



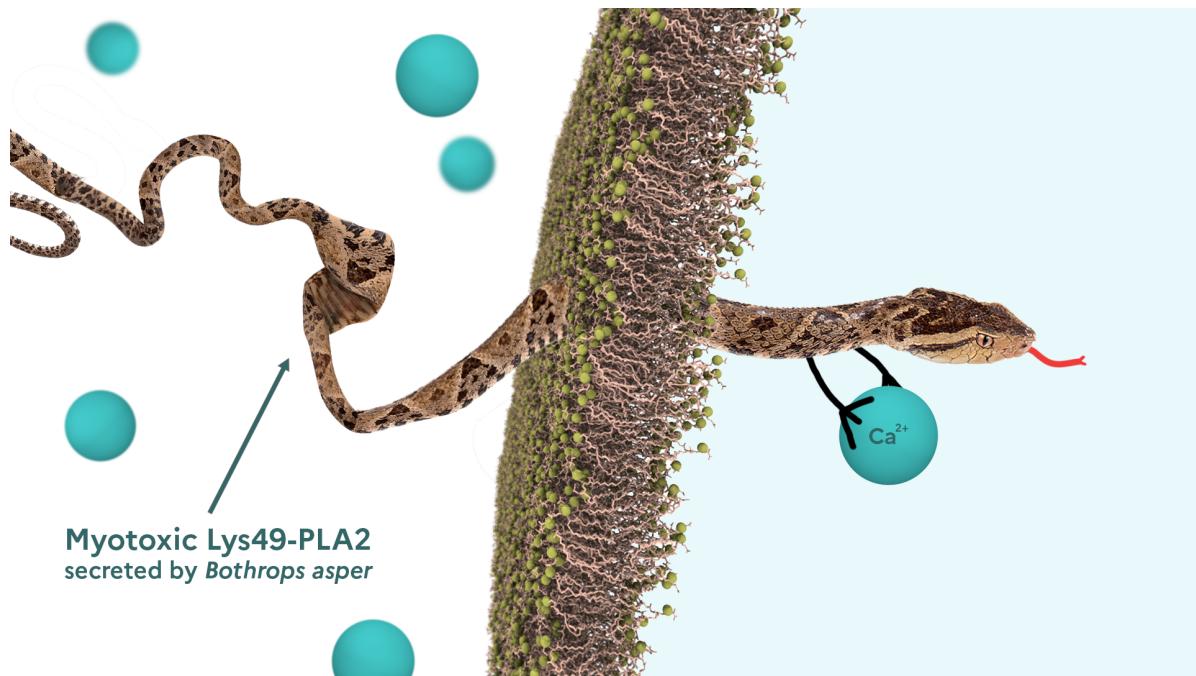
| PSL

U.PORTO

## Second year Internship Report – Engineering Cycle

A study of the mechanism of myotoxicity of snake venom PLA<sub>2</sub>-like toxins with molecular dynamics simulations

Antoine GISSLER



April 1<sup>st</sup> – August 19<sup>th</sup> 2022

Department of Chemistry and Biochemistry – Universidade do Porto  
Tutor: Pedro A. Fernandes, João T. S. Coimbra

ÉCOLE NATIONALE SUPÉRIEURE DE CHIMIE DE PARIS



# Table of contents

<b>Acknowledgments</b>	<b>2</b>
<b>Glossary</b>	<b>2</b>
<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>4</b>
1.1 Introductory note . . . . .	4
1.2 Host organization . . . . .	4
1.3 Subject of the internship . . . . .	5
1.4 Tasks, organization and strategy . . . . .	5
1.5 Mission (stakes and added value) . . . . .	6
<b>2 Theoretical background and methods</b>	<b>7</b>
2.1 About Molecular Dynamics . . . . .	7
2.2 Lys49-PLA <sub>2</sub> -like protein . . . . .	10
2.3 Studied system . . . . .	11
2.4 Umbrella sampling . . . . .	13
2.5 List of experiments . . . . .	14
<b>3 Realized work and obtained results</b>	<b>15</b>
3.1 Equilibration of initial systems . . . . .	15
3.1.1 Control model system . . . . .	15
3.1.2 Experimental model system . . . . .	16
3.2 Translocation of a calcium ion through the model membrane . . . . .	18
3.2.1 Pull method . . . . .	19
3.2.2 Insertion method . . . . .	21
3.3 Translocation of a calcium ion through the peptide-containing membrane . . . . .	22
3.3.1 Obtention of the Potential of Mean Force for POPS . . . . .	22
3.3.2 Study of the asymmetry of the PMF profile . . . . .	24
<b>4 Conclusion and future perspectives</b>	<b>27</b>
4.1 Conclusions on the subject and discussion . . . . .	27
4.2 Personal conclusion . . . . .	27
<b>Bibliography</b>	<b>28</b>
<b>Appendix</b>	<b>30</b>
A. The 20 amino acids . . . . .	30
B. Structure of a force field file . . . . .	31
C. Structure of a parameters file for simulations . . . . .	32
D. Validation of Hydrogen Mass Repartitioning on a Lipid/Protein System . . . . .	33

An online version of this report with supplementary information (*figures, animations, scripts...*) is available on [https://github.com/antoinegsler/Internship\\_Report-2022](https://github.com/antoinegsler/Internship_Report-2022)

# Acknowledgments

I would like to thank first Professor Pedro A. Fernandes for offering me the possibility to do my internship in the laboratory, and João Coimbra for supervising me throughout these last months. Thanks to your help and trust, I was able to get new knowledge in Computational Chemistry (which is for sure going to help me later), as well as realizing experiments, which I hope will help you in your project.

I would like to thank Rui, Carola, Samah, Filipa, Sophie and João again for their kindness, their support and their tips to visit Portugal taste the best Francesinha and Pastel de Nata in Porto.

I would also like to thank the rest of the lab for their kindness.

# Glossary

**ADMET:** Absorption, Distribution, Metabolism, Excretion and Toxicity

**COM:** Center of Mass

**DFT:** Density Functional Theory

**MD:** Molecular Dynamics

**MM:** Molecular Mechanics

**PLA<sub>2</sub>:** Phospholipase A<sub>2</sub>

**PMF:** Potential of Mean Force

**POPS:** 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine

**POPC:** 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine

**QM:** Quantum Mechanics

**US:** Umbrella Sampling

# Abstract

The viperidae *Bothrops asper* is widespread across South America, and secretes an important quantity of toxic venom for its victims, including human. Consequently, it is a necessity to carry out studies about the mechanisms of action of venom leading to the death of cells. In this perspective, I was able to use molecular dynamics to study the mechanism of myotoxicity of a PLA<sub>2</sub>-like toxin, which is a major constituent of the venom. The use of such methods, thanks to progresses that entitled scientists to model biological systems, will enable us to have a better understanding of the myotoxicity.

By following a hypothesis found in literature, the permeation of calcium ions through membranes (similar to the ones found in skeletal cells), contaminated with the active sequence of the protein, is studied. It was possible to observe a decrease of the energetical barrier for the permeation event of about 60 kJ/mol: thus bringing more understanding on the toxicity.

The results found during this internship will also foster studies of similar proteins for the creation of novel antibiotics, or even to find new treatments against cancer, by destroying preferably cancer cells (which exhibit more negative lipids on the outer leaflet of the membrane than others).

# Résumé

La vipère *Bothrops asper* est très répandue, notamment en Amérique du Sud, et dispose d'un venin dangereux pour ses victimes, dont l'Homme fait partie. Il est alors nécessaire d'étudier plus en profondeur les mécanismes suivant l'injection du venin dans le corps jusqu'à la mort des cellules affectées. Dans cette perspective, il m'a été rendu possible de travailler sur le mécanisme de myotoxicité des toxines principales contenues au sein du venin (appelées PLA<sub>2</sub>), celui-ci restant à ce jour peu compris. Cette étude se fera par le biais de simulations de dynamique moléculaire, les progrès effectués dans le domaine permettant désormais d'envisager des modélisations de systèmes biologiques.

En suivant une hypothèse formulée dans la littérature, la perméation des ions calcium à travers une membrane répliquant celle des cellules musculaires, "contaminée" par le peptide actif, est étudiée. La diminution de la barrière énergétique dans le cadre de la perméation des ions calcium constatée, de l'ordre de 60 kJ/mol, donne de nouvelles clés de compréhension concernant la toxicité.

Les résultats de ce stage permettent d'envisager l'application de protéines similaires dans le cas de nouveaux médicaments antibiotiques, ou encore de traitements contre le cancer, en détruisant préférentiellement ces cellules (plus négatives que les autres sur la membrane extérieure).

# 1. Introduction

## 1.1 Introductory note

Molecular dynamics simulations have become of great interest in the last decades. This technique allows researchers to predict chemical properties and processes in given systems without having to do experiments. It is used in the pharmaceutical field, to perform massive screening of interaction between molecules and biomolecular targets (typically proteins), in order to provide potential drug candidates. Thus, it reduces costs and prevents any danger (for instance, highly exothermic reactions) during laboratory phases.

With the improvement of computing solutions, and thus the constant reduction of calculus time, laboratories and companies see the computational field as a viable option, allowing to obtain results close to experiments in quite a short time.

I have chosen to conduct an internship in this domain, as I believe that this field is developing fast and brings revolutionary solutions for old and recent problems. Moreover, I always liked quantum physics and quantum chemistry, which entitle us to understand more about what is happening between atoms at the atomic and sub-atomic scale, something that is still under heavy study. Computational chemistry is still quite broad, and can be applied to a lot of sectors. I decided to focus on Biochemistry as I enjoyed visiting the Computational Biochemistry department during my last internship at Galapagos, a biotechnology group specialized in developing medicines targeting diseases benefiting from low medical coverage.

## 1.2 Host organization

The University of Porto (Universidade do Porto, Figure 1.1) is located in the center of Porto, in Portugal. It was created back in 1911 and is the second biggest university of Portugal. It welcomes every year more than 30 000 students, and 2 500 academic and scientific staff. [1]



Figure 1.1: Logo of the University of Porto

It is possible to follow courses on a broad range of fields, from litterature, to science. Three major campuses are spread throughout the city, in which we can find more than 14 faculties. The University of Porto is also present on an international scale, with the welcoming of more than 6 000 students coming from abroad, and more than 2 000 of their students doing a mobility programme every year.

My internship took place in the Faculty of Science (FCUP), which is located in the campus of Campo Alegre, in the center of the city of Porto. It contains six departments : Biology, Computer Science, Physics and Astronomy, Geosciences, Environment and Spatial Planning, Mathematics and Chemistry and Biochemistry. It also includes 19 research centers, covering the research in the previously cited domains. The research group where I had the occasion of working during the last five months was the Computational Biochemistry Group, belonging to the Department of Chemistry and Biochemistry and part of the REQUIMTE laboratory. This group now counts three university staff members, seven researchers and more than ten MSc and PhD students.

### 1.3 Subject of the internship

One of the research lines currently being studied in the host laboratory are venoms, and more particularly snake venoms. Indeed, this is an important subject, as about 5.4 million snake bites occur each year, causing the death of around 100 000 people, with approximately four times more amputations [2]. This subject does not often benefit from the coverage it needs, and it is today estimated that due to insufficient protective measures, hundreds of thousands of snake bites that could have been avoided are still observed [3]. There is thus an important need for antivenoms; however, some toxins lack scientific studies.

Attention was brought to a particular snake: *Bothrops asper* (see Figure 1.2). It can mostly be found in Central America, but also to some extent in South America. It weights a few kilograms, and has a length of about 1.2 to 1.8 meters [4]. It is considered to be responsible for at least half of all snakebites where it can be found [5], due to its defensive behaviour. When biting, *Bothrops asper* injects in general a quantity of venom of approximately 500 mg [6] ; the protein composition of this venom is described in Table 1.1.



Figure 1.2: Photography of *Bothrops asper* [7]

Protein family	Caribbean	Pacific
PLA <sub>2</sub>	28.8	45.1
Zn <sup>2+</sup> -metalloproteinase	41.0	44.0
Serine proteinase	18.2	4.4
L-amino-acid oxidase	9.2	4.6
Medium-sized disintegrin	2.1	1.4
C-type lectin-like	0.5	0.5
CRISP	0.1	0.1
DC-fragments	< 0.1	< 0.1

Table 1.1: Occurrence of proteins in *Bothrops asper* venom (in percentage) for the Caribbean and the Pacific versants [8]

As it is possible to see in Table 1.1, PLA<sub>2</sub> represent up to a third of the proteins in the venom. It is probably the most studied proteins in snake venoms, as they can be found in most of them and are one of the most toxic in the list. Indeed, they are the ones responsible for the immobilization of the prey, and can even lead to its death. They are composed of a long chain of between 119 and 134 amino acids (*see Appendix A for the list of amino acids*) and show enzymatic activity. Due to the constant evolution of living species during time, these phospholipases have mutated and now constitute a vast family of enzymes [9]. What makes them generally toxic is their ability to hydrolyze glycerophospholipids, causing a rapid necrosis of skeletal muscle fibers. However, some PLA<sub>2</sub> proteins present in *Bothrops asper* venom do not contain this characteristic, due to a mutation (an aspartic acid is replaced by a lysin at the 49<sup>th</sup> position), but are still responsible for toxicity. The goal is now to understand which mechanism could lead to the death of the cell when the toxin binds to the sarcolemma, using modern tools.

Thus, I had the occasion to work on the following subject : "*A study of the mechanism of myotoxicity of snake venom PLA<sub>2</sub>-like toxins with molecular dynamics simulations*".

### 1.4 Tasks, organization and strategy

The work that has been given to me consists in elaborating biological membrane models that are similar (in their constitution and their behaviour) to those in muscle cells in human bodies, using computational tools. Some of those membranes will contain part of the PLA<sub>2</sub>-like toxin: indeed, a small sequence of amino acids in the C-terminal region of the protein is able to induce cell death (*see section 2.2 for more details*). We know that PLA<sub>2</sub>-like protein increase the concentration of calcium in the muscle cells, but the mechanism used by the toxin to facilitate the crossing of calcium through the membrane is not known. The goal will then be to create

simulation of calcium ion translocation through models of biological membranes (*with varying composition*), in the absence and presence of the toxin, to observe membrane behaviour, but also to obtain energy profiles which could help confirming or denying the hypothesis of direct involvement of the toxin in calcium influx.

To this end, the following tasks were done:

1. Understand how the tools and the environment work, by following Amber and GROMACS tutorials
2. Read existing litterature on PLA<sub>2</sub>-like proteins
3. Gather information about previous studies on the translocation of ions through lipid membranes to find experimental data, recurrent parameters, recommended force fields and possible issues
4. Find the typical membrane composition of muscle cells (through experimental studies), see how it is simulated in other MD studies
5. Elaborate a protocol for the translocation of ions through the membrane (for reproducibility purposes): methods and conditions (temperature, molarities...)
6. Study the translocation of calcium ions through a model membrane
7. Study the translocation of calcium ions through a membrane with the peptide

However, this is only a small aspect of the entire project: here is not considered the binding of the protein on the surface of the membrane for instance, as it is another work that is done in parallel by another member. Thus, a continuous dialogue has to be established in order to have uniform protocols and results ; as the objective is to publish our discoveries if it is possible. Periodic meetings with the group allowed this, whilst also being the occasion to receive advices.

## 1.5 Mission (stakes and added value)

A prevalent hypothesis exists concerning the molecular mechanism of the intense myotoxic activity of snake venom PLA<sub>2</sub>-like myotoxins: the toxins preferably binds to lipids bearing negatively charged headgroups. Furthermore, as the toxin binds to the myocyte membrane, it facilitates the passing of calcium ions through it, causing the death of the cell. Thus, the mission will be to prove or deny this hypothesis, with Molecular Dynamics simulations.

To do so, an examination of a model of a myocyte membrane will be created with the help of literature to find the right lipids with adequate proportion. The interaction between the membrane and the toxin will then be studied; and from this step, it will be possible to quantitatively examine the variation of permeability induced by the toxin, as well as the membrane distortion. Umbrella-sampling simulations will be carried out in a model system, containing no peptide, and in the "real" system, in order to obtain the free energy profiles for the translocation of calcium ions.

This work will provide a better understanding of some unknown but common mechanisms happening in an important number of snakebite envenomings, which could provide new treatments against snake venom, and in a larger scale give a new perspective for the treatment of diseases impacting the composition of cell membranes, such as cancers does by showing a higher concentration of phosphatidylserines in the outer leaflet of the membrane [10].

## 2. Theoretical background and methods

### 2.1 About Molecular Dynamics

The goal of Molecular Dynamics (MD) is to obtain time-dependent properties on a molecular system. Indeed, the complete understanding of mechanics at the atomic scale relies on the analysis of the dynamics of the system, as every evolution is characterized by a duration, an energy variation and a conformational change. A temporal analysis can in certain cases bring answers to hypotheses on the way a reaction happens, as well as the reactants used, in a qualitative (*a clear viewing of the transformation going on*) and quantitative (*enables the obtention of energetic diagrams*) way.

Molecular phenomena have very different timescales, and can vary from a few picoseconds to seconds. In the following table are listed some characteristic phenomena (*involving biological systems such as proteins*) and their typical time scales.

Time scale	Amplitude	Description
$10^{-15} - 10^{-12}$ s	0.001 – 0.1 Å	Bond stretching, angle bending...
$10^{-12} - 10^{-9}$ s	0.1 – 10 Å	Loop motion, collective motion
$10^{-9} - 10^{-6}$ s	1 – 100 Å	Folding in small peptides, helix coil transition
$10^{-6} - 10^{-1}$ s	10 – 100 Å	Protein folding

Table 2.1: Some phenomena, amplitude and time scales, based on Carlo Adamo's lectures

According to the model that needs to be studied, the length of the simulation can be totally different: the longest ones being usually performed for biochemistry purposes. And as the time of the simulation is correlated with the computing time, the precision given to the model will be different according to the model (when it is possible): different levels of precision exist depending of what is studied

- **Ab-initio methods:** This is the most precise, as well as the most costly method. No empirical parameters are given, electrons and nuclei are used to solve the Schrödinger equation
- **Density Functional Theory (DFT):** This method concentrates on the electron density of the system to obtain a complete determination of molecular properties
- **Semi-empirical methods:** These methods are still based on quantum mechanics, but some empirical parameters are used to improve calculation performance
- **Molecular Mechanics:** This time, only atoms are considered: electrons and nuclei are assimilated to a mass and a charge, and principles of classical mechanics are applied to them to determine the evolution of the system
- **Quantum Mechanics/Molecular Mechanics (QM/MM) hybrid methods:** When the quantum effects are relevant for the study, but the system is too large to be entirely treated using Quantum Mechanics, one can sample a small part of the system with QM, and still take into account the effects of the surroundings with MM (*to account for polarization, electrostatics and mechanical forces*)

In our case, since we are dealing with large biological systems, and we will be dealing mostly with the dynamics of atoms and molecules, we will be using Molecular Mechanics methods and molecular dynamics simulations.

As told before, the idea is to determine the global dynamics of the system by initializing an initial system, and letting it follow the laws of Newton (*such as  $\vec{F} = m \cdot \vec{a}$* ). As the electronic activity is not taken into account explicitly, it will thus not be possible to see the formation and breaking of chemical bonds. Numerical methods are used for integrating dynamics equations: positions and other values for the system are then discretized with a time step, whose value is of tremendous importance for the stability of the system. Indeed, it is necessary to have it small enough so that every phenomenon is taken into account (*otherwise everything would be unstable and could crash*), but big enough to avoid having a too big computing time. As molecular mechanics and molecular dynamics do not account for electrons explicitly, the fastest motion is the bond vibration involving hydrogen atoms ( $\simeq 10 - 100$  fs), as these atoms are light compared to others in the system. We typically choose a timestep which is 10 times lower than the fastest frequency motion in the system: making it in general lower or equal to 1 fs. Some approximations allow to obtain higher timesteps, such as constraining bonds involving hydrogen atoms (2 fs), or group atoms together into big beads, like it is done using coarse-grained methods (*allowing to reach extended timesteps which can go up to 100 fs*) [11]. Thus, for each increment of the time step, forces on every atom are evaluated with their environment, leading to their acceleration. An integration allows to update the velocity, a supplementary integration leads to new coordinates for every atom. This cycle is repeated until total simulation time is reached. However, this time is also dependant of what is sampled: as shown in Table 2.1, phenomena can have very different orders of time scale. Yet, the probability of an event happening generally requires to be taken into account: making a  $10^{-1}$  s simulation might not be enough to see protein folding, as one cannot be sure it will instantly happen. This is what happens for reactions having to cross an energy-demanding barrier: one way would be to increase simulation time (costing a lot more computing time), the other would be to bias the simulation to make the phenomenon favorable. This is the approach that we will use later to make the calcium ion go through the membrane with Umbrella Sampling.

Dynamics equations rely on basic principles that need to be respected: energy conservation, linear momentum conservation, angular momentum conservation and time reversibility.

At any moment, the system can be described by its energy potential. It corresponds to the energy difference between the current conformation and the lowest energy state, using the principles of Molecular Mechanics. Indeed, total potential results from the sum of the potential of all the possible interactions of atoms between each others :

$$V_{\text{system}} = V_{\text{stretch}} + V_{\text{bend}} + V_{\text{torsion}} + V_{\text{van der Waals}} + V_{\text{Coulomb}} \quad (2.1)$$

Each of these terms define a type of interaction. Stretching, bending and torsion are part of bonded interactions, that is to say that these depend on atoms that are connected by one or more bonds. A representation of these phenomena is provided in Figure 2.1:

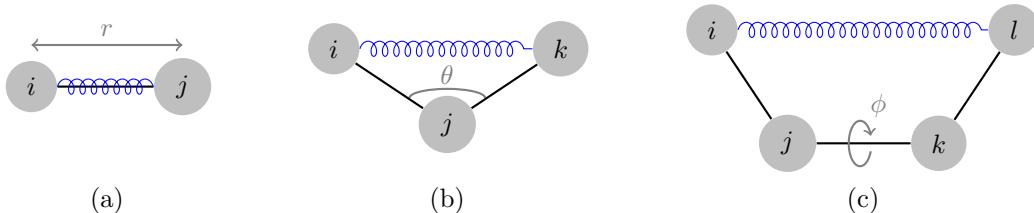


Figure 2.1: The three bonded interactions: stretch (a), bend (b) and torsion (c)

**Stretch:** Corresponds to the energy needed to elongate a bond, starting from its reference distance  $r_0$ . It is well described by the Morse potential, which is often limited to its second order Taylor serie when  $r \simeq r_0$ , for the sake of simplicity (as bond lengths cannot fluctuate a lot).

$$V_{\text{stretch}}(r) = D_{ij} \cdot \left[ 1 - e^{-\alpha_{ij}(r-r_0)} \right]^2 \simeq k_{ij} (r - r_0)^2 \text{ with } k_{ij} = D_{ij} \cdot \alpha_{ij}^2$$

With  $D_{ij}$  the energy needed to dissociate the bond between the atoms  $i$  and  $j$ , and  $\alpha_{ij}$  a fitting parameter, widening or shortening the curve. It appears that bonds can be approximated to harmonic springs when close to their equilibrium.

**Bend:** Corresponds to the energy needed to change the angle  $\theta$  made between three atoms away from its reference value  $\theta_0$ . Again, this potential is harmonic.

$$V_{\text{bend}}(\theta) = \frac{k_{ijk}}{2} (\theta - \theta_0)^2$$

**Torsion:** Corresponds to the energy needed to make a torsion on the  $j-k$  bond from the reference value  $\phi_0$ . As we can rotate the bond with high values of  $\theta$  without risking breaking the bond, the model of a harmonic spring is no longer valid. Since rotating this bond of  $360^\circ$  (or a multiple of  $360^\circ$ ) should have the same energy as not rotating the bond, we have to get a periodic potential:

$$V_{\text{torsion}}(\phi) = \frac{V_n}{2} [1 + \cos(n\phi - \phi_0)]$$

**van der Waals:** This interaction accounts for the attraction and repulsion of instantaneous and induced dipoles caused by the movement of electrons around the nuclei in atoms: for the sake of simplicity, they are modelled with a Lennard-Jones potential :

$$V_{\text{Van der Waals}} = \sum_{i < j} \left[ \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^6} \right]$$

The van der Waals interaction happens for atoms in between molecules, and even for atoms far enough within the same molecule (this condition is considered fulfilled when atoms are separated by at least 3 bonds).

**Coulomb:** It correspond to the electronic interaction of non-bonded atoms (*i.e.* separated by at least 3 bonds or not in the same molecule). Depending on each atom's charge, the interaction can be attractive or repulsive. The potential is derived from the Coulomb force:

$$V_{\text{Coulomb}} = \sum_{i < j} \frac{q_i q_j}{4\pi\epsilon r_{ij}}$$

With  $q$  the charge of the considered atom,  $r_{ij}$  the distance between both atoms  $i$  and  $j$ , and  $\epsilon$  the dielectric constant of the medium.

In the case of biological systems, reaching more than tens of thousands of atoms (*can go up to millions*), the entirety of non-bonded interactions cannot be taken into account. Thus, only atoms in a certain range (called the cut-off range) from the atom will be considered for calculations, in order to reduce calculations without sacrificing too much accuracy.

All the empirical parameters depend on the type of atoms considered (the parameter for a  $sp^2$  carbon might not be the same for  $sp^3$  for instance). Thus, the more complex the system is, the more parameters will exist. The ensemble of these empirical parameters form what is called a **force field**, and is generally specific to a kind of molecules. Indeed, there is no "perfect" force field, and they are only judged on their performance to mimick what reproduce experiments, or high-level QM calculations. For instance, according to the documentation of AMBER, the force field **ff14SB** is parametrized to work with proteins, **TIP3P** is designed to describe water molecules and ions, **Lipid17** will be best for lipids, etc... (*an example of the parameters of a forcefield is given in Appendix B.*)

In order to perform these calculations, several programs have been developped, the most used being GROMACS and Amber. Their job is to take an input (*i.e. the initial structure, the force field with its empirical parameters, and the conditions of the simulations*) and to automatically compute the final structure (and if asked the trajectory, or other simulation informations). During the internship, we have employed the GROMACS software, as it is faster than Amber, and has all the options required.

## 2.2 Lys49-PLA<sub>2</sub>-like protein

The studied protein comes from *Bothrops asper*'s venom. The most common approach to obtain its structure for simulations is to extract it from the venom, multiply it, purify it, and carry out an X-Ray diffraction experiment. Other technics are also available, such as NMR, cryogenic electronic microscopy, etc... From the obtained spectrum and its peaks, it is possible to guess the composition and the structure of the protein. Fortunately, this experimental work has already been carried out, and the final structure file is available on the RSCB Protein Data Bank, under the PDB ID 1Y4L [12, 13]. The tertiary structure of the protein is displayed using the VMD software [14] in Figure 2.2:

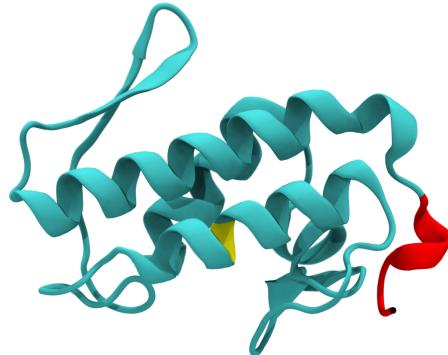


Figure 2.2: Structure of *Bothrops asper* myotoxin II

Two regions are highlighted in this structure: the first one (represented in yellow) is the Lys49 amino acid. As told before, this amino acid is the result of a substitution, and is the origin of the creation of PLA<sub>2</sub>-like toxins. The second zone (represented in red) is our zone of interest: indeed, it is suspected that only 13 amino acids in the whole protein are inserted into the membrane and are responsible for the myotoxicity [15]. They would affect the lipids and make it easier for calcium ions to get through, by lowering the existing energy barrier. We thus extract the sequence from the protein in order to have a lighter system; the rest of the protein having little to no affinity with the membrane, or being bound to the outer layer. We obtain the following sequence: KKYRYYLKPLCKK, which gives the following representation:

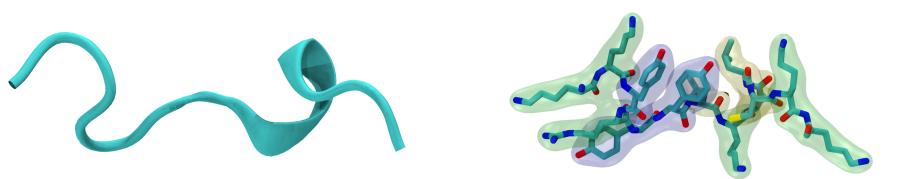


Figure 2.3: Structure of the 13 amino acids sequence: in cartoon (left) and with sticks (right). Positively charged amino acids are highlighted in green, polar amino acids in blue, non-polar amino acids in yellow

The sequence is composed of many positively charged amino acids, such as Lysine (K) or Arginine (R). These amino acids will thus be stabilized in negatively rich regions, such as the ones close to the headgroups of the lipid bilayer. Since these amino acids are mostly found at the extrema of the peptide, it is possible to guess that the peptide will be more stable when it is elongated parallel to the lipids, with the extrema placed around the phosphorus atoms and carboxyl groups. In this way, the polar amino acids also are closer to the water bulk, avoiding energy-expensive water penetration. A .pdb file is created from this sequence and used in the further sections. For all the reasons given earlier, and in order to avoid unnecessary computing cost, the peptide will thus initially be placed in this manner.

In order to study the effect of the amino acid composition, but also the effects of charged amino acids, a mutated peptide will also be introduced, where half of the structure will be left intact and the other half will mirror the first half, yielding this sequence: KKYRYYLYYRYKK.

## 2.3 Studied system

For this project, two categories of systems will be created:

- An **experimental model system**, where we will study the permeation of calcium through a membrane that would have been infected with the peptide
- A **control model system**, where we will study the permeation of calcium through a regular membrane without being infected with the peptide

In each model system, different membrane composition will be tested, to see the influence of the constituents on the behaviour. Indeed, sarcolemma of skeletal muscle cells are the cells where the peptide preferably binds, thus the composition of the membrane we will create has to match experiments. We know that sarcolemma is composed in its majority of phospholipids: phosphatidylcholine (PC, 45.5% of total phospholipid), phosphatidylethanolamine (PE, 22.4%), phosphatidylserine (PS, 17.6%) and sphingomyelin (14.4%). PE and sphingomyelin are not really of interest here, as they bear positive headgroups which does not present a particular affinity with the positive amino acids contained in the peptide, nor the calcium ion. As a matter of consequence, PE and sphingomyelin will be replaced by PC, for the sake of simplicity. In order to complete these phospholipids, palmitoyl-oleoyl (PO) acyl chains were considered, as it is commonly used to mimick cell membranes ; thus POPS and POPC lipids (represented in Figure 2.4) were the two lipids of interest.

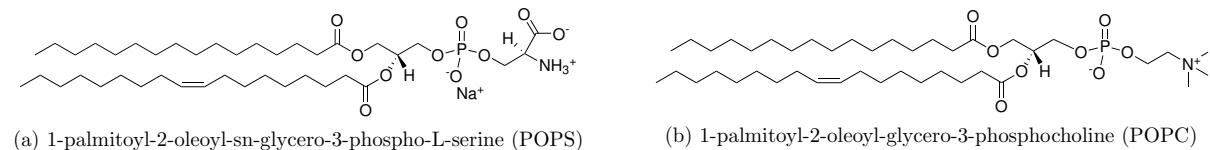


Figure 2.4: Considered lipids in the study

As told previously, the prevalent hypothesis relies on the implication of the protein and negative lipids. As a matter of consequence, several models of membrane have to be modelled to verify such assumption: two extreme cases (*membrane entirely made out of POPS or POPC*), and one more realistic case (*membrane made out of 85% of POPC and 15% of POPS*).

Number of lipids in the simulation will vary accordingly with the model. Indeed, with the insertion of a cell-penetrating peptide, it is possible to expect a deformation of the surface. To avoid having effects too close to edges (and thus affecting periodicity), an important number of lipids is required: here we chose 64 lipids per leaflet (*i.e. 128 in total*). However, as this effect is not expected to happen in control models, the size of the system can be reduced to favor faster calculations: 36 lipids per leaflet (*72 in total*) were considered.

Other than the lipid composition of the membrane, same conditions were applied for both models, in order to be able to compare results and for reproducibility purposes. They can be found in Table 2.2.

Parameter	Value
Temperature	310.15 K
Water molecules per lipid	60
[NaCl]	0.15 M

Table 2.2: Parameters considered for every simulation

All structures used for simulations are created using Charmm-GUI [16–18], an easy and online tool for generating system inputs, and in particular membrane systems [19–22]. It assembles the lipids into a bilayer, solvates with water and chosen ions, and can insert proteins in the membrane. The protein .pdb file is imported; while lipids and other residues are given by Charmm-GUI. Previously cited system options were The resulting system was then used for molecular dynamics simulations in GROMACS 2018: the TIP3P water model was employed ; and the CHARMM36m force field parameters [23] were used for all the atoms, as it provides good results for proteins and lipids, and as it has already proven its accuracy in previous reviews. Most of the molecular dynamics parameters were automatically given by Charmm-Gui and left intact: Verlet cut-off scheme, non-bonded interaction cut-off radius of 1.2 nm, Particle Mesh Ewald (PME) scheme for long-range electrostatics, LINCS constraint algorithm for bonds involving a hydrogen atom. Hydrogen Mass Repartitioning [24] was implemented in the system to allow faster calculations (*a validation of the method is provided in Appendix D.*), thus entitling for a time step of 4 fs.

In order to equilibrate each of the created system, the following protocol was used:

- **Energy minimization:** using the *steepest descent* algorithm, was performed for 5000 steps
- **NVT:** Gradual annealing from 0K to target temperature (310.15 K), whilst forcing the system to keep the same number of particles, volume and controlling the temperature (NVT), using the *Berendsen* coupling thermostat
- **NPT:** Gradual increase of pressure to the correct value (1 bar) whilst forcing the system to keep the same number of particles and temperature, and controlling the pressure (NPT), thanks to *Berendsen* coupling barostat and thermostat
- **Production:** with fixed temperature and pressure (*NPT ensemble*) using the *velocity-rescale* thermostat and the semi-isotropic *Parinello-Rahmann* barostat

During the first steps of the process (*energy minimization, thermal equilibration and pressure equilibration*), position restraints are also applied to phosphorus atoms, contained in lipids. These restraints are useful in the first steps of equilibration for the system not to collapse. Once the system is stable enough, and at the correct temperature and pressure, restraints can be gradually lowered, and removed. The thermostat and barostat used (*Berendsen* for both) is also used for equilibration purposes only, and is replaced for all the following simulations.

All these instructions are given to GROMACS in a .mdp file (*an example of the structure is available in Appendix C.*), as well as with the initial structure.

It is possible to follow the evolution of equilibration by extracting some relevant parameters and comparing them to litterature and experimental data: pressure, density, thickness of the membrane, area per lipid but also deuterium order parameter. Area per lipid is the average area of the  $x - y$  plane divided by the number of lipids. The deuterium order parameter  $S_{CD}$  of a phospholipid acyl chain is defined as an indicator of the orientation of the C-H bond vector with respect to the lipid normal axis. Let  $\theta$  the angle between the C-H bond and the lipid normal axis, we obtain Equation 2.2 :

$$S_{CD} = \frac{1}{2} \langle 3 \cos^2(\theta) - 1 \rangle \quad (2.2)$$

With  $\langle \cdot \rangle$  being the average for every carbon atom of the lipid chains over time.

Once the system is prepared, we can insert a calcium ion in the solute (by replacing two sodium ions for the global charge to remain equal). To do so, the ion is added at a specified position by the user, and the interactions between the residue and the system are gradually switched on during a new set of equilibration.

## 2.4 Umbrella sampling

The goal is now to obtain the free energy profile of calcium permeation through the membrane. However, due to the different properties of the ion and the membrane, the energy barrier to cross is non negligible (in the orders of multiple  $k_B T$ ): the probability of seeing this event in a nanosecond scale is reduced (according to the Boltzmann probability distribution). Simulating in these conditions is thus impossible for this system.

In order to sample qualitatively each state of the transition (even those with high energy) in a short time range, umbrella sampling can be used. Introduced by Torrie and Valleau in 1977 [25], its principle rely on the use of biased potentials to sample a specific degree of freedom of the system. By adding restraints, the studied degree of freedom can be forced to stay in values that would have been sampled very quickly otherwise. The biased potential has to be low in the zone that will get sampled, and should rapidly increase as soon as the degree of freedom attempts to diverge: a harmonic potential is also used for its simplicity (*see the equations below*).

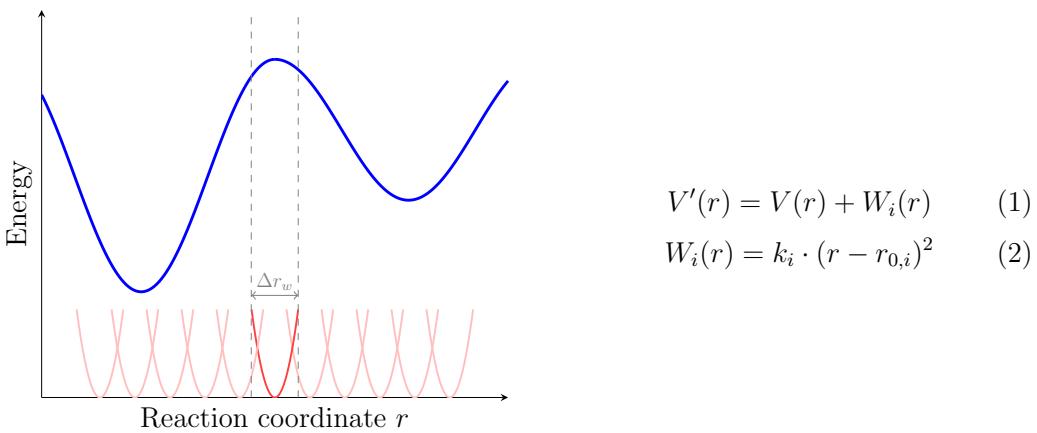


Figure 2.5: The main potentials in Umbrella Sampling: in blue is represented the free energy  $V(r)$  of the reaction along a reaction coordinate and in red the biasing potentials  $W_i(r)$

This process is repeated periodically along the reaction coordinate to sample every state of the system (usually every 0.1 nm): in the end, from a single system are created multiple sub-systems where the sampled residue is restrained in a different value in the degree of freedom ; these sub-systems are called **windows**. A simulation can thus be run in every window to obtain the behavior of the degree of freedom and its relative energy ; giving in the end the free energy of the system along the reaction coordinate after re-assembling every part of the puzzle, and recovering the true probability of the degree of freedom by subtracting the biasing effect of the added harmonic potential.

The value of the harmonic constant  $k_i$  is not left to chance: it has to be strong enough to force the sampled degree of freedom to stay in the chosen value, but also loose enough to allow window overlapping (otherwise more windows would be necessary, which would imply a higher computing cost).

In order to perform umbrella sampling, it is necessary to have already starting configurations with the sampled residue at the right place. In order to do so, two options are available:

- **Pull method:** Extracting the different configurations from a steered MD simulation: the residue can be pulled (by a virtual spring) along the reaction coordinate
- **Insertion method:** Generating as many configurations as windows by inserting the residue at each spot

For our study, it was decided to generate windows from both options in order to provide a comparative review, which was not yet provided in the litterature:

- For the pull method, the calcium ion was inserted as described previously at  $z = 3.8$  nm, without any restraint on the  $x - y$  plane, but on the  $z$  axis (need to start precisely at 3.8 nm). The ion was then pulled in the  $-3.8 \leq z \leq 3.8$  nm zone, with a pull rate of 0.1 nm/ns and a spring harmonic constant of 2000 kJ.mol $^{-1}$ .nm $^{-2}$ . Every 0.1 nm, the configuration was extracted and saved for umbrella sampling.
- For the insertion method, the calcium ion was inserted every 0.1 nm from  $z = 3.8$  nm to  $z = -3.8$  nm with restraints on the  $z$  axis but also on the  $x - y$  plane (to guide the ion). The input files for each window were prepared using the equilibrated system and a Python script

For each one of the methods, 77 windows were generated. A harmonic potential of  $k = 1500$  kJ.mol $^{-1}$ .nm $^{-2}$  was used for the purpose of umbrella sampling, along the  $z$  axis. Each window was run for 200 ns: 20 ns of equilibration under the NPT ensemble, followed by 180 ns of production for data gathering purposes ; totalling 15.4  $\mu$ s of total simulation time per system. In the end, the force required to keep the ion at its position (along the  $z$  axis) inside each independent window, all with different values of  $z$ , is used to get the free energy  $\Delta G$ , thanks to the Weighted Histogram Analysis Method (WHAM) [26] implementation in GROMACS [27]. Uncertainties were estimated using the bootstrap technique already present in the `gmx wham` command. To this effect, a number of 100 bootstraps were used.

## 2.5 List of experiments

As several parameters might have a non-negligible influence on the outcome of the simulation, it is required to carry out experiments where all of them are changed separately. In total, eight simulations were created, exploring every possibility we have.

Experiment	1	2	3	4
Peptide	✗	✗	✗	✓
Mutated peptide	—	—	—	✗
Method used	Pulling	Insertion	Insertion	Pulling
Lipids used	POPS	POPS	POPC	POPS

Experiment	5	6	7	8
Peptide	✓	✓	✓	✓
Mutated peptide	✗	✗	✓	✗
Method used	Insertion	Insertion	Insertion	Insertion
Lipids used	POPS	POPC	POPS	PC/PS

Table 2.3: Summary of all experiments carried out

# 3. Realized work and obtained results

## 3.1 Equilibration of initial systems

### 3.1.1 Control model system

The systems are generated using Charmm-GUI, each one resulting in around 22500 atoms. They were then imported into GROMACS and equilibrated following the protocol given previously in section 2.3. After the final step of equilibration, every system had reached convergence (*values of parameters were stable*), resulting in the final form of the bilayer, shown in Figure 3.1:

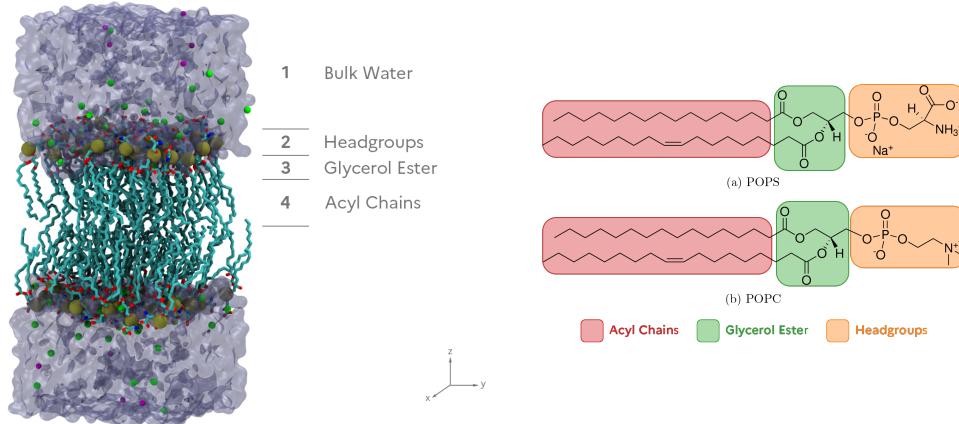


Figure 3.1: Overview of the system and its characteristic regions (*left*) and the regions represented on the chemical structure of POPS and POPC lipids (*right*); Water molecules are represented as a grey surface, phospholipids as sticks and colored by atom type, with P shown as a brown sphere, and  $\text{Na}^+$  and  $\text{Cl}^-$  ions respectively as green and purple spheres

We can observe that the system is symmetric, which is coherent with the input parameters. It is possible to see, along the  $z$  axis, a succession of layers of elements ; from this point it is possible to split our system in 4 different regions, having from top to bottom: the bulk water, the headgroups (containing the Phosphorus atoms and in the case of POPS the carboxyl group), the glycerol ester and the acyl chains (also called tails). All these groups have different densities, as visible on Figure 3.2:

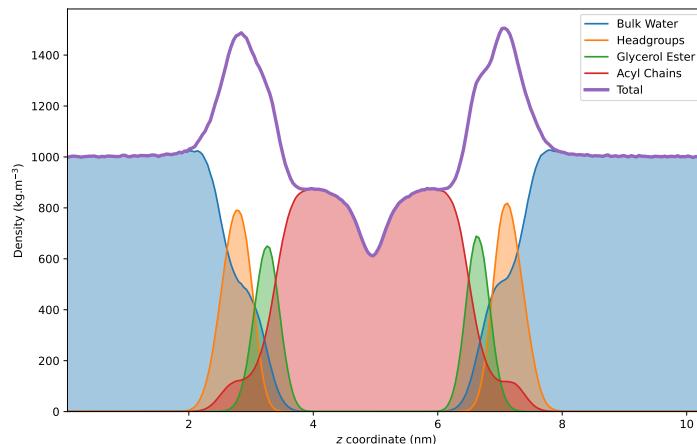


Figure 3.2: Density profile of the POPS system

Although it is possible to observe similar density profiles for all hydrated bilayer systems, some differences do exist when the lipid composition is modified. The following table lists some of the parameters that are typically analyzed in the simulations of hydrated bilayers, and that can be compared with experimental values in the literature:

Lipids	POPS	POPC
Box size $x$ ( $\text{\AA}$ )	$45.68 \pm 0.11$	$47.96 \pm 0.11$
Box size $y$ ( $\text{\AA}$ )	$45.68 \pm 0.11$	$47.96 \pm 0.11$
Box size $z$ ( $\text{\AA}$ )	$103.05 \pm 0.48$	$94.99 \pm 0.43$
Bilayer thickness ( $\text{\AA}$ )	41.80	38.74
Area per lipid ( $\text{\AA}^2$ )	$57.96 \pm 0.20$	$63.89 \pm 0.21$
↪ Data from litterature	$55 \text{ \AA}^2$ at 300K [28]	$65 \pm 1.3 \text{ \AA}^2$ [29]

Table 3.1: Simulation size and area per lipid for each lipid considered

Values obtained with our simulations are similar to the ones found in the litterature, which supports the well-convergence of our system. The difference between box sizes might be explained with the difference between headgroups (*see Figure 2.4*): as phosphatidyl serine headgroups bear at the same time positive and negative charges, it entitles a tighter connexion between lipids, thus reducing the area per lipid. As a matter of consequence, the tails of the PS lipids will have to take less place on the  $x - y$  plane, making them more straight. This will increase the bilayer thickness (*and the total height of the box, as shown in the table*), as well as the order parameter (*as more constraint on the tails implies more order*), as we can see in Figure 3.3.

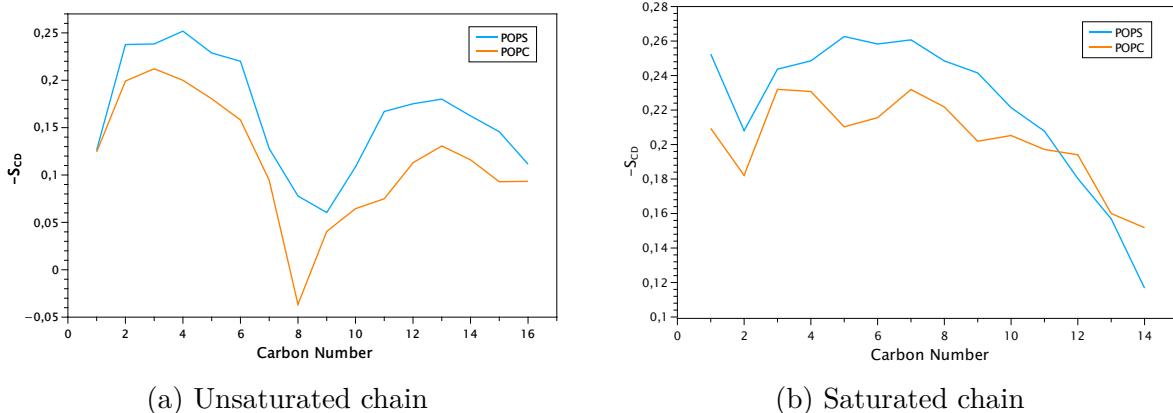


Figure 3.3: Order parameter for the two lipid tails, for POPC and POPS

These results also fit the ones found in the litterature, both for POPS and POPC [30], which supports our force field choice (Charmm36m).

### 3.1.2 Experimental model system

For these simulations, the sequence of interest from the Lys49-PLA<sub>2</sub> protein was extracted and considered as a transmembrane protein. Indeed, it has been found that the peptide 115-129 has a similar activity compared to the whole protein [31]. The peptide main axis was aligned with the lipids in Charmm-GUI, as it was demonstrated to be its most stable position for optimal interaction and as previously highlighted in literature. Systems were generated and resulted for each lipid into around 40000 atoms. 500 ns of production were used during equilibration to make sure the system had converged, as the insertion of a peptide affects many properties. The actual shape of the model is quite similar to the control one, that we can see in Figure 3.1:

the real difference being localized around the peptide. A close-up look inside the membrane in Figure 3.4 reveals a local deformation of the membrane:

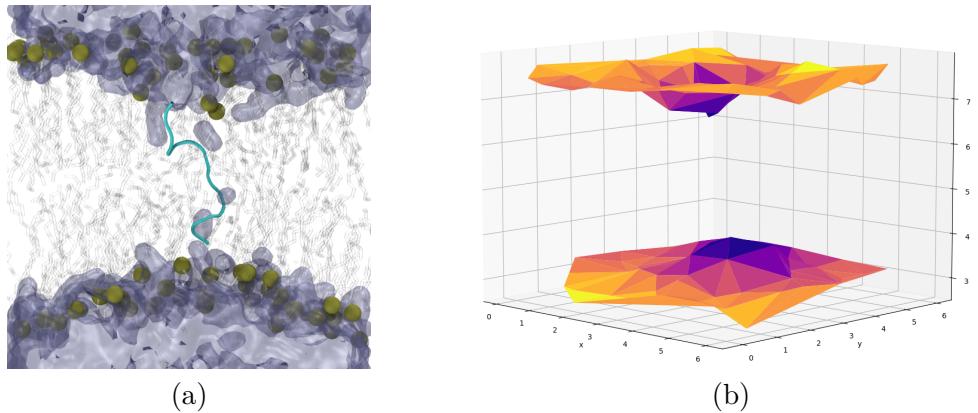


Figure 3.4: (a) Inside view of the membrane ; (b) Plot of the position of P atoms using colors according to their height

With these two captures, two phenomena are exposed:

- **A localized thinning** of the membrane in a radius of about 1 nm around the center of mass of the protein (*which corresponds to the area occupied by the atoms of the protein*).
- **A water leakage** inside the hydrophobic layer: several water molecules are constantly hydrating some parts of the protein

Polar and positively charged amino acids are the cause for these defects. Indeed, as some of the positive amino acids (*in this case the lysines*) are located at both ends of the peptide, they will easily interact with the headgroups of the lipids, with stronger interactions between the peptide and POPS, as it bears two negative sites, as opposed to only one in the case of POPC. Thus, lipid headgroups will get closer to the peptide due to their affinity, lowering the thickness locally. Moreover, the positive charges will need to remain hydrated, even though in a hydrophobic region: the entrance of water inside the bilayer will be facilitated, and the polar amino acids will help to stabilize them. As the calcium ion also needs to remain hydrated, it can be expected that the presence of water will help to increase the permeation through the membrane. The Figure 3.4 (a) also sheds the light on the fact that the positive amino acids do not share the same alignment: as a matter of consequence, the calcium ion that translocates could benefit from the polar amino acids all around the peptide, no matter its side of approach in the bilayer. The fact that polar and positively charged amino acids are not aligned also means that the intrusion will not allow the formation of a direct water pore from one side of the membrane to the other (*also visible in Figure 3.4 (a)*). The following table lists the same typical parameters as used previously for the study of the control models. We define the mean thickness as the thickness of the whole bilayer, and the local thickness by the thickness of the bilayer close to the peptide (*i.e. only considering lipids affected by the thinning*).

Lipids	POPS	POPS/POPC	POPC
<b>Box size x (Å)</b>	60.25	62.65	63.59
<b>Box size y (Å)</b>	60.25	62.65	63.59
<b>Box size z (Å)</b>	105.23	99.40	96.54
<b>Mean thickness (Å)</b>	42.37	39.23	39.70
<b>Local thickness (Å)</b>	34.40	33.58	31.66
<b>Area per lipid (Å<sup>2</sup>)</b>	56.72	61.33	63.17

Table 3.2: Simulation size and area per lipid for each lipid considered

It is possible to see that the thinning is equivalent in both POPC and POPS models: the thickness of the membrane in this region seems to be lowered by about 7.5 Å. The deuterium order parameter is also locally affected by the insertion of water and peptide inside the membrane: as there is less order, the value decreases.

As expected, the system composed of 15% POPS / 85% POPC has results that are in between the two extreme cases (either full POPS or full POPC). Still, the thickness of the membrane around the peptide is closer to the value of the full-POPS membrane than it is to the other one; leading to think that the peptide could cause the aggregation of the PS lipids.

This hypothesis is also supported by the fact that the distance between the center of mass (COM) of the peptide and the P atom has a decrease tendency after some time for some PS lipids, in each leaflet, as observable on Figure 3.5.

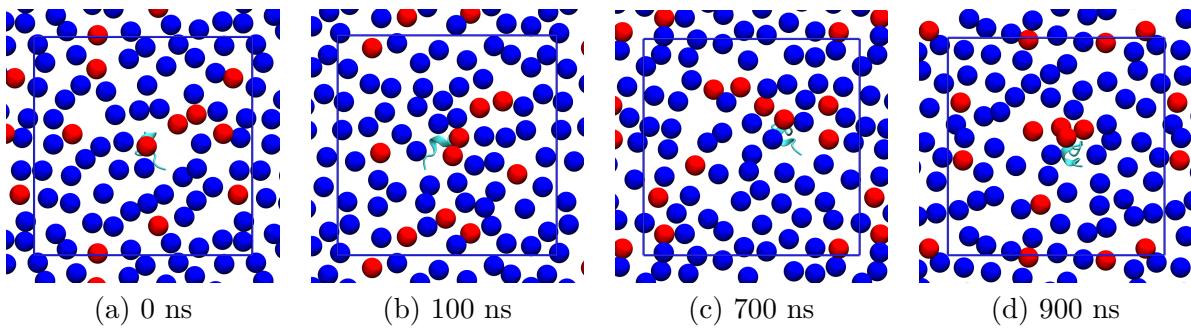


Figure 3.5: Captures of the top leaflet of the POPS/POPC system at various times in simulation. Blue balls represent POPC headgroups, red balls represent POPS headgroup; blue lines delimit the periodic box.

This aggregation phenomenon should reduce the number of possibilities for the approach of calcium ions (*as told earlier, electrostatic interactions are acting as an attractive force between the positive calcium ion and the negative phospholipids*) coming close to the skeletal cells, which favors its closeness with the peptide and is in favor for an increased permeation. Nevertheless another hypothesis, which was not within the scope of this internship (*but is explored by researchers in this project*), supposes a higher concentration of proteins in a same area. Thus, considering the water leakage, and similarly to what can be seen for multiple studies [32, 33], proteins could gather and form a continuous water channel, making it possible for ions and other hydrophilic species to get from one side of the membrane to the other. According to the captures of the simulation in Figure 3.4, if a channel were to exist, it would use the barrel-stave model: meaning that there would be no significant distortion of the membrane (*lipids would stay organized in leaflets*) [34].

## 3.2 Translocation of a calcium ion through the model membrane

Because the translocation of an ion through a bilayer requires an energy which is usually higher than  $k_B T$ , it is very unlikely to see it happen in regular MD simulations. Moreover, in order to generate the free energy profile of the translocation, it is necessary to sample an important number of configurations of the ion through the bilayer, as depending on the type of residue that is translocated, and depending to the nature of the bilayer, profiles can have very different shapes [35].

In order to observe the translocation and to analyse it, performing a biased molecular dynamics simulation was necessary. The following two sections will discuss the two different ways of generating configurations (*windows*), and will be the occasion to dress a review of both methods.

### 3.2.1 Pull method

The first approach we considered was pulling the ion, using a virtual spring, through the system. This method has the advantage of producing a continuous trajectory, from which windows can be extracted. This method also has the advantage of benefiting from an important amount of documentation: even one of GROMACS tutorials is about pulling a peptide away from the rest of the protofibril to perform Umbrella Sampling [36, 37]. First, we studied the membrane containing only POPS lipids; some captures of the trajectory can be observed on Figure 3.6.

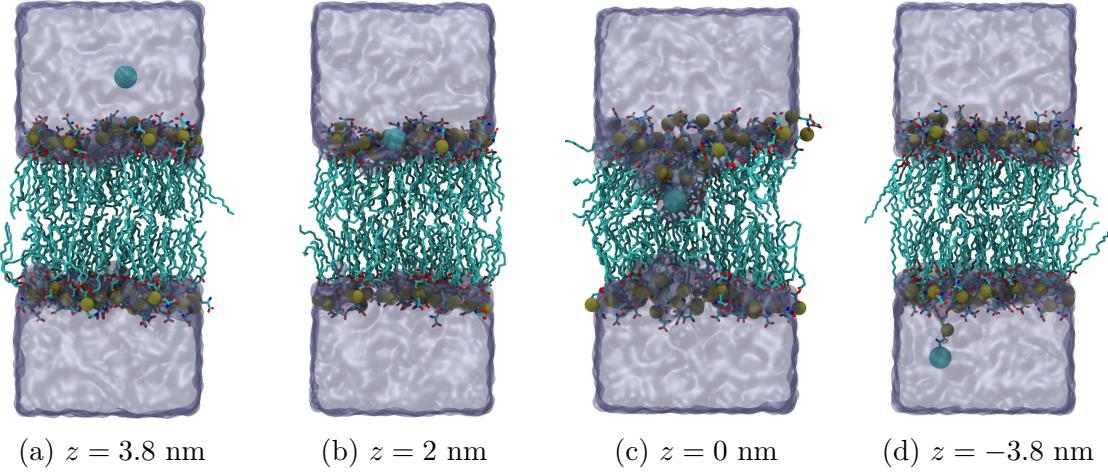


Figure 3.6: Captures of the calcium ion in a POPS system at various levels of penetration,  $z = 0$  being the center of the bilayer (a video of the simulation is available online)

These captures show us that the translocation of the ion causes an ion-induced defect [38], as visible in capture (c). This means that the ion remains constantly hydrated, by carrying water molecules. However, as visible in the following images, this does not cause the formation of a water pore, as the lipids tend to get back to their initial position after the ion gets further in the membrane. The same phenomenon is visible for the lower leaflet.

However, an interesting behavior, which is not well described in literature, was obtained. Indeed, as the ion gets deeper in the membrane, it keeps its interaction with the carboxyl (*if existing*) and phosphate groups of some lipids (the number depends of the lipid) and also pulls them. For a few lipids, the pulling even entitles them to pass from a leaflet to another: this phenomenon is called flip-flop. It is possible to guess that the flip-flop requires an important amount of energy, as it requires the displacement of more atoms and water molecules (*the headgroups also bear charges*). This event is an artifact due to the high speed imposed to the crossing, and would never happen in a real physiological system. A closer view of lipid flip-flop for POPC and POPS is available in Figure 3.7.

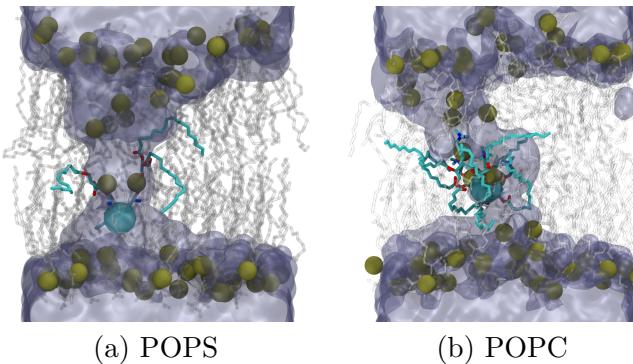


Figure 3.7: Capture of the pulling simulation with the calcium ion at  $z = -1.2$  nm  
Lipids pulled with the ion are highlighted in color, brown balls correspond to Phosphate atoms

The number of lipids being pulled with the calcium ion is not fixed, and will vary with the pulling rate and the lipid itself. Indeed, as POPS lipids bear more negative groups than POPC lipids, the number of lipids required to stabilize the positive ion will be less important with POPS; and this is what we observe with the Figure above. By repeating this simulation with different rates (*with POPS*), the Table 3.3 was obtained.

Pull rate (nm/ns)	1.00	0.10	0.01
Number of lipid flip-flops	1	2	4

Table 3.3: Number of lipid flip-flops observed when changing the pulling rate of the calcium ion

These results are coherent with what could have been expected: indeed, the slower the calcium ion gets pulled, the more equilibrated the defect will be. Thus, as the calcium ion cannot move close to the phospholipids' headgroups, being under a heavy restraint caused by the virtual spring pulling it, the phospholipids are inclined to go in the direction of the ion. However, we can expect the number of lipid flip-flop caused by a single ion to be limited to a maximal value, as pulling a lipid is energetically demanding and as the calcium only bears a +2 charge; but this study was not performed during the internship because of a lack of time and to avoid using unnecessary computing time.

The event of lipid translocation has the consequence of breaking the symmetry, as the composition of each leaflet will no longer be identical as soon as the calcium ion starts to drag lipids. After generating, equilibrating and gathering production data for each window, the WHAM algorithm implemented in GROMACS was used to reunite all conformations in one profile, yielding the Potential of Mean Force (PMF); Figure 3.8 is obtained.

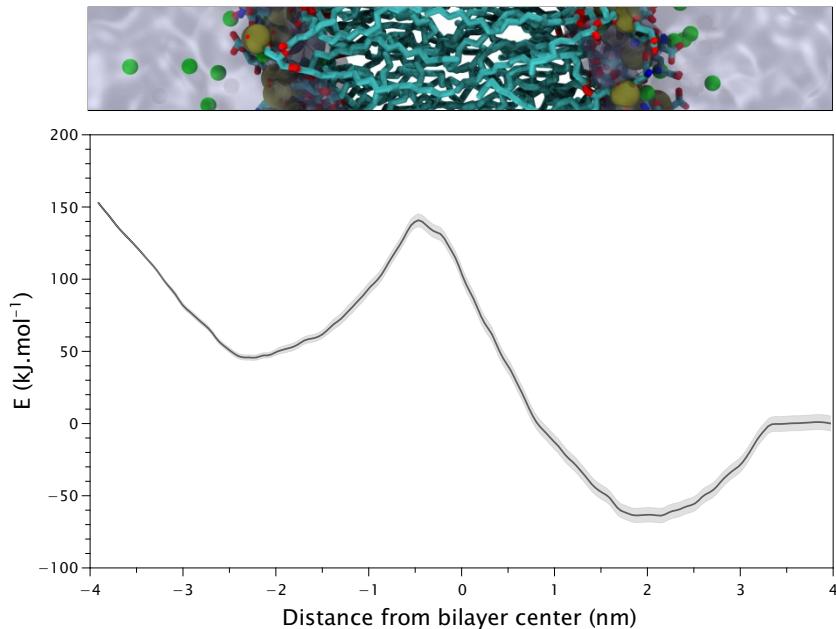


Figure 3.8: Potential of Mean Force associated to the translocation of a calcium ion through a full POPS membrane, using the pull method (*statistical error is represented in gray*)

Using this plot, it is possible to distinguish the main regions (*at least for positive values of z*) of the bilayer based on the behavior of the free energy : bulk water corresponds to the constant free energy; headgroups are associated with the pit, and esters as well as acyl chains are the central peak of the profile (*from 0 to 2 nm*).

What can be noticed is that the plot is not symmetric, even though the system is (*which should*

*imply the symmetry on every parameter).* Indeed, as it was said earlier, the fact the calcium ions dragged lipids increased the required force applied to the ion, and as a matter of consequence the total free energy. It is important to acknowledge the presence of this asymmetry, which is often left undiscussed in litterature (*authors often simulate only half of the translocation and assume the symmetry of their profile to save computing time*).

### 3.2.2 Insertion method

While the previous method yields as expected results with low statistical error, its way of generating windows implies that each window has a bias caused by the path the ion previously followed: such as the lipid dragging. However, although it is important to acknowledge the presence of such phenomenon as explained previously, the goal of this work is mainly to get the free energy profile of the ion translocation, without any other residue affecting the energy. To give an independent behavior to every window, it was decided to generate them individually, by inserting the calcium ion every  $\Delta z = 0.1$  nm. No other constraint along the  $x - y$  was added, because of the symmetry of the system along this plane: thus no matter where the ion would be, it would not make a difference (*energetically speaking*). After equilibrating and gathering production data for each window, the WHAM algorithm was used as previously for both POPS and POPC systems; giving Figure 3.9.

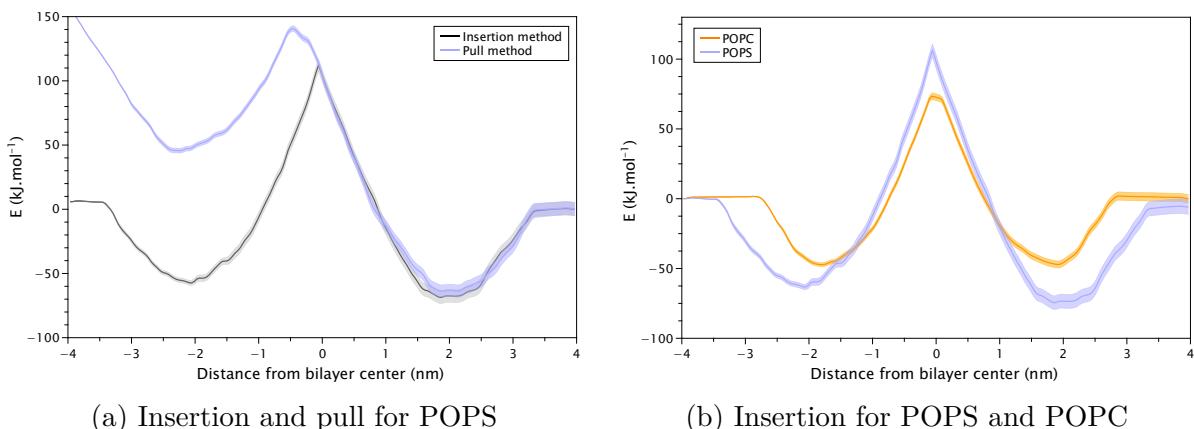


Figure 3.9: PMF curves (a) comparing insertion and pull methods and (b) comparing the two extreme lipid composition cases with the insertion method

With this method, it is possible to observe profiles that are more symmetric, even without performing any sort of symmetrization with the WHAM algorithm. Indeed, the insertion of calcium ion could be compared to a simulation where the same ion would be pulled at an infinitely low speed; hence allowing better equilibration and removing artifacts induced by pulling. The PMF profile in Figure 3.9 (a) also highlights that both methods provide identical free energies, with an energetic difference between bulk water and the membrane/water interface for PS lipids of  $67.5 \text{ kJ.mol}^{-1}$  as well as an identical slope inside the bilayer; which conforms the accuracy of this method.

Insertion profiles still exhibit a bit of asymmetry (*PMF profile for negative z values is not the symmetrical image of the positive z profile, and extrema can be shifted*), although the most important asymmetry is noticed for POPS. Indeed, difference can be noticed at the pits, with  $\Delta E_{\text{pits}} = 9.9 \text{ kJ.mol}^{-1}$ . Still, this value remains negligible (*statistical error zones still meet each other, which is satisfying*) and can be explained: this zone is marked by the strong interactions between the calcium ion and the headgroups, hence bringing unavoidable systematic sampling errors, that would disappear after several hundreds nanoseconds of equilibration. Using POPC, these energy differences are limited to a few tenths of  $\text{kJ.mol}^{-1}$ .

While looking at Figure 3.9 (b), it is possible to see that the pits are narrower and less deep for POPC. This can be explained with the nature of the headgroups: as shown in Figure 2.4, POPS lipids bear carboxyl groups at their end, allowing interactions with calcium directly in bulk water. Moreover, these entitle for a stronger interaction, thus lowering even more the energy. For POPC, the lipids bear amines  $\text{NH}_3^+$  at their end, thus removing the favorable interaction that the calcium had in the water phase. Inside the bilayer, the ion can bind with Phosphate, lowering the energy. The PMF energy at the center of bilayer ( $z = 0 \text{ nm}$ ) is also reduced with POPC (see *Table 3.4 on page 24 for precise values*), which can be associated with the higher area per lipid and the lower thickness of the bilayer for this composition: the translocation of the calcium ion requires less water to get through, and the intrusion of water is easier as lipids are more spaced. As a consequence, it is thus easier for the calcium ion to get through the membrane.

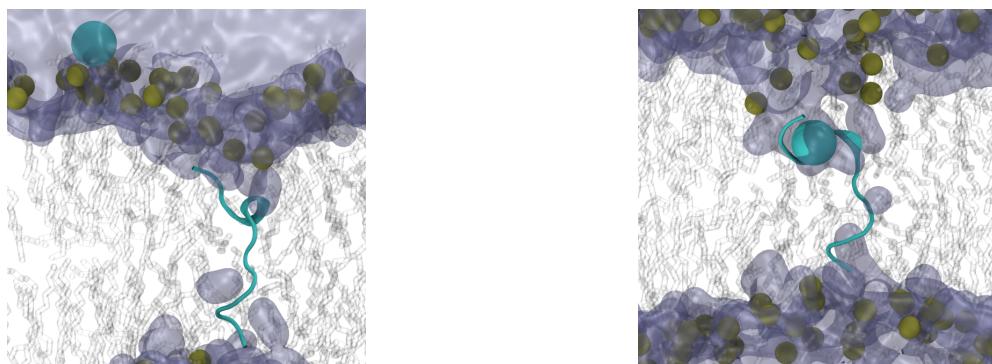
As a matter of consequence, insertion simulations will give a better result for the profiles and should be chosen when realizing PMF of permeation of ions in membranes, even though the computing time is increased. Using these profiles, we can then obtain the permeation energies for the calcium ion, by looking at the energy of the energetic barrier. This value is higher for POPS ( $\Delta E_{\text{permeation, POPS}} = 169.43 \text{ kJ.mol}^{-1}$ ) than POPC ( $\Delta E_{\text{permeation, POPC}} = 120.71 \text{ kJ.mol}^{-1}$ ). Umbrella sampling simulations using a mixture of POPS and POPC are still being currently performed, but results in between those of POPS and POPC can be expected, as it was the case for previous structural parameters considered.

### 3.3 Translocation of a calcium ion through the peptide-containing membrane

Last section has been the occasion to validate the usage of the insertion method, which gives better results. Hence, we have decided to use this method for the following simulations.

#### 3.3.1 Obtention of the Potential of Mean Force for POPS

As the protein preferably binds to negative phospholipids, it was decided to perform Umbrella Sampling first on a membrane containing only POPS lipids. During the insertion process of the calcium ion in the membrane, it was possible to observe two events :



(a) In bulk water, the ion does not tend to align with the peptide, but rather to align with the closest headgroups it can reach. In the POPC simulation, the ion goes on a random part of the surface (*no particular attraction, P atoms are buried under  $\text{NH}_3^+$  groups*)

(b) Inside the membrane (*and no matter its composition*), the ion is always located within a radius of less than 1 nm next to the peptide. In some cases (*as shown in the picture*), the ion is even in contact with the peptide.

Figure 3.10: Captures of windows containing the ion (a) at the water/membrane interface and (b) inside the membrane

Thus, when the ion passes through a membrane close to the peptide (so that it can form interactions with it), it will preferably go close to the peptide, in order to benefit from the molecules of water that are present and from the presence of hydrophilic groups. This creates a more favorable environment for the ion, which should lower the energy implied in the translocation event. However, if the ion passes through the membrane "far" from the peptide (*thus having no possibility to benefit from the peptide and its surroundings*), energy would be similar to the model membranes. Fortunately, the calcium ion is a positive ion, and like the protein, it should be attracted to POPS phospholipids. Moreover, as the peptide also favors the aggregation of negative phospholipids around it, then the calcium ion should also be attracted by the zone where the peptide is.

A simulation where the ion is pulled was also realized; a light restraint guiding the ion around the peptide was applied in bulk water. Indeed, applying no restraint in the case of a full POPS or POPC membrane could lead to simulation where the ion would be too far from the peptide to feel any difference in the permeation. For reproducibility purposes, this restraint was also applied to the system having a mix of POPS and POPC lipids. All these simulations also demonstrated the affinity for the peptide surroundings inside the membrane (*calcium followed the path created by the water intrusion*), hence giving trust that the resulting configurations extracted from the insertion method were reliable. The pull simulation also showed the presence of lipid flip-flop. Both animations of pulling and insertion methods are available online.

As told earlier, the fact that the ion stays close to the water molecules inside the bilayer is a good sign indicating that the free energy barrier should be lowered. To make sure of this, Umbrella Sampling was performed on the configurations obtained with the insertion method, and the resulting profile is shown in Figure 3.11.

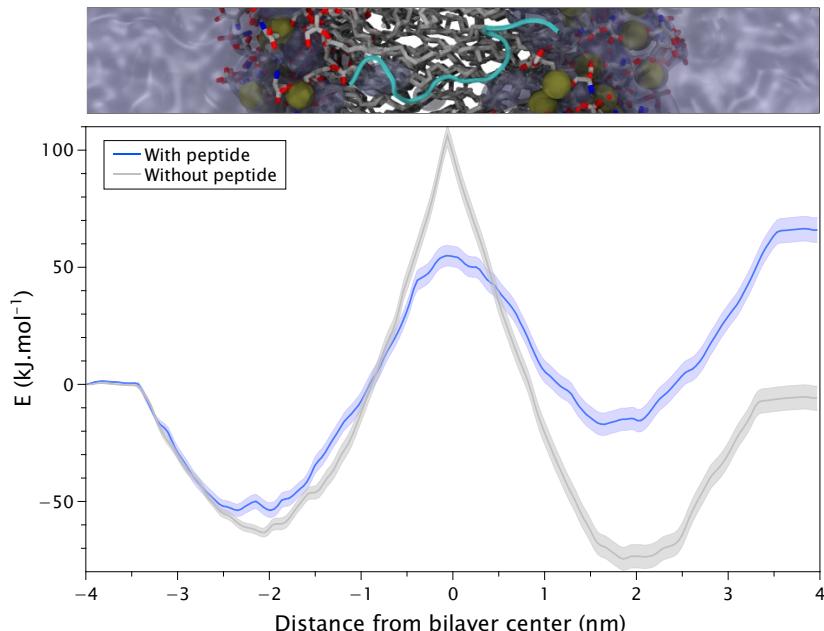


Figure 3.11: Potential of Mean Force associated to the translocation of a calcium ion through a full POPS membrane containing the sequence 115-129 of the Lys49-PLA<sub>2</sub> like protein and in the same control membrane, using the insertion method

The profile obtained is at first quite surprising: the free energy for bulk water is not identical on each side (*even though it should be*), and the symmetry is lost. Yet, it is already easy to see that the energy barrier at  $z = 0$  nm is reduced, no matter the side where the origin of the PMF is defined. In function of the side of insertion of the ion, the following free energy barriers can

be found:

- $\Delta E_{\text{barrier},-z} = 108.19 \text{ kJ}\cdot\text{mol}^{-1}$  for negative  $z$  values,
- $\Delta E_{\text{barrier},+z} = 69.70 \text{ kJ}\cdot\text{mol}^{-1}$  for positive  $z$  values

The peptide was inserted so that the rest of the protein should be on the negative  $z$  side: making this side the outer layer of the cell membrane: thus it should be considered that the energy barrier for the translocation of a calcium ion through the membrane is around  $108.19 \text{ kJ}\cdot\text{mol}^{-1}$ . To be entirely sure about this result, it would be interesting to perform more tests on this regard.

Although the peptide was never said to be symmetrical, the shape of the profile still led to many interrogations. By studying the evolution of the profile along the production time (*the longer the simulation time, the more accurate the result*), it is possible to observe that the profile starts by being symmetric, and then evolves to what we know (Figure 3.12).

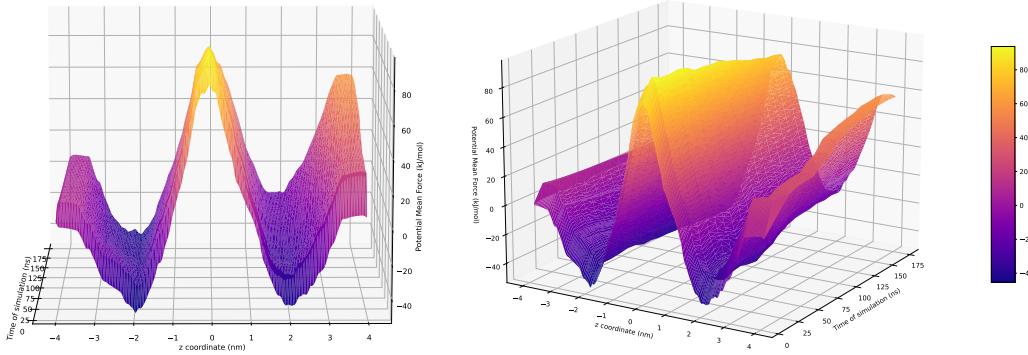


Figure 3.12: 3D-colormap showing the evolution of the Potential of Mean Force through the time of simulation

More than a simple asymmetry, the presence of the peptide seems to have a specific effect limited to a half of the profile (*in this case for positive  $z$  values*). For instance, the energy difference between water and headgroups is now :

- $\Delta E_{\text{water} \rightarrow \text{headgroups},-z} = -52.54 \text{ kJ}\cdot\text{mol}^{-1}$  for negative  $z$  values,
- $\Delta E_{\text{water} \rightarrow \text{headgroups},+z} = -81.09 \text{ kJ}\cdot\text{mol}^{-1}$  for positive  $z$  values

A simulation using the pull method was performed (*not shown*), and gave the same value for  $\Delta E_{\text{water} \rightarrow \text{headgroups},+z}$ , confirming the results obtained earlier. This difference can be explained by the presence of the peptide: as visible in Figure 3.10 and Figure 2.3, the top part of the peptide has a small helix. As the amino acids at the extremas of the peptide have positive charges, the presence of a helix will have the consequence of increasing the local charge, which will require more headgroups to counter. When the calcium ion, still in water, comes closer to the membrane surface right above the peptide, it will be difficult for the lipids to interact with both residues, thus increasing the free energy. When the ion comes at the depth of the interface, the lipids can stabilize both the peptide and the ion, giving a local minimum.

### 3.3.2 Study of the asymmetry of the PMF profile

Umbrella Sampling was also performed for the POPC membrane containing the peptide, the resulting profile is shown in Figure 3.13:

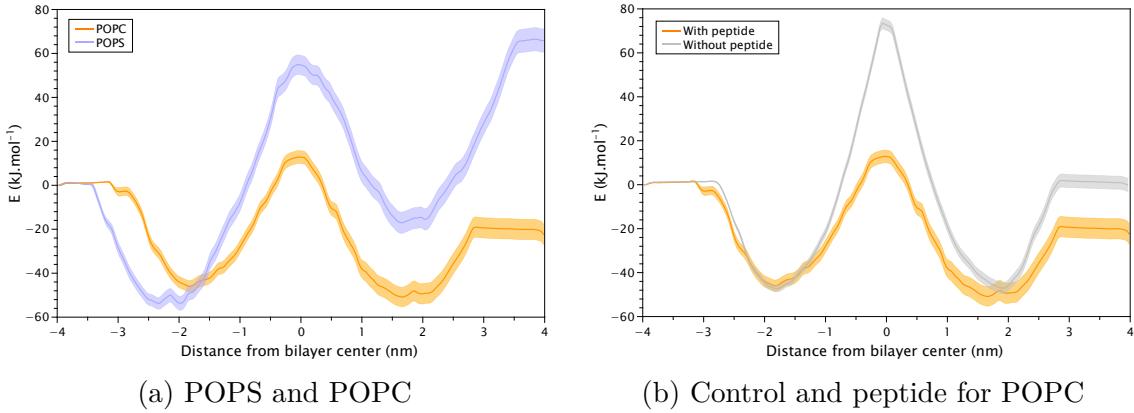


Figure 3.13: PMF curves (a) comparing different lipid compositions and (b) comparing the presence and absence of peptide for POPC

It is possible to see that the energy increase for positive values of  $z$  is no longer observed here, which can be explained by the fact that POPC lipids do not expose a negative charge at the surface of the membrane. We can expect that the 20 kJ.mol $^{-1}$  difference between both sides in bulk water is still due to the peptide. It is also possible to see with Figure 3.13 (b) that the energy barrier is also lowered in the case of POPC: in the case of a PS/PC mixture, it can thus be deduced that the energy barrier would also be lowered. All characteristic energies for the POPC membrane are available in Table 3.4.

However, this simulation still yields an asymmetric profile. In order to make sure that the unique source of asymmetry is the peptide, a new simulation is carried out, where the peptide is replaced by the mutated sequence: KKYYRYYLYYRYKK. The same experimental conditions were applied, and Figure 3.14 was obtained:

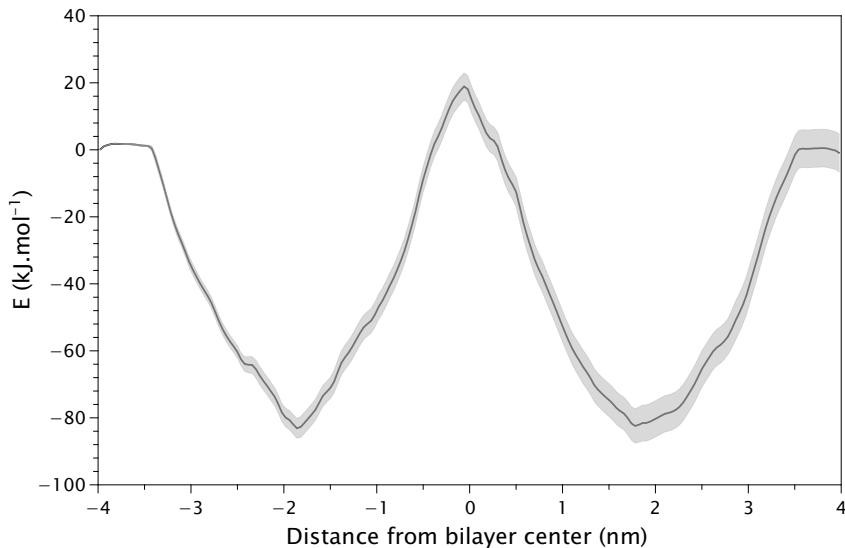


Figure 3.14: Potential of Mean Force associated to the translocation of a Calcium ion through a full POPS membrane containing the mutated peptide, using the insertion method

It is possible to observe that using a symmetrized version of the peptide, the potential of mean force is now symmetric, thus proving that the peptide is the main reason for the asymmetry. Using this mutation of the peptide, we find a permeation barrier of  $\Delta E_{\text{barrier}} = 101.67$  kJ.mol $^{-1}$ . This result is coherent with the result obtained for the original peptide. Thus, by symmetrizing the other part of the peptide (*thus giving the sequence KKCLPKLKPLCKK*), it could be expected

to have a smaller barrier, around  $70 \text{ kJ.mol}^{-1}$ . Simulations are currently being performed with this new peptide, and should be done in September.

		$z < 0$		$z > 0$	
		Control	Peptide	Control	Peptide
<b>POPS</b>	Bulk water	0	0	-5.37	66.21
	Headgroups	-62.99	-52.54	-73.42	-14.88
	Bilayer center	106.13	54.82		
<b>POPC</b>	Bulk water	0	0	0.26	-20.52
	Headgroups	-47.22	-45.61	-47.03	-50.05
	Bilayer center	73.48	12.87		

Table 3.4: Energies (in  $\text{kJ.mol}^{-1}$ ) of characteristic regions of profiles, for various lipid compositions. The energy at  $z = -4$  was arbitrarily set to 0 for every profile

# 4. Conclusion and future perspectives

## 4.1 Conclusions on the subject and discussion

During this internship, Umbrella Sampling simulations were performed on different models of membranes, containing the myotoxic sequence of the Lys49 PLA<sub>2</sub>-like protein secreted in *Bothrops Asper*'s venom. These experiments allowed to discover that the insertion of the peptide reduced the free energy barrier of the translocation of calcium both for POPS and POPC membranes, thus increasing the permeation for those ions, which are directly linked to the necrosis of cells. Simulations for membranes having a mix of POPS and POPC lipids are still running, but their results are expected to be similar and in between both extreme cases that were studied (*either POPS or POPC*).

However, the barrier still remains quite high, and the "massive influx" announced in the litterature would not happen only by the insertion of the single peptide. What could explain such event could be the aggregation of more proteins, hypothesis which can be supported by the important quantity of venom secreted by the viper. Indeed, this could stabilize the formation of a water pore, although this was never observed in experimental studies. The aggregation could also create a phenomenon of transient water pore. Other members of the project are currently working on coarse-grained simulations where multiple peptides are present.

This work was also the occasion of using Hydrogen Mass Repartitioning (*see Appendix D*) for Umbrella Sampling on a protein/lipid system, which was not deeply studied. This allowed the performance to be doubled, while the accuracy remained suitable.

## 4.2 Personal conclusion

It has been a real pleasure to work on this project, which I found to be very exciting. It made it possible for me to learn more and develop my skills in Computational Chemistry with advanced researchers, whilst doing something that would be useful in the future for further research on *Bothrops asper*'s venom. It also allowed to confirm my wishes for my future career, as I aspire to work in this domain.

This internship also was my first occasion to write a scientific paper, which I found very interesting.

# Bibliography

- [1] U. of Porto, U.Porto at a Glance, Accessed on 2022-04-21, **2022**, <https://sigarra.up.pt/up/en/p/u-porto-em-sintese>.
- [2] H. S. Bawaskar, P. H. Bawaskar, P. H. Bawaskar, *Journal of the Royal College of Physicians of Edinburgh* **2021**, *51*, 7–8.
- [3] World Health Organization, Snakebite envenoming, **2021**, <https://www.who.int/news-room/fact-sheets/detail/snakebite-envenoming>.
- [4] Animal Diversity Web, Bothrops asper: INFORMATION, [https://animaldiversity.org/accounts/Bothrops\\_asper/#physical\\_description](https://animaldiversity.org/accounts/Bothrops_asper/#physical_description).
- [5] M. Sasa, D. K. Wasko, W. W. Lamar, *Toxicon* **2009**, *54*, 904–922.
- [6] R. Otero-Patiño, *Toxicon* **2009**, *54*, 998–1011.
- [7] B. Gratwicke, Fer-de-lance Bothrops asper, **2011**, <https://www.flickr.com/photos/briangratwicke/5414237015/>.
- [8] A. Alape-Girón, L. Sanz, J. Escolano, M. Flores-Díaz, M. Madrigal, M. Sasaa, J. J. Calvete, *Journal of Proteome Research* **2008**, *7*, 3556–3571.
- [9] B. Lomonte, J. M. Gutiérrez, *Acta chimica Slovenica* **2011**, *58*, 647–659.
- [10] W. Szlasa, I. Zendran, A. Zalesiska, M. Tarek, J. Kulbacka, *Journal of Bioenergetics and Biomembranes* **2020**, *52*, 321–342.
- [11] P. J. Bond, J. Holyoake, A. Ivetac, S. Khalid, M. S. Sansom, *Journal of Structural Biology* **2007**, *157*, 593–605.
- [12] M. T. Murakami, E. Z. Arruda, P. A. Melo, A. B. Martinez, S. Calil-Eliás, M. A. Tomaz, B. Lomonte, J. M. Gutiérrez, R. K. Arni, *Journal of Molecular Biology* **2005**, *350*, 416–426.
- [13] RSCB Protein Data Bank, Crystal structure of Bothrops asper myotoxin II complexed with the anti-trypanosomal drug suramin, <https://www.rcsb.org/structure/1y4l>.
- [14] W. Humphrey, A. Dalke, K. Schulten, *Journal of Molecular Graphics* **1996**, *14*, 33–38.
- [15] M. Cintra-Francischinelli, P. Pizzo, L. Rodrigues-Simioni, L. A. Ponce-Soto, O. Rossetto, B. Lomonte, J. M. Gutiérrez, T. Pozzan, C. Montecucco, *Cellular and Molecular Life Sciences* **2009**, *66*, 1718–1728.
- [16] S. Jo, T. Kim, V. G. Iyer, W. Im, *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.
- [17] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *Journal of Computational Chemistry* **2009**, *30*, 1545–1614.
- [18] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda, W. Im, *Journal of Chemical Theory and Computation* **2016**, *12*, 405–413.
- [19] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda, W. Im, *Journal of Computational Chemistry* **2014**, *35*, 1997–2004.

- [20] S. Jo, J. B. Lim, J. B. Klauda, W. Im, *Biophysical Journal* **2009**, *97*, 50–58.
- [21] S. Jo, T. Kim, W. Im, *PLoS ONE* **2007**, *2*, e880.
- [22] J. Lee, D. S. Patel, J. Ståhle, S.-J. Park, N. R. Kern, S. Kim, J. Lee, X. Cheng, M. A. Valvano, O. Holst, Y. A. Knirel, Y. Qi, S. Jo, J. B. Klauda, G. Widmalm, W. Im, *Journal of Chemical Theory and Computation* **2019**, *15*, 775–786.
- [23] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, A. D. MacKerell, *Nature Methods* **2016**, *14*, 71–73.
- [24] Y. Gao, J. Lee, I. P. S. Smith, H. Lee, S. Kim, Y. Qi, J. B. Klauda, G. Widmalm, S. Khalid, W. Im, *Journal of Chemical Information and Modeling* **2021**, *61*, 831–839.
- [25] G. Torrie, J. Valleau, *Journal of Computational Physics* **1977**, *23*, 187–199.
- [26] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, P. A. Kollman, *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
- [27] J. S. Hub, B. L. de Groot, D. van der Spoel, *Journal of Chemical Theory and Computation* **2010**, *6*, 3713–3720.
- [28] P. Mukhopadhyay, L. Monticelli, D. P. Tieleman, *Biophysical Journal* **2004**, *86*, 1601–1609.
- [29] N. Kuerka, M. P. Nieh, J. Katsaras, *Biochimica et Biophysica Acta - Biomembranes* **2011**, *1808*, 2761–2771.
- [30] T. J. Piggot, J. R. Allison, R. B. Sessions, J. W. Essex, *Journal of Chemical Theory and Computation* **2017**, *13*, PMID: 28876925, 5683–5696.
- [31] K. C. Giannotti, E. Leiguez, V. Moreira, N. G. Nascimento, B. Lomonte, J. M. Gutiérrez, R. L. D. Melo, C. Teixeira, *BioMed Research International* **2013**, *2013*, DOI 10.1155/2013/807982.
- [32] L. Thøgersen, B. Schiøtt, T. Vosegaard, N. C. Nielsen, E. Tajkhorshid, *Biophysical Journal* **2008**, *95*, 4337–4347.
- [33] I. L. Karle, *Proceedings of the American Philosophical Society* **1999**, *143*, 28–41.
- [34] L. T. Nguyen, E. F. Haney, H. J. Vogel, *Trends in Biotechnology* **2011**, *29*, 464–472.
- [35] J. L. MacCallum, D. P. Tieleman, *Current Topics in Membranes* **2008**, *60*, 227–256.
- [36] J. A. Lemkul, D. R. Bevan, *Journal of Physical Chemistry B* **2010**, *114*, 1652–1660.
- [37] J. Lemkul, *Living Journal of Computational Molecular Science* **2019**, *1*, DOI 10.33011/livecoms.1.1.5068.
- [38] I. Vorobyov, T. Olson, J. Kim, R. Koeppe, O. Andersen, T. Allen, *Biophysical Journal* **2014**, *106*, 586–597.
- [39] A.-L. Chemistry, Amino Acids, Proteins and DNA, Accessed on 2022-04-01, <https://alevelchemistry.co.uk/notes/amino-acids-proteins-and-dna/>.
- [40] C. W. Hopkins, S. L. Grand, R. C. Walker, A. E. Roitberg, *Journal of Chemical Theory and Computation* **2015**, *11*, 1864–1874.
- [41] P. Mukhopadhyay, L. Monticelli, D. P. Tieleman, *Biophysical Journal* **2004**, *86*, 1601–1609.

# Appendix

## A. The 20 amino acids

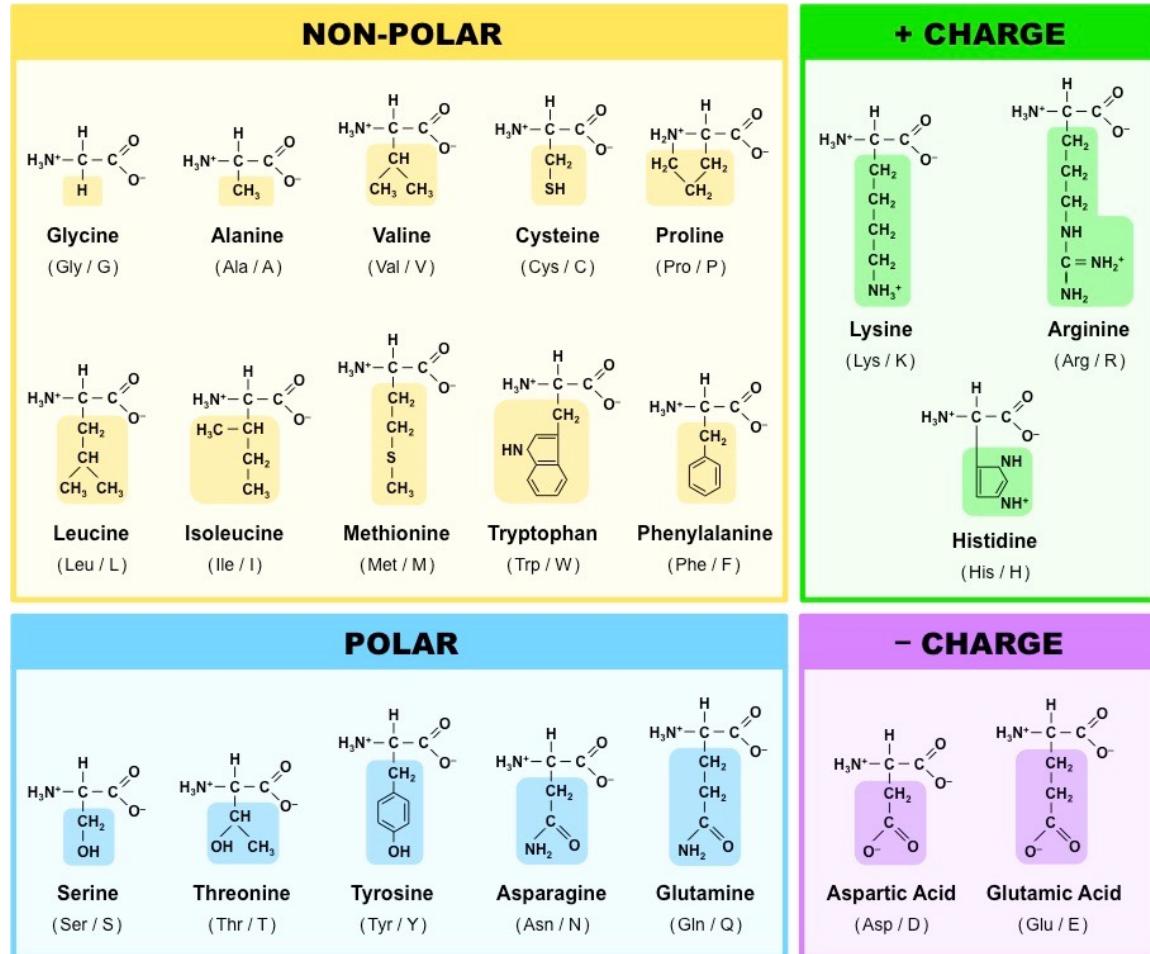


Figure A.1: The twenty amino acids [39]

## B. Structure of a force field file

Due to its important length, the file could not entirely be shown in the report. Instead, only a few lines of each category are shown, to understand the principle.

```

forcefield.itp

[ defaults ]
; nbfunc      comb-rule      gen-pairs      fudgeLJ fudgeQQ
1           2           yes        1.000000     1.000000

[ atomtypes ]
; name at.num mass   charge  ptype    sigma   epsilon ;      sigma_14      epsilon_14
H          1     1.0080  0.310   A       4.00013524445e-02  1.924640e-01
HA1         1     1.0080  0.090   A       2.38760856462e-01  1.882800e-01
HA2         1     1.0080  0.090   A       2.38760856462e-01  1.422560e-01
HA3         1     1.0080  0.090   A       2.38760856462e-01  1.004160e-01
HAL1        1     1.0080  0.090   A       2.35197261589e-01  9.204800e-02
HAL2        1     1.0080  0.090   A       2.38760856462e-01  1.171520e-01
HP          1     1.0080  0.115   A       2.42003727796e-01  1.255200e-01 ;  2.42003727796e-01  1.255200e-01
CCL          6     12.0110 0.340   A       3.56359487256e-01  2.928800e-01
N          7     14.0070 -0.290   A       3.29632525712e-01  3.368000e-01 ;  3.29632525712e-01  4.184000e-04
NC2          7     14.0070 -0.700   A       3.29632525712e-01  8.368000e-01
NH1          7     14.0070 -0.470   A       3.29632525712e-01  8.368000e-01 ;  2.76178602624e-01  8.368000e-01
NH3          7     14.0070 -0.300   A       3.29632525712e-01  8.368000e-01
NH3L         7     14.0070 -0.300   A       3.29632525712e-01  8.368000e-01
O          8     15.9990 -0.510   A       3.02905564168e-01  5.020800e-01 ;  2.49451641079e-01  5.020800e-01
O2L          8     15.9994 -0.780   A       3.02905564168e-01  5.020800e-01
OT          8     15.9994 -0.834   A       3.15057422683e-01  6.363864e-01
SOD          11    22.9898  1.000   A       2.51367073323e-01  1.962296e-01
PL          15    30.9740  1.500   A       3.83086448800e-01  2.447640e+00
S          16    32.0600 -0.230   A       3.56359487256e-01  1.882800e+00
CLA          17    35.4500 -1.000   A       4.04468018036e-01  6.276000e-01
CAL          20    40.0800  2.000   A       2.43571709540e-01  5.020800e-01

[ nonbond_params ]
; i       j       func      sigma   epsilon
CLA      SOD      1     3.32394311738e-01  3.51037600000e-01
SOD      O2L      1     2.81523994932e-01  3.13883680000e-01
SOD      OBL      1     2.78851298778e-01  3.13883680000e-01
SOD      OCL      1     2.87760285959e-01  3.13883680000e-01

[ bondtypes ]
; i       j       func      b0      Kb
CA       CA      1     1.375000e-01  2.552240e+05
CEL1     CEL1     1     1.340000e-01  3.681920e+05
CEL1     CTL2     1     1.502000e-01  3.054320e+05
CP1      C       1     1.490000e-01  2.092000e+05
CP2      CP1     1     1.527000e-01  1.861880e+05
CP2      CP2     1     1.537000e-01  1.861880e+05
CP3      CP2     1     1.537000e-01  1.861880e+05
CT1      C       1     1.490000e-01  2.092000e+05
CT2      CA      1     1.490000e-01  1.924640e+05
CT2      CT1     1     1.538000e-01  1.861880e+05
CT2      CT2     1     1.530000e-01  1.861880e+05
CT3      C       1     1.490000e-01  2.092000e+05
CT3      CT1     1     1.538000e-01  1.861880e+05
CTL1     CCL     1     1.522000e-01  1.673600e+05
CTL1     CTL2     1     1.538000e-01  1.861880e+05
CTL1     HAL1     1     1.111000e-01  2.585712e+05
CTL1     HBL     1     1.080000e-01  2.761440e+05

[ pairtypes ]
; i       j       func      sigma1-4      epsilon1-4
H       CP1     1     1.89271432669e-01  8.97368027066e-02
H       CP2     1     1.89271432669e-01  8.97368027066e-02
H       CP3     1     1.89271432669e-01  8.97368027066e-02
H       CT1     1     1.89271432669e-01  8.97368027066e-02

[ angletypes ]
; i       j       k       func      th0      Kth      s0      Kub
CA       CA       CA      5     1.200000e+02  3.3472000e+02  2.4162000e-01  2.9288000e+04
CP2     CP1       C      5     1.123000e+02  4.3513600e+02  0.0000000e+00  0.0000000e+00
H       NH1       C      5     1.230000e+02  2.8451200e+02  0.0000000e+00  0.0000000e+00
OSL     CTL1      CTL2     5     1.101000e+02  6.3345760e+02  0.0000000e+00  0.0000000e+00

[ dihedraltypes ]
; i       j       k       l       func      phi0      Kphi      mult
C       CT1      CT2      CA      9     1.800000e+02  5.355520e+00      1
CA       CA       CT2      CT1     9     0.000000e+00  4.476880e+00      1
H       NH1       C       CP1     9     1.800000e+02  1.046000e+01      2
O       C       CT1      CT2     9     0.000000e+00  5.857600e+00      1

[ dihedraltypes ]
; i       j       k       l       func      q0      Kq
C       HC       HC      NC2     2     0.000000e+00  0.0000000e+00
N       C       CP1      CP3     2     0.000000e+00  0.0000000e+00
NC2     X       X       C       2     0.000000e+00  3.765600e+02
NH1     X       X       H       2     0.000000e+00  1.673600e+02

```

## C. Structure of a parameters file for simulations

The file shown here corresponds to the set of parameters inputed in GROMACS, listing all the conditions that must be applied to the system. Here, as described later, is presented the equilibration of a system under a NPT ensemble.

Parameters given for a simulation under a NPT ensemble	
define	= -DPOSRES -DPOSRES_FC_BB=50.0 -DPOSRES_FC_SC=0.0 -DPOSRES_FC_LIPID=0.0 -DDIHRES -DDIHRES_FC=0.0
integrator	= md
dt	= 0.002
nsteps	= 5000000
nstxtcout	= 5000
nstvout	= 5000
nstfout	= 5000
nstcalcenergy	= 100
nstenergy	= 1000
nstlog	= 1000
;	
cutoff-scheme	= Verlet
nstlist	= 20
rlist	= 1.2
vdwtype	= Cut-off
vdw-modifier	= Force-switch
rwdw_switch	= 1.0
rwdw	= 1.2
coulombtype	= PME
rcoulomb	= 1.2
;	
tcoupl	= berendsen
tc_grps	= SOLU MEMB SOLV
tau_t	= 1.0 1.0 1.0
ref_t	= 310.15 310.15 310.15
;	
pcoupl	= berendsen
pcoupltype	= semiisotropic
tau_p	= 5.0
compressibility	= 4.5e-5 4.5e-5
ref_p	= 1.0 1.0
refcoord_scaling	= com
;	
constraints	= h-bonds
constraint_algorithm	= LINCS
continuation	= yes
;	
nstcomm	= 100
comm_mode	= linear
comm_grps	= SOLU_MEMB SOLV

The file is separated into 6 different parts :

- The first one defines the general conditions of the simulation, such as the integrator used (*here is used md, which corresponds to the leap-frog algorithm*), the time step of integration, the number of steps and the frequency of outputs. Specific constraints (*to certain atoms or groups of atoms*) are also defined here (*usually at the first line so that the user does not miss them*)
- The second part contains the parameters for non-bonded interactions (*Coulomb and van der Waals interactions*) and the neighbour searching: the cut-off type (the potential can be abruptly turned to zero or progressively) and its radius for each interaction
- The third part is related to the temperature coupling: the type of coupling, the atom groups considered, their time constant for coupling and the desired temperature
- The fourth part is related to pressure coupling, and works the same as temperature
- The fifth part is related to generalized constraints for bonds
- The last part is relative to the center of mass motion removal for certain groups

## D. Validation of Hydrogen Mass Repartitioning on a Lipid/Protein System

When doing Molecular Dynamics, laws of mechanics are being used, linking the forces applied to the system to its accelerations. Thus, in order to obtain the updated position of atoms, integrations using numerical methods have to be performed. The timestep for these integrations cannot be randomly chosen, and has to follow some rules in order to take into account every information. Indeed, when the timestep is too important, some movements can be forgotten, as their frequency motion in the system will be higher. As molecular mechanics and molecular dynamics do not account for electrons, the fastest motion is the bond vibration involving hydrogen atoms ( $\simeq 10 - 100$  fs), as these atoms are light compared to others in the system.

We typically choose a timestep which is 10 times lower than the fastest frequency motion in the system: making it in general lower or equal to 1 fs. However, as systems in biochemistry can exceed tens of thousands of atoms pretty easily, having a low time step is at the cost of a long calculation time. For instance, using the same forcefield and parameters as in our system below, a time step of 2 fs yields a performance of calculation of less than 60 ns/day, which tremendously slows down our efficiency when it comes to launch simultaneously dozens of simulations of hundreds of nanoseconds.

In order to solve this issue, it would be possible to buy more machines, or to buy faster CPUs and GPUs; however this implies a high financial cost and sometimes better components do not exist yet. Another solution is to reduce calculation time, by increasing the time step. In order to do so, one technique has been developed, called Hydrogen Mass Repartitioning (HMR): its purpose is to rebalance masses between heavy atoms and its hydrogen atoms, making them heavier and thus reducing the bond vibrations frequency. Regularly, hydrogen atoms' masses are increased to 3 atomic units (instead of 1 a.u.), allowing to use a time step of 4 fs (see *Figure D.2*) [40].

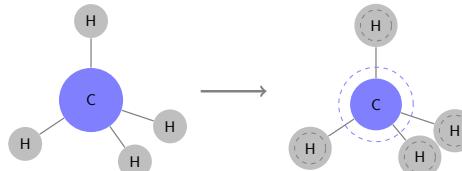


Figure D.2: Scheme representing Hydrogen Mass Repartitioning on a simple example ( $\text{CH}_4$ )

With this method, we can thus expect a doubled performance, and thus a calculation time which would be divided by a factor of 2. But it is first needed to verify that this method does not alter our results before going further. To do so, the same system as in our study was used, with the following protocol (Table D.1):

Experiment	1	2	3	4
Lipids	POPS	POPS	POPS	POPS
Lipids per leaflet	36	36	36	36
Water molecules per lipid	60	60	60	60
Temperature	303.15 K	303.15 K	303.15 K	303.15 K
[NaCl]	0.15 M	0.15 M	0.15 M	0.15 M
HMR	No	No	Yes	Yes
Modifier	Switch	Shift	Switch	Shift
Time step	2 fs	2 fs	4 fs	4 fs

Table D.1: Protocols (Switch = Force-switch and Shift = Potential-shift-Verlet)

The influence of the cut-off modifier was also investigated, as shown in the Table D.1. Indeed, both modifiers behave differently (*see Figure D.3*): the shift modifier will force the potential to get to the desired value (*often 0*) instantaneously, whereas the switch potential will get the potential to the limit value smoothly (*in general using a linear function*). Let  $V(r)$  be a non-bonded potential, then we have:

$$V_{\text{shift}}(r) = \Pi(r - r_c) \times [V(r) - V(r_c)] \text{ with } \Pi(r - r_c) = \begin{cases} 1 & \forall r < r_c \\ 0 & \forall r > r_c \end{cases} \quad (\text{D.1})$$

$$V_{\text{switch}}(r) = \chi(r) \times V(r) \text{ with } \chi(r) = \begin{cases} \frac{1}{r - r_{c,0}} & \forall r < r_{c,0} \\ \frac{r - r_{c,1}}{r_{c,0} - r_{c,1}} & \forall r \in [r_{c,0}, r_{c,1}] \\ 0 & \forall r > r_{c,1} \end{cases} \quad (\text{D.2})$$

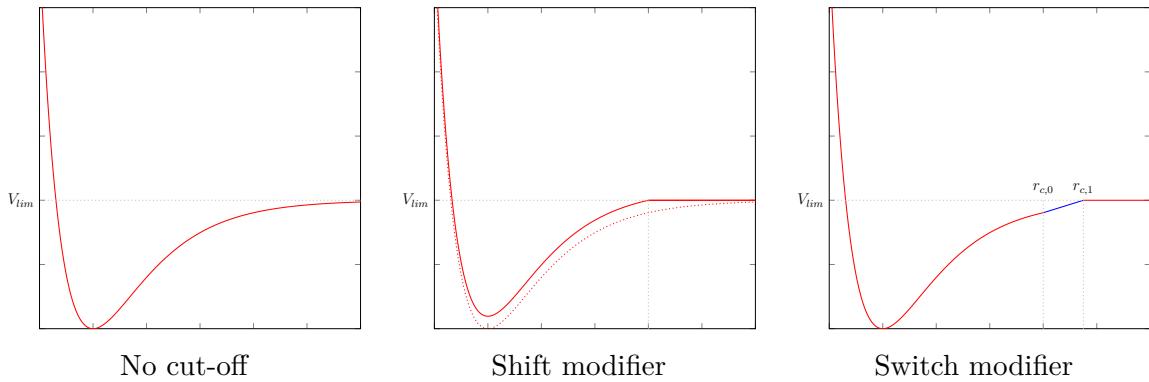


Figure D.3: Types of cut-off modifiers

Having no cut-off was not considered in this study, as it is already known to have a poorer performance compared to the two others suggested above. However, a quantitative study on the effect of one cut-off modifier was not found in litterature.

CHARMM-GUI input files were used for the equilibration without any modification, only production simulation was expanded to make sure to reach convergence; the following results were obtained.

### Performance of calculations

This data is calculated by GROMACS at the end of a simulation, and corresponds to the ratio of simulation time over the time needed to obtain the final result. This performance is obtained during a production run, *i.e.* without any restraint. Table D.2 is obtained

Experiment	1	2	3	4
Performance (ns/day)	68.612	74.627	155.918	167.862

Table D.2: Performance of calculation according to the experiment carried out

With this data, we can observe a doubling in performance when HMR is applied to the system. This seems coherent due to the fact that the time step was doubled. It is also possible to see that the cut-off modifier has an influence of about 8 % on performance, with and without HMR.

### Deuterium order of acyl chains

As the constituents of the membrane are modified, we can expect some changes in system-dependant parameters, such as the deuterium order parameter. We calculate this parameter with MEMBRAINY (GROMACS `gmx order` command does not provide reliable results when using unsaturated chains) and plot the result, we obtain Figure D.4

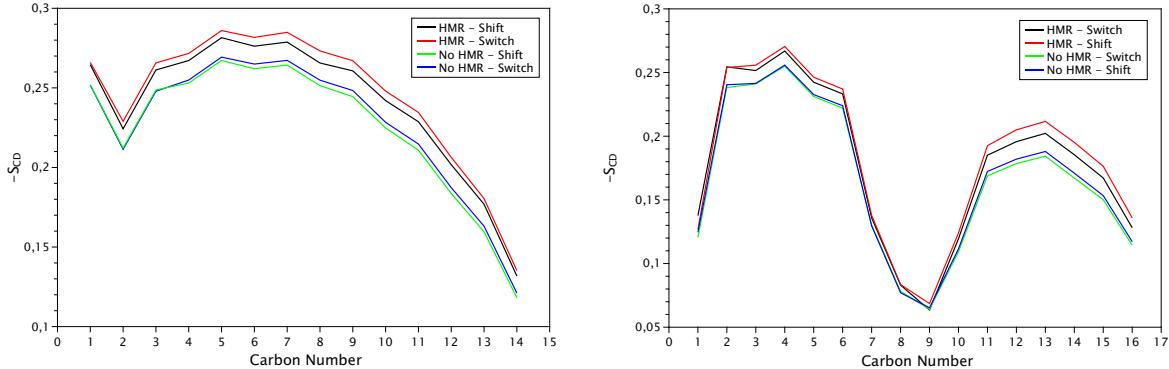


Figure D.4: Deuterium order parameter for saturated (left) and unsaturated (right) acyl chains for every simulation

Values obtained are close to the ones from experiment [41]. We can observe that curves corresponding to the experiments using HMR have slightly higher values ( $\Delta S_{CD} \simeq 0.02$  for high values of  $S_{CD}$ ), but still follow the same tendency. As higher values of the parameter usually mean that lipids are more ordered, this means that lipids are less inclined to be deorganized, its acyl chains being constrained to be more linear. This seems right accordingly with the increase of Hydrogen atoms' masses.

### Area per lipid and other parameters

Other parameters, relative to the membrane itself, were considered: the area per lipid, the membrane thickness, and the diffusion of lipids along the membrane surface (*not considering lipid flip-flop*). Their values for every simulation are shown in Table D.3.

Experiment	1	2	3	4
Area per lipid ( $\text{\AA}^2$ )	55.00	55.21	56.71	52.99
Membrane thickness (nm)	43.29	42.70	42.21	43.97
Lateral diffusion ( $10^{-5} \text{ cm}^2 \cdot \text{s}^{-1}$ )	$1.42 \pm 1.04$	$0.578 \pm 0.623$	$1.35 \pm 0.08$	$0.649 \pm 0.262$

Table D.3: Values of different characteristic parameters according to the experiment carried out

All those values are in a similar range for each parameter (except for experiment 4, which corresponds to HMR+Shift). Thus, these parameters also highlight the fact that applying HMR does not affect the systems.

As a matter of conclusion, this validates the usage of HMR in lipid systems. In order to complete this study, the study of permeation of calcium ions with and without HMR was performed. Unfortunately, simulations were not completed in time to be included in this report.



Chimie Paris



ParisTech

| PSL

— U.PORTO