715 – 10722, 2015.
- [73] Karteek K. Bejagam, Giacomo Fiorin, Michael L. Klein, and Sundaram Balasubramanian. Supramolecular polymerization of benzene-1,3,5-tricarboxamide: A molecular dynamics simulation study. *Journal of Physical Chemistry B*, 118(19):5218–5228, 2014.

- [74] Matteo Garzoni, Matthew B. Baker, Christianus M.A. Leenders, Ilja K. Voets, Lorenzo Albertazzi, Anja R.A. Palmans, E. W. Meijer, and Giovanni M. Pavan. Effect of H-Bonding on Order Amplification in the Growth of a Supramolecular Polymer in Water. *Journal of the American Chemical Society*, 138(42):13985–13995, 2016.
- [75] Davide Bochicchio and Giovanni M. Pavan. From Cooperative Self-Assembly to Water-Soluble Supramolecular Polymers Using Coarse-Grained Simulations. *ACS Nano*, 11(1):1000–1011, 2017.
- [76] Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical Review Letters*, 111(23):1–5, 2013.
- [77] Pratyush Tiwary, Vittorio Limongelli, Matteo Salvalaglio, and Michele Parrinello. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences of the United States of America*, 112(5):E386–E391, 2015.
- [78] Matteo Salvalaglio, Pratyush Tiwary, Giovanni Maria Maggioni, Marco Mazzotti, and Michele Parrinello. Overcoming time scale and finite size limitations to compute nucleation rates from small scale well tempered metadynamics simulations. *Journal of Chemical Physics*, 145(21), 2016.
- [79] Matteo Salvalaglio, Pratyush Tiwary, and Michele Parrinello. Assessing the reliability of the dynamics reconstructed from metadynamics. *Journal of Chemical Theory and Computation*, 10(4):1420–1425, 2014.
- [80] Mickaël Lelimosin, Vittorio Limongelli, and Mark S.P. Sansom. *Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations*, volume 138. 2016.
- [81] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.*, 100:020603, Jan 2008.
- [82] Anna L. de Marco, Davide Bochicchio, Andrea Gardin, Giovanni Doni, and Giovanni M. Pavan. Controlling Exchange Pathways in Dynamic Supramolecular Polymers by Controlling Defects. *ACS nano*, 15(9):14229–14241, sep 2021.
- [83] Piero Gasparotto, Davide Bochicchio, Michele Ceriotti, and Giovanni M. Pavan. Identifying and tracking defects in dynamic supramolecular polymers. *The Journal of Physical Chemistry B*, 124(3):589–599, 2020. PMID: 31888337.
- [84] Sung Ho Jung, Davide Bochicchio, Giovanni M. Pavan, Masayuki Takeuchi, and Kazunori Sugiyasu. A block supramolecular polymer and its kinetically

- enhanced stability. *Journal of the American Chemical Society*, 140(33):10570–10577, 2018. PMID: 30056699.
- [85] Aritra Sarkar, Ranjan Sasmal, Charly Empereur-Mot, Davide Bochicchio, Srinath V.K. Kompella, Kamna Sharma, Shikha Dhiman, Balasubramanian Sundaram, Sarit S. Agasti, Giovanni M. Pavan, and Subi J. George. Self-Sorted, Random, and Block Supramolecular Copolymers via Sequence Controlled, Multicomponent Self-Assembly. *Journal of the American Chemical Society*, 142(16):7606–7617, 2020.
 - [86] Aritra Sarkar, Tejmani Behera, Ranjan Sasmal, Riccardo Capelli, Charly Empereur-Mot, Jaladhar Mahato, Sarit S. Agasti, Giovanni M. Pavan, Arindam Chowdhury, and Subi J. George. Cooperative Supramolecular Block Copolymerization for the Synthesis of Functional Axial Organic Heterostructures. *Journal of the American Chemical Society*, 142(26):11528–11539, 2020.
 - [87] Nicolas M. Casellas, Sílvia Pujals, Davide Bochicchio, Giovanni M. Pavan, Tomás Torres, Lorenzo Albertazzi, and Miguel García-Iglesias. From isodesmic to highly cooperative: Reverting the supramolecular polymerization mechanism in water by fine monomer design. *Chemical Communications*, 54(33):4112–4115, 2018.
 - [88] Reid C. Van Lehn, Prabhani U. Atukorale, Randy P. Carney, Yu Sang Yang, Francesco Stellacci, Darrell J. Irvine, and Alfredo Alexander-Katz. Effect of particle diameter and surface composition on the spontaneous fusion of monolayer-protected gold nanoparticles with lipid bilayers. *Nano Letters*, 13(9):4060–4067, 2013.
 - [89] Sebastian Salassi, Ester Canepa, Riccardo Ferrando, and Giulia Rossi. Anionic nanoparticle-lipid membrane interactions: The protonation of anionic ligands at the membrane surface reduces membrane disruption. *RSC Advances*, 9(25):13992–13997, 2019.
 - [90] Paraskevi Gkeka, Lev Sarkisov, and Panagiotis Angelikopoulos. Homogeneous hydrophobic-hydrophilic surface patterns enhance permeation of nanoparticles through lipid membranes. *Journal of Physical Chemistry Letters*, 4(11):1907–1912, 2013.
 - [91] Reid C. Van Lehn and Alfredo Alexander-Katz. Free energy change for insertion of charged, monolayer-protected nanoparticles into lipid bilayers. *Soft Matter*, 10(4):648–658, 2014.
 - [92] Stefania Sabella, Randy P. Carney, Virgilio Brunetti, Maria Ada Malvindi, Noura Al-Juffali, Giuseppe Vecchio, Sam M. Janes, Osman M. Bakr, Roberto Cingolani, Francesco Stellacci, and Pier Paolo Pompa. A general mechanism for

- intracellular toxicity of metal-containing nanoparticles. *Nanoscale*, 6(12):7052–7061, 2014.
- [93] Panagiotis Angelikopoulos, Lev Sarkisov, Zoe Cournia, and Paraskevi Gkeka. Self-assembly of anionic, ligand-coated nanoparticles in lipid membranes. *Nanoscale*, 9(3):1040–1048, 2017.
 - [94] Paolo Pengo, Maria Şologan, Lucia Pasquato, Filomena Guida, Sabrina Pacor, Alessandro Tossi, Francesco Stellacci, Domenico Marson, Silvia Boccardo, Sabrina Priel, and Paola Posocco. Gold nanoparticles with patterned surface monolayers for nanomedicine: current perspectives. *European Biophysics Journal*, 46(8):749–771, dec 2017.
 - [95] Reid C. Van Lehn and Alfredo Alexander-Katz. Pathway for insertion of amphiphilic nanoparticles into defect-free lipid bilayers from atomistic molecular dynamics simulations. *Soft Matter*, 11(16):3165–3175, 2015.
 - [96] Jonathan K. Sheavly, Joel A. Pedersen, and Reid C. Van Lehn. Curvature-driven adsorption of cationic nanoparticles to phase boundaries in multicomponent lipid bilayers. *Nanoscale*, 11(6):2767–2778, feb 2019.
 - [97] Xiaoqian Lin, Xubo Lin, and Ning Gu. Optimization of hydrophobic nanoparticles to better target lipid rafts with molecular dynamics simulations. *Nanoscale*, 12(6):4101–4109, 2020.
 - [98] Ester Canepa, Sebastian Salassi, Anna Lucia de Marco, Chiara Lambruschini, Davide Odino, Davide Bochicchio, Fabio Canepa, Claudio Canale, Silvia Dante, Rosaria Brescia, Francesco Stellacci, Giulia Rossi, and Annalisa Relini. Amphiphilic gold nanoparticles perturb phase separation in multidomain lipid membranes. *Nanoscale*, 12(38):19746–19759, 2020.
 - [99] Gerrit Van Meer, Dennis R. Voelker, and Gerald W. Feigenson. Membrane lipids: Where they are and how they behave. *Nature Reviews Molecular Cell Biology*, 9(2):112–124, 2008.
 - [100] Eric S. Melby, Arielle C. Mensch, Samuel E. Lohse, Dehong Hu, Galya Orr, Catherine J. Murphy, Robert J. Hamers, and Joel A. Pedersen. Formation of supported lipid bilayers containing phase-segregated domains and their interaction with gold nanoparticles. *Environmental Science: Nano*, 3(1):45–55, 2016.
 - [101] Nicolas Puff, Chiho Watanabe, Michel Seigneuret, Miglena I. Angelova, and Galya Staneva. Lo/Ld phase coexistence modulation induced by GM1. *Biochimica et Biophysica Acta - Biomembranes*, 1838(8):2105–2114, 2014.

- [102] Reinier Oropesa-Nuñez, Silvia Seghezze, Silvia Dante, Alberto Diaspro, Roberta Cascella, Cristina Cecchi, Massimo Stefani, Fabrizio Chiti, and Claudio Canale. Interaction of toxic and non-toxic HypF-N oligomers with lipid bilayers investigated at high resolution with atomic force microscopy. Technical Report 29.
- [103] Manuela Leri, Francesco Bemporad, Reinier Oropesa-Nuñez, Claudio Canale, Martino Calamai, Daniele Nosi, Matteo Ramazzotti, Sofia Giorgetti, Francesco S. Pavone, Vittorio Bellotti, Massimo Stefani, and Monica Bucciantini. Molecular insights into cell toxicity of a novel familial amyloidogenic variant of β 2-microglobulin. *Journal of Cellular and Molecular Medicine*, 20(8):1443–1456, 2016.
- [104] Ross S. Gunderson and Aurelia R. Honerkamp-Smith. Liquid-liquid phase transition temperatures increase when lipid bilayers are supported on glass. *Biochimica et Biophysica Acta - Biomembranes*, 1860(10):1965–1971, 2018.
- [105] Reid C. Van Lehn and Alfredo Alexander-Katz. Energy landscape for the insertion of amphiphilic nanoparticles into lipid membranes: A computational study. *PLoS ONE*, 14(1):1–19, 2019.
- [106] Raktim Dasgupta, Markus S. Miettinen, Nico Fricke, Reinhard Lipowsky, and Rumiana Dimova. The glycolipid GM1 reshapes asymmetric biomembranes and giant vesicles by curvature generation. *Proceedings of the National Academy of Sciences*, 115(22):5756–5761, may 2018.
- [107] Ruo-Xu Gu, Helgi I. Ingólfsson, Alex H. de Vries, Siewert J. Marrink, and D. Peter Tieleman. Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations. *The Journal of Physical Chemistry B*, 121(15):3262–3275, apr 2017.
- [108] Reid C. Van Lehn, Maria Ricci, Paulo H.J. Silva, Patrizia Andreozzi, Javier Reguera, Kislun Voitchovsky, Francesco Stellacci, and Alfredo Alexander-Katz. Lipid tail protrusions mediate the insertion of nanoparticles into model cell membranes. *Nature Communications*, 5(1):4482, dec 2014.
- [109] Prabhani U. Atukorale, Zekiye P. Guven, Ahmet Bekdemir, Randy P. Carney, Reid C. Van Lehn, Dong Soo Yun, Paulo H. Jacob Silva, Davide Demurtas, Yu Sang Yang, Alfredo Alexander-Katz, Francesco Stellacci, and Darrell J. Irvine. Structure-Property Relationships of Amphiphilic Nanoparticles That Penetrate or Fuse Lipid Membranes. *Bioconjugate Chemistry*, 29(4):1131–1140, 2018.
- [110] Paulo F.F. Almeida. Thermodynamics of lipid interactions in complex bilayers. *Biochimica et Biophysica Acta - Biomembranes*, 1788(1):72–85, 2009.

- [111] Jonathan Barnoud, Giulia Rossi, and Luca Monticelli. Lipid Membranes as Solvents for Carbon Nanoparticles. *Physical Review Letters*, 112(6):068102, feb 2014.
- [112] Jonathan Barnoud, Giulia Rossi, Siewert J. Marrink, and Luca Monticelli. Hydrophobic Compounds Reshape Membrane Domains. *PLoS Computational Biology*, 10(10), 2014.
- [113] Wade F. Zeno, Alice Rystov, Darryl Y. Sasaki, Subhash H. Risbud, and Marjorie L. Longo. Crowding-induced mixing behavior of lipid bilayers: Examination of mixing energy, phase, packing geometry, and reversibility. *Langmuir*, 32(18):4688–4697, 2016.
- [114] Christine S. Scheve, Paul A. Gonzales, Noor Momin, and Jeanne C. Stachowiak. Steric pressure between membrane-bound proteins opposes lipid phase separation. *Journal of the American Chemical Society*, 135(4):1185–1188, 2013.
- [115] Chian Sing Ho, Nawal K. Khadka, Fengyu She, Jianfeng Cai, and Jianjun Pan. Influenza M2 Transmembrane Domain Senses Membrane Heterogeneity and Enhances Membrane Curvature. *Langmuir*, 32(26):6730–6738, 2016.
- [116] Lars V. Schäfer, Djurre H. De Jong, Andrea Holt, Andrzej J. Rzepiela, Alex H. De Vries, Bert Poolman, J. Antoinette Killian, and Siewert J. Marrink. Lipid packing drives the segregation of transmembrane helices into disordered lipid domains in model membranes. *Proceedings of the National Academy of Sciences of the United States of America*, 108(4):1343–1348, 2011.