2018-03-07

Modeling of Biological Membranes by a Coarse-grained Approach

Delgado Magnero, Karelia Hortencia

UNIVERSITY OF CALGARY

Modeling of Biological Membranes by a Coarse-grained Approach

by

Karelia H. Delgado Magnero

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

GRADUATE PROGRAM IN BIOLOGICAL SCIENCES

CALGARY, ALBERTA

MARCH, 2018

Abstract

In this thesis, Martini coarse-grained simulations were performed to investigate the dynamic behavior of lipids and proteins in two different membrane-protein systems. In the first study, I simulated a chromatophore membrane. The model consists of the cytochrome-bc_1 from *Rhodobacter sphaeroides* embedded in an asymmetric distribution of three lipid species. Analyses of a simulation carried out over 40 microseconds highlight specific lipid-protein interactions that are of biological relevance for ongoing research. The second study works towards the modeling of biological membranes at a larger scale. We developed a protocol to build and simulate complex models more representative of physiological membranes, in terms of crowdedness of proteins and lipid diversity. Several eukaryotic plasma membrane models were generated. Analyses of a system consisting of 150 proteins with more than 60 lipid species and simulated for 6 microseconds are discussed.

Keywords: biological membranes, coarse-grained simulation, molecular dynamics simulation, Martini model, large-scale membrane models, lipid-protein interactions.

Acknowledgments

I would like to thank all the people who have enriched my knowledge about simulations and biological processes and who have contributed in some form or another to finalizing my thesis.

I would like to extend a sincere thank you to my supervisor Dr. Peter Tieleman for providing me with challenging and lovely research topics. For giving me the opportunity to attend several conferences and to collaborate with different research groups. Thank you for replying almost instantaneously to my emails and for providing your profound knowledge on every single conversation that we had. Thank you for your kind introduction the day of my defense, I will never forget it.

The work presented in this thesis would not be the same without the participation of Dr. Valentina Corradi and Dr. Gurpreet Singh. Thank you for all your help and guidance during my studies. After our conversations, I always felt motivated to continue with sometimes tricky tasks. Thank you Vale for all the time you dedicated to me during the last few years.

Thank you to Dr. Abhishek Singharoy, it has been a pleasure to work with you. Thank you for being an approachable collaborator, always positive and for being very inspiring at work.

Thank you to former and current members of the Biocomputing group, I will remember the beautiful moments, experiences we had and shared. Thank you Joyce for your help all the times I needed it. Thank you to the examiners of my thesis Dr. Noskov, Dr. MacCallum and Dr. Kusalik. and the support staff at the University of Calgary. Thank you to our collaborators at Dr. Marrink's group in Groningen, where I spent a great time studying, I appreciate your kindness, your welcoming, as well as to Dr. Thewalt's group in Vancouver.

My appreciation goes to the CREATE Training Program in Bionanomachines, the Alberta Innovates-Technology Futures (AITF) Scholarship, and the awards provided by the Department of Biological Sciences, which funded my studies and research.

My heartfelt thanks goes to my family, the primary driving force of my life. Thank you so much to all my friends for your constant support and dedication to me. They all know who they are.
And finally, a special thank you goes to my husband Ale. This stage is completed! We made it!

Table of Contents

List of Tables

List of Figures

List of Symbols, Abbreviations and Nomenclature

Symbol	Definition
MD	Molecular Dynamics
CG	Coarse-grained
ICM	Intracytoplasmic membrane
PG	Phosphatidylglycerol
PE	Phosphatidylethanolamine
PC	Phosphatidylcholine
R.	*Rhodobacter*
ns	nanoseconds
µs	microseconds
nm	nanometer
bc_1	Cytochrome bc_1
ISP	Iron-Sulfur Protein
cyt. c_1	Cytochrome c_1
EGFR	Epidermal growth factor receptor
Kv1.2	Voltage-dependent Shaker potassium channel 1.2
GluA2	AMPA-sensitive glutamate receptor 2
DAT	Dopamine transporter
AQP1	Aquaporin-1
P-gp	P-glycoprotein
Glut1	Glucose transporter
Na,K-ATPase	Sodium, potassium pump

"Walker, there is no path, you make the path as you go"

Antonio Machado

Chapter One: Introduction and Goals

Biological membranes define cellular boundaries and separate the organelles from the cellular interior and from one another. Composed of numerous membrane proteins and different classes of lipids, membranes have essential roles both as structural and functional units of the cell [1, 2]. Disruptions of certain lipid-lipid and lipid-protein interactions can impact cellular processes that are crucial for life and health [3]. Some of these disruptions are responsible for diseases including diabetes, cancer, neurodegenerative disorders, and cardiovascular pathologies. New approaches to combat diseases are focused on generating therapies based on the regulation of lipid-lipid and lipid-protein interactions [1, 4]. Understanding the interplay between these macromolecules is a major focus of ongoing research.

However, the full comprehension of lipid-protein interplay in membranes of living cells is hampered by the lack of experimental methods to describe their dynamic organization and structural heterogeneity with the required complexity and spatiotemporal resolution. In this context, computer simulations, especially using the Martini coarse-graining (CG) model, have become a powerful tool to investigate the dynamic behavior of biological membrane components and their interactions at a near-atomic resolution [5, 6]. The reduction of details and complexity of the CG model allows simulations to reach length and time scales that are inaccessible for an atomistic level of resolution [7, 8].

In this thesis, computer simulations using the Martini CG force field [9] are performed to investigate the dynamic behavior of lipids and proteins in two biological membrane models. The work presented here shows the current trends in molecular simulations of membrane proteins. They include (i) the study of lipid-protein interactions in a membrane with a single protein and (ii) the building and modeling of more complex membranes containing multiple copies of several proteins and a variety of lipid species.

1.1 Lipid-protein interaction study in a membrane with a single protein

The study presented in Chapter 3 is part of a multiscale simulation project. The full photosynthetic organelle of purple bacteria, named chromatophore, was simulated for 500 ns at an atomistic level of detail (**Figure 1-1**) (Singharoy et al., unpublished). The chromatophore vesicle contains numerous copies of five integral membrane proteins embedded in a lipid mixture that mimics a chromatophore membrane [10]. Due to the complexity of the system, which has more than 100 million atoms, lipid-protein interactions could not be evaluated with the required temporal resolution. In Chapter 3, a Martini CG simulation was carried out for 40 microseconds ***to equilibrate the lipids around the bc_1 complex and assess how lipids interact with and organize around the protein over an extended timescale***. From the CG simulation, lipids that preferentially interact with specific regions of the protein were identified, thus adding molecular details to available experimental data on this topic. The results of this long simulation will be used to improve the atomistic chromatophore model by representing the reorganization of lipid around the protein as obtained with CG simulations.

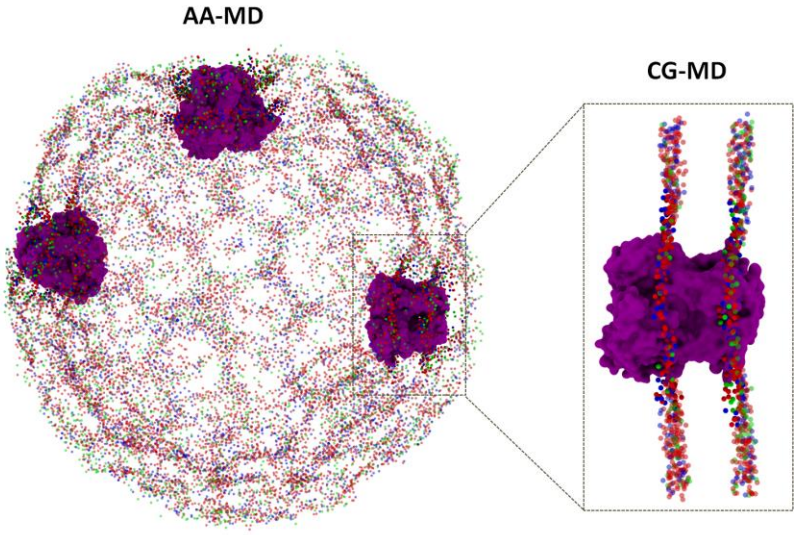

Figure 1-1 Snapshots of the chromatophore membrane. Left panel, the atomistic model of the chromatophore vesicle simulated by Singharoy et al. (unpublished). Right panel, CG model simulated in this thesis to study the lipid reorganization around the bc_1-protein. Head groups of lipids in the chromatophore membrane are shown in spheres with different colors; the bc_1-complexes are depicted in purple with a surface representation.

1.2 Building and simulating complex plasma membrane models

Experimental and computational studies of lipid-protein systems composed of a single membrane protein and a few lipids to solvate the proteins are frequently used to understand the interaction between these biomolecules [11]. However, these models are quite far from membranes of living cells. The work presented in Chapter 4 is part of an extensive project intended to build, simulate, and analyze more realistic eukaryotic plasma membrane models (**Figure 1-2**). The goal of the study presented in this dissertation is ***to develop a protocol to build and simulate eukaryotic plasma membrane models with different proteins and lipid compositions***. Here, I developed and tested a protocol based on the plasmabuilder, an in-house software, and the execution of several steps of molecular dynamics simulations to build large membrane models suitable for simulations. The protein-membrane systems contain up to ten different types of eukaryotic membrane proteins (e.g., transporters, channels, enzymes, and receptors), placed in a membrane model containing various lipid species found in the plasma membrane of eukaryotic cells. In the simulation setup, membrane proteins are present in different ratios, for a total of ca. 150 protein molecules, embedded in more than 60,600 lipids. One of the models was simulated for 6 microseconds. The results provide the first molecular insights into the lateral organization of such a complex system.

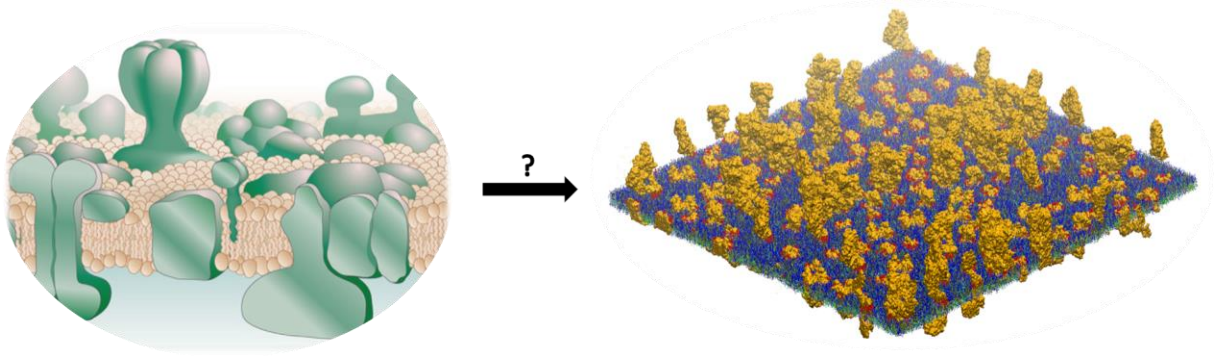

Figure 1-2 Representation of biological membranes. From the text book image to a molecular model ready for simulation. The tool and steps required to generate complex models consisting of different membrane proteins and numerous lipid species are explained in Chapter 4. The left figure was adapted from [12] with permission from Springer Nature.

The thesis is organized in five chapters. The present chapter describes the two main goals of the thesis. Chapter 2 provides the biological and theoretical background of biomembranes and molecular dynamics simulations. Chapter 3 outlines the simulation and analyses performed on the bc_1-system. Chapter 4 describes the first steps in the modeling and analysis of a large-scale eukaryotic plasma membrane and discusses the main challenges faced while building more realistic systems to study lipid-protein interactions. Finally, Chapter 5 summarizes the main results of this thesis.

Chapter Two: Biological and Theoretical Background

2.1 Biological membranes

Lipids and proteins are the main components of biological membranes (**Figure 2.1**). Membrane proteins carry out a large variety of cellular functions and represent approximately 30% of the human proteome [13] and 50% of the existing drugs targets [14]. Considered a key target for therapeutic purposes, these proteins act as transporters or channels to transfer solutes from one side of the membrane to other, as receptors involved in cell signal transduction, or as enzymes to accelerate chemical reactions and facilitate energy conversion. Additionally, membrane proteins can stimulate or mediate the interaction between macromolecular complexes, intracellular compartments or cells.

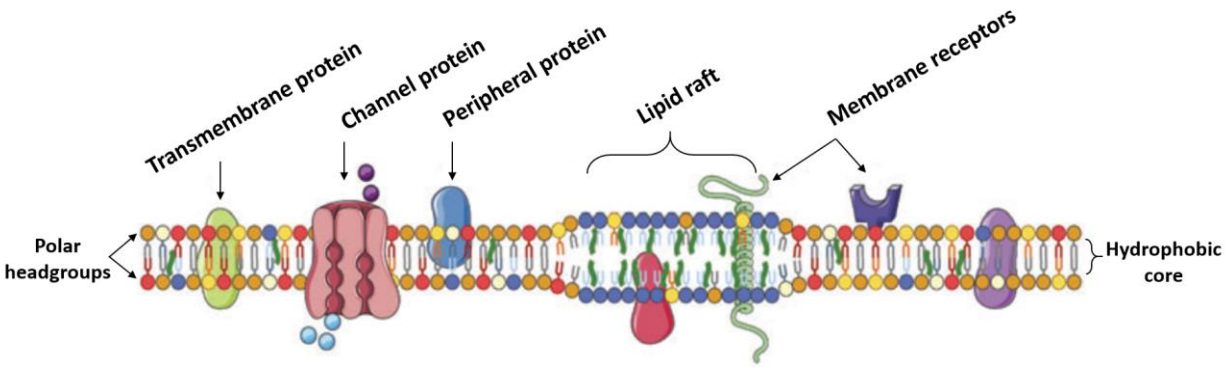

Figure 2-1 Structural heterogeneity of biological membranes. Membranes proteins embedded in a lipid bilayer. Their different color represents the diversity of lipids and proteins. Picture adapted from [2] with permission from John Wiley and Sons.

Membrane proteins are classified as integral (i.e. permanently attached to the membrane - their extraction requires disruption of the membrane) or peripheral (i.e. temporarily attached to the lipid bilayer or other integral proteins - they can be easily removed from the membrane) [15]. Integral membrane proteins may span the whole width of the membrane (polytopic integral proteins) or be associated with only one leaflet of the bilayer (monotopic integral proteins). Membrane proteins are found in two architectures: α-helix bundles or β-barrels. Proteins consisting of α-helix bundles are present in almost all cell

membranes while β-barrels are mainly found in the outer membrane of the bacterial cell and mitochondria [15].

Lipids are amphipathic molecules composed of a hydrophilic polar head and a hydrophobic core (**Figure 2-1**). Their organization in a lamellar structure is the basis of all cellular membranes. Cell membranes are composed of three main classes of lipids, which include glycerophospholipids, sphingolipids, and sterols, like cholesterol in animals (**Figure 2-2**). Lipids differ in their polar headgroup, which defines the lipid class, but mainly in the apolar hydrocarbon chain moiety that forms the hydrophobic core of the bilayer. The apolar hydrocarbon chain can be either saturated (no double bonds between carbons) or unsaturated (presence of one or more carbon-carbon double bonds). Unsaturation introduces a kink in the acyl chain, which increases its shape and volume and therefore affects the lipid packing in the membrane. The chemical nature of the lipid acyl chain such as the number of unsaturations and length of the chains varies depending on the type of membrane and its environment. For instance, when exposed to high or low temperatures the acyl chain composition changes to keep the natural fluidity of the membrane by adapting to the new conditions [15].

Lipid composition varies across organelles, cell types and in disease states, indicating that their function in the membrane and interactions with proteins are not random at all [1, 16-19]. As an example of lipid diversity, sterols and sphingolipids are not present in prokaryotic cells, while the negatively charged lipid phosphatidylglycerol (PG) is found in bacterial and plant membranes but not in eukaryotic membranes other than mitochondria [15]. The diversity of lipid composition and the fact that changes in the structure of one lipid will perturb the lipid-lipid network makes the study of biological membranes where hundreds of lipids contribute to the overall membrane properties such as fluidity, curvature, membrane thickness, lipid domain formation, surface charged and bilayer coupling challenging [20].

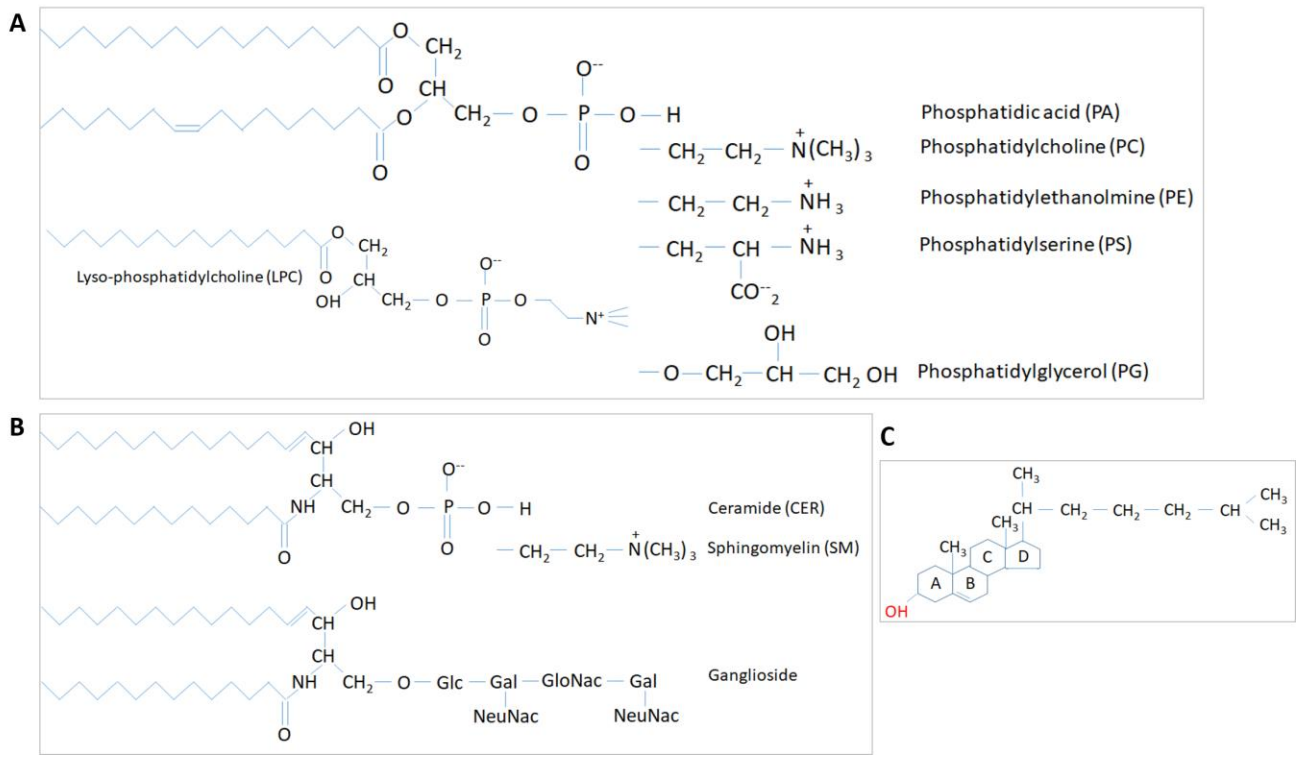

Figure 2-2 Major classes of membrane lipids. A) Glycerophospholipids, **B)** Sphingolipids, **C)** Cholesterol. Examples of lipid polar headgroups are shown.

In a bilayer, lipids undergo rotational, lateral and transverse motion [15]. The rotational motion refers to the lipid movement with respect to its axis, which influences the interactions with neighboring lipids. Lateral motion describes lipid diffusion within a monolayer and takes place when lipids exchange their position with neighboring lipids through Brownian motion. Transverse diffusion, usually called "flip-flop", represents the movement of lipids between the lipid leaflets. Except for some lipids that have very small polar headgroups and can quickly pass through the nonpolar center of the membrane, flip-flop is a slow event that typically only takes place in the presence of specific proteins or due to abnormal deformation of the bilayer [15].

2.2 Lipid-protein interactions

For many years, proteins were the priority in studies of biological membrane and lipids were considered basically a solvent for hydrophobic proteins, but during the last few years, the complexity of the lipid environment started to be recognized. Lipids and proteins are closely related. Changes in the lipid bilayer composition affect the function and structure of membrane proteins [21, 22], while membrane proteins can induce changes in the organization and distribution of lipids in the bilayer. The role of lipids in membrane protein functions has been described as (i) specific, when lipids directly bind to the protein surface, influencing the protein activity and/or stability [23], (ii) non-specific, when physical properties of the membranes such as surface tension, thickness, curvature, and fluidity affect the protein function or (iii) a combination of both [24]. It is essential to choose an accurate membrane model to obtain biologically relevant conclusions about the lipid-protein interplay in the system of interest.

Biological membranes are characterized by their dynamism, the asymmetric distribution of lipids in the bilayer leaflets, the protein crowding in the membrane, the existence of bulky extra-membranous protein domains which may stimulate protein-protein interactions, the irregular thickness of the bilayer due to the membrane lipids perturbation by the protein, and the deformation that a membrane might suffer due to changes in lipid compositions, or protein binding or insertion in the membrane [25].

Experimental methods have been applied to explore the molecular details, composition, and organization of proteins and lipids in biological membranes [26]. High-resolution mass spectrometry, for instance, is an efficient technique that has provided quantitative and structural information about the lipid composition of both eukaryotic plasma membranes and organelle membranes [20, 27]. This method has also been used to identify membrane proteins, such as those found in the plasma membrane of small tissues [28] or chromatophore membranes [29]. Other structural biology methods such as X-ray crystallography, nuclear magnetic resonance (NMR), and cryo-electron microscopy (cryo-EM) have allowed a better understanding of the complexity and heterogeneous nature of cellular membranes [30, 31]. However, to perform these experiments the membrane components need to be purified, or the

sample needs to be frozen, which makes these techniques not suitable for studying and representing the dynamics of the lipid-protein interactions as occurring in membranes of living cells.

Super-resolution fluorescence microscopy techniques have opened new opportunities to investigate the organizational principles of biological membranes [26, 32]. These techniques achieve spatial resolutions below 200 nm, but only a few species of molecules can be observed at the same time, while the rest are invisible to the measurement. The temporal resolution is in the order of milliseconds, and as a consequence, the spatial distribution of membrane molecules, such as the formation of transient clusters, may not be accurately captured. Therefore, although these techniques are valuable tools to study the dynamics of membrane components in living cells, they are limited by constraints in spatial and temporal resolution. In addition, both a high spatial and a high temporal resolution cannot typically be obtained at the same time. Thus, knowledge of the dynamic organization of cellular membranes remains somewhat elusive.

Computational modeling has emerged as a powerful tool to complement experimental methods [33]. In general, this technique uses computational resources to simulate the dynamics of molecular systems. In the case of this thesis, this allows the investigation of the dynamic behavior of membrane components and their interactions at a spatiotemporal resolution that can not be reached by any other biophysical method. The main advantage of molecular dynamics techniques is its ability to reach a length scale small enough to follow the movements of every atom of a system and/or far enough to study the behavior of models with sizes comparable to those reached by current imaging techniques (tens of nm).

2.3 Molecular modeling

Molecular modeling methods study the structure and function of molecules at an atomistic level of details. Based on computational models and their simulations, this technique can be used to support, guide, and explain results from experimental data [34]. To find the most likely molecular configurations of a system, which is associated with the minimum free energy arrangements of the atoms, it is necessary to have a function that describes its potential

energy as a function of the quantity and types of atomic nuclei involved (Z) as well as their coordinates (R) and their electrons (r) as follow:

$$E = E(Z, \overrightarrow{R}, \vec{r}) \quad (2.1)$$

This function, called potential energy surface or hypersurface, describes the way that energy varies with the coordinates. The fundamental problem of molecular modeling methods lies in solving the mathematical expression of such hypersurfaces. Quantum mechanics describes phenomena that occur on an electronic scale. This fact determines the higher accuracy of quantum models in the solution of the potential energy hypersurface of a molecular system. However, the high number of atoms contained in biological systems limits the use of quantum mechanics to model them.

2.3.1 Molecular dynamics method

Molecular dynamics (MD) simulation is a powerful technique widely used to study the motion of atoms and consequently, the dynamics of biological complexes [7, 35]. This method is based on the classic description of particles, where the electronic motions are ignored, and the potential energy of the system is calculated considering only the nuclear positions. A set of rules, called force field, defines how particles associate; then the movement of all the atoms in a system are traced over a given time. Several programs packages are available to perform MD simulations including GROMACS [36], AMBER [37], CHARMM [38] and NAMD [39]. In this thesis, all the MD simulations were performed using the GROMACS software package [36, 40, 41].

A force field is chosen depending on the components of the system and the question to be addressed. Several force fields are available with different properties and parameters [42]. Commonly, in all force fields interactions between atoms are classified as bonded or non-bonded interactions. The former includes the interactions between atoms that are connected between one, two and three bonds and the latter describes the interactions between any pair of atoms that are not connected and are within a given cut-off radius. The bonded interactions comprise bond, angles and dihedrals terms, while nonbonded interactions consist of Coulombic terms for charged particles and terms for repulsion and attraction interactions described by the Lennard-Jones potential. A typical potential energy function is described as:

$$\boldsymbol{V} = \sum_{bonds} \frac{1}{2} k_{ij}^{b} \left(r_{ij} - b_{ij}^{0}\right)^{2} + \sum_{angles} \frac{1}{2} k_{ijk}^{\theta} \left(\theta_{ijk} - \theta_{ijk}^{0}\right)^{2} + \sum_{dihedrals} k^{\varphi} \left(1 + \cos\left(n(\varphi - \varphi^{0})\right)\right) + \sum_{\substack{i<j \\ LJ}} \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \sum_{\substack{i<j \\ Coulomb}} \frac{q_i \, q_j}{4\pi\epsilon_0 \, r_{ij}} \quad (2\text{-}2)$$

Here k^{b}, k^{θ}, k^{φ} denote the force constants for bonds, angles, and dihedrals, respectively, r_{ij}, is the distance between atoms *i* and *j*, q_i is the partial charge on atom *i*, A_{ij} and B_{ij} are the Lennard-Jones parameters, n is the dihedral multiplicity and b^{0}, θ^{0}, φ^{0} are equilibrium values for the bond lengths, angles and dihedral angles.

In the molecular dynamics method, all the molecules of the system are put into a virtual box, and the force on every atom is calculated by the negative gradient of the potential energy function:

$$\mathrm{F_i} = -\frac{\delta \mathrm{V}}{\delta \mathrm{r_i}} \quad (2\text{-}3)$$

Given the analytical form of the potential energy function of the system, the initial position of the atoms and the atomic masses (m), the new position of the atoms are calculated solving the classical equations of motion ($F = ma$). Several algorithms based on finite difference methods are available for integrating the equations of motion. In these methods, the integration is broken into small stages separated by a fixed time step Δt. The forces acting on each atom are assumed to be constant during the time interval. The leap-frog algorithm is a commonly used numerical integrator for this purpose. Once the force acting on each particle is calculated, $F(t)$, the velocities (v) and positions (r) of the atoms are determined as follow:

$$v\left(t + \tfrac{1}{2}\Delta\mathrm{t}\right) = v\left(t - \tfrac{1}{2}\Delta\mathrm{t}\right) + \frac{\Delta\mathrm{t}}{m} F(t) \quad (2\text{-}4)$$

$$r(t + \Delta\mathrm{t}) = r(t) + \Delta t v(t + \tfrac{1}{2}\Delta t) \quad (2\text{-}5)$$

Before performing molecular dynamics simulation, it is usually recommended to minimize the energy of the system, to avoid large forces that can cause the failure of the simulation. Steepest descent is a well-known minimization method that consists of moving each

atom a short distance in the direction of a decrease in energy [40]. The minimization steps are performed until reaching a minimum energy arrangement within some specified tolerance. Then, using the final structure from the minimization step, the positions and velocities of atoms as a function of time are obtained by solving the classical equations of motion iteratively for each particle with very small-time steps, commonly in the order of femtoseconds for all-atoms simulations.

Simulations in this thesis were carried out with a constant number of atoms, constant temperature and constant pressure (NPT ensemble). The temperature depends on the kinetic energies of all the atoms in the system. During the integration, cut-off effects and numerical rounding errors can lead to drift in the kinetic energy of all atoms. As a result, the system tends to heat up during the simulation. In the same way, the total pressure of a system depends on the forces exerted by the atoms and their positions, which determines the size of the simulation box and whether the simulation box expands or contracts. A common method used to ensure that the average temperature and pressure of the system are maintained during the simulation is to couple the system to an "external bath" that is fixed at the desired temperature and pressure [43].

For temperature coupling, both Berendsen thermostat [43] and V-rescale method [44] were used in this thesis. For the Berendsen algorithm, the system is weakly coupled to an external heat bath with the desired temperature T_{bath}. When the temperature T, calculated from the system's total kinetic energy, deviates from the desired temperature, it is corrected by

$$\frac{\mathrm{dT}}{\mathrm{dt}} = \frac{1}{\tau_t}(\mathrm{T_{bath}} - \mathrm{T}), \qquad (2\text{-}6)$$

where the time constant or coupling parameter τ_t determines the rate of this correction.
Then, the velocities are scaled at each time step by the scaling factor

$$\lambda^2 = 1 + \frac{\delta t}{\tau_t}\left(\frac{\mathrm{T_{bath}}}{\mathrm{T}} - 1\right), \qquad (2\text{-}7)$$

where δt is the simulation time step. By rescaling the velocities, the average temperature is maintained around the desired temperature value [40, 43].

Semi-isotropic pressure coupling, where the X,Y dimensions of the box are scaled separately from the Z dimension, was used for all the simulations presented in this thesis. The total pressure is adjusted by scaling the simulation box size. Like the Berendsen thermostat, a Berendsen barostat was used to maintain the desired pressure [43]

$$\frac{dP}{dt} = \frac{1}{\tau_p}(P_{bath} - P), \quad (2\text{-}8)$$

P_{bath} is the desired pressure and τ_p the coupling parameter.
Then, the volume of the box is scaled at each time step by the scaling factor

$$\lambda = 1 - \beta_T \frac{\delta t}{\tau_p}(P_{bath} - P), \quad (2\text{-}9)$$

where δt is the simulation time step, β_T is the isothermal compressibility, and the positions of all the atoms are scaled by a factor $\lambda^{\frac{1}{3}}$ [43].

After scaling the velocities and coordinates of the atoms, the equations of motion are updated with the new values to maintain the desired temperature and pressure. All these procedures are repeated for a given number of steps, and the positions of all atoms saved at regular intervals, generating the trajectory of the system which can be viewed and analyzed.

Several approximations are made in the molecular dynamics technique. For instance, because electronic properties are averaged out in classical mechanics, the breaking and forming of chemical bonds cannot be studied. Therefore, the force fields used to describe the interactions are not perfect and depend on the accuracy of the chosen parameters. Nonbonded interactions are only calculated for the atoms within a selected cut-off distance and not until infinity, to maintain an acceptable computational effort. Simulations are limited by the number of atoms present in the box, so they reproduce only small patches of actual systems. To allow a proper integration of Newton's equations of motions, very small-time steps need to be used.

This fact makes the integration computationally expensive, which limits the amount of sampling that can be performed.

Reaching adequate sampling in a simulation is one of the main concerns in molecular dynamics [45]. The sampling time should be longer than the time scale on which the biological phenomenon of interest occurs. This allows a more accurate sampling of the phase space and therefore, to access and explore the configurations that are relevant to the biological event being studied. However, despite continuing increases in computational power and the availability of high-performance clusters, sufficient simulation time is not always achievable when performing equilibrium MD simulations. Many biological processes take place at time and scales that are not accessible or not feasible to study by current atomistic simulations. In some cases, simpler models creating by coarse-graining methods may allow for a better sampling of the system, increasing both the time and length scales that are accessible for a given biomolecular system.

2.3.2 Martini model and force field

In the study of biological systems, atomic details are not always necessary to describe the phenomenon of interest, which has encouraged the development of coarse-graining (CG) force fields [35]. In CG simulations, atoms are grouped into functional groups, reducing the number of particles in the system, and defining a smoother energy landscape which leads to faster dynamics and the use of larger time steps. Therefore, the computational effort required to perform simulations is reduced and both the size and simulation time of the system can be simultaneously increased.

Martini is a coarse-grained force field used in the simulation of a broad range of biological systems [8, 46]. In the Martini model, on average four non-hydrogen atoms plus accompanying hydrogens are represented by one interaction site or CG bead (**Figure 2-3**) [9]. For instance, four real water molecules are represented by a CG bead. Ions are also described by a CG bead, which represents both the ions and their first hydration shell. To represent ring-like groups such as found in cholesterol and aromatic amino-acids the four-to-one mapping is

too coarse; therefore up to two-to-one non-hydrogen atoms are mapped into one CG bead, providing a higher resolution of these molecules.

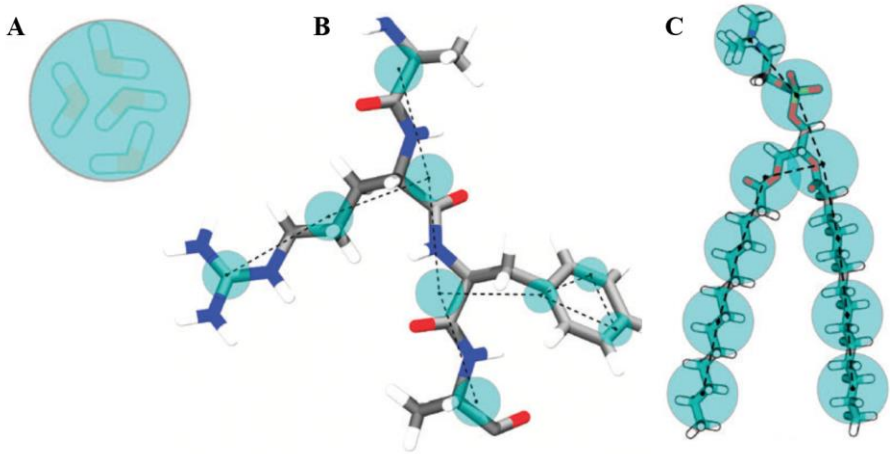

Figure 2-3 Example mappings of some molecules to the Martini model : A) Standard water molecule, B) Peptide including a phenylalanine residue, C) Phospholipid. Picture adapted from (Marrink and Tieleman, 2013) [8].

In the Martini model, the chemical nature of the mapped structure is described by four main bead types: polar, non-polar, apolar and charged. Within the same type, subtypes describe the capability of the groups to form hydrogen bonds and their degree of polarity. The Lennard-Jones potential describes nonbonded interactions between beads. The well-depth ε of the LJ function is adjusted to obtain different interaction strengths, to simulate both strongly polar as well as hydrophobic interactions. As described in the potential energy function, charged groups interact through a modified Coulombic energy and bonded interactions are described by the standard functions for bonds, angles, and dihedrals.

It is important to note that in the Martini model, the absence of hydrogen bonds prevents the realistic folding of proteins. In practice, the secondary structure of proteins is fixed and cannot change during the simulation. An elastic network approach, called Elnedyn, has been developed to improve the stability of the tertiary structure of proteins by keeping the protein close either to a particular native state or to the initial configuration of the atomistic protein used to generate de CG model [47].

The potential use of CG-MD simulations to study membrane system has been examined in several reviews [8, 48]. Hedger et al. summarize examples of CG and atomistic simulations of single membrane protein systems [11]. The simulations enabled the identification of lipid binding and/or interaction sites on integral membrane proteins that reproduce experimental findings such as cholesterol binding on G-protein coupled receptors (GPCRs) and phosphatidylinositol 4,5 bisphosphate (PIP_2) binding sites on receptors, ion channels, and transporter proteins [11]. Other studies confirm the successful role of CG simulation in accurately predicting lipid binding sites seen in X-ray structures such as the case of the cardiolipin binding sites found on cytochrome bc_1 of mitochondrial membranes [49]. Recent studies show the feasibility of using large-scale CG-MD simulations to investigate lipid-lipid and lipid-protein interactions at the near-atomistic level of resolution [6, 7].

Chapter Three: Cytochrome bc_1 in a chromatophore membrane

3.1 Introduction

Photosynthesis is performed by some bacteria, green algae and plants to capture, transfer, and convert solar energy into chemical energy, which drives the synthesis of ATP for fueling cellular metabolism. The energy conversion process is regulated by the cooperation of hundreds of proteins, lipids and other molecules. Investigations of natural light-harvesting systems have provided considerable insights for the development of bio-hybrid and artificial solar devices, improving the harvesting of energy from sunlight to produce solar power [50, 51].

Purple bacteria and their chromatophores

Purple bacteria are likely the earth's most primitive and simplest photosynthetic organisms. Their photosynthetic apparatus, also known as a chromatophore, has been widely studied and characterized [52]. Chromatophores are formed by intracytoplasmic membranes (ICM) which house and leverage the energy production reactions accomplished by an assembly of pigment-protein photosynthetic complexes (**Figure 3-1**). The overcrowding of chromatophores in the cell interior occurs to harvest sunlight when the purple bacteria experience low illumination levels; this is considered an adaptation to the light starvation regime of these bacteria [53]. Chromatophores are composed mainly of five integral membrane proteins embedded in a lipid bilayer (**Figure 3-1**) and contain cofactors that interact with the photosynthetic proteins and participate in the collection and transfer of the solar energy.

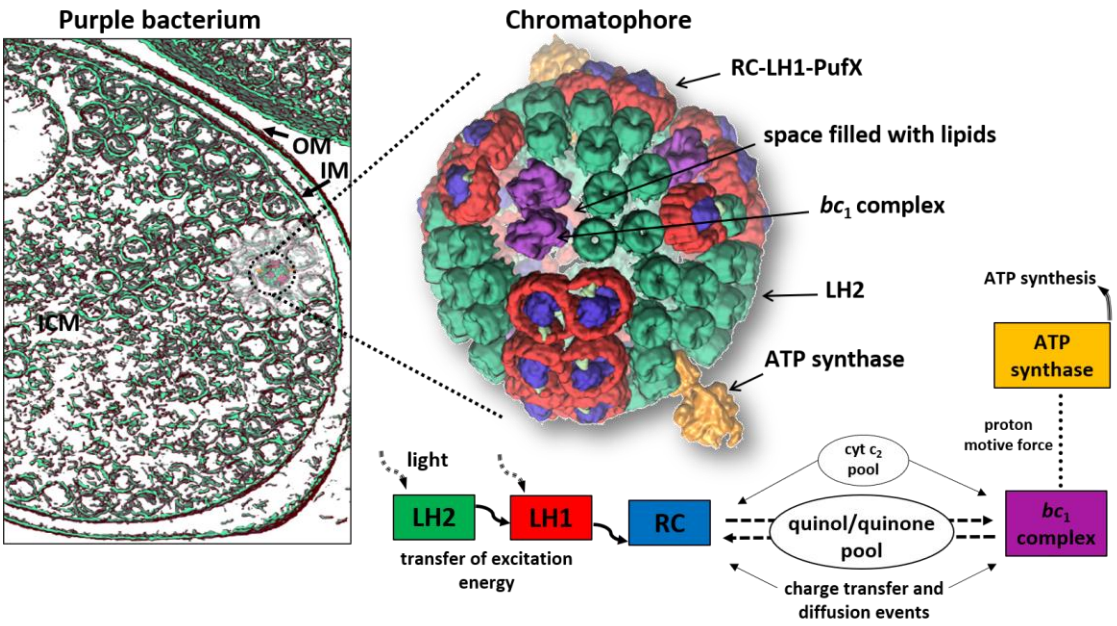

Figure 3-1 Chromatophore for purple bacterium *R. sphaeroides*. The left panel shows a cross-section of a purple bacterium. The intracytoplasmic space (ICM), outer membrane (OM) and inner membrane (IM) are emphasized. Right panel: close-up of one of the chromatophore vesicles. Proteins are shown in different colors and include: light harvesting complexes I and II (LH1 and LH2), reaction center (RC), cytochrome bc_1 (bc_1 complex), ATP synthase. In some species an additional protein called PufX is also expressed. The lower panel summarizes the energy conversion processes carried out in the chromatophore vesicle. Figures modified from [53], used with permission of Dr. Abhishek Singharoy.

A variety of quantum and classical processes that cover a broad range of time and length scales takes place during photosynthesis (**Figure 3-1**, lower panel) [53, 54]. The photosynthetic cycle starts when photons are absorbed by light-harvesting pigments (bacteriochlorophylls and carotenoids) found in LH2 and LH1 proteins [54]. After light absorption, LH2 transfers its excitation energy to LH1, which passes the excitation energy to the RC. In the RC, the electron transfer across the membrane is initiated, two protons and two electrons from the cytoplasm are transferred to a mobile charge carrier quinone to form quinols. Subsequently, the quinols diffuse through the lipid bilayer to the bc_1 complex. Once in bc_1, protons and electrons are released into the periplasm, which results in the formation of an electrochemical proton gradient. In the process, the soluble charge carrier cytochrome 2 (cyt c_2) returns electrons from the bc_1 complex to the RC, which closes the cycle. The produced proton

gradient is the proton motive force that powers the synthesis of ATP by ATP synthase. The right lower panel of **Figure 3-1** shows a schematic representation of these energy conversion processes. Further information about energy harvesting and conversion steps in a chromatophore model can be found in reference [53]. The physiological organization of photosynthetic proteins in the membrane is fundamental for efficient light capture and energy transfer [55, 56]. Several studies have been performed on the ICMs of various species of purple bacteria to explore the morphologies of chromatophores (see review [57]). Interestingly, the shape varies among species. Small vesicles are found in *Rhodobacter* (*R.*) *sphaeroides*, flat lamellar discs in *Phaeospirillum molischianum* and tubular structures have been identified in certain *R. sphaeroides* mutants [58-60]. Taking advantage of the different shapes, the ICMs of purple bacterial species are also models for experimental and computational methods in the investigation of organelle membrane architecture.

The architecture of the chromatophores found in *R. sphaeroides* has been investigated by several imaging techniques, and currently, its structure is known at near-atomic detail (**Figure 3-1**) [29, 61]. The spherical structure of these chromatophores is caused by the membrane-bending properties and the arrangement of LH2 and LH1-RC-PufX complexes in the vesicle, which cause local curvature and are found in convex regions of the membrane [29, 62-64]. However, the bc_1 complex along with that of the ATP synthase has not been visualized in any experimental techniques, and therefore, their location in the chromatophore membrane is still quite unclear. Recent experimental studies have proposed that bc_1 resides in the concave and relatively flat 'neck' regions of the chromatophore membrane. It has also been suggested that the area around bc_1 complexes is enriched in lipids, while the rest of the photosynthetic proteins are found closer packed in the membrane, leaving less space for lipids between them (**Figure 3-1**) [29, 61].

Chandler et al. [62, 64] performed all-atom simulations of 20 ns to study the curvature properties of LH2, LH1, bc_1 complex, and LH1-RC-PUfX dimer, embedded in a model of the chromatophore membrane. It was suggested that the bc_1 complex might satisfy the curvature of the chromatophore bilayer without inducing any curvature by itself. In contrast, the other

complexes were capable of inducing curvature in the membrane, although the extent of the curvature showed a dependency on protein packing. The authors pointed out that the distribution of charged residues on the protein surface plays an important role in the membrane shape induced by the proteins. However, the short simulation time of these studies limits the amount of sampling to define the lipid environment around the proteins accurately.

Despite the uncertainty about the precise location and protein arrangement around the bc_1 complex and ATP synthase in chromatophores of purple bacteria, Sener and co-workers [53] built an atomistic model to study the functional processes that are carried out in this photosynthetic vesicle [53, 54]. Consisting of a spherical membrane with an inner radius of 30 nm, the model has a total of ca. 100 million atoms, including proteins, co-factors, quinones, quinols, lipids, ions and water (**Figure 3-1**). Firstly, a protein-only model of the chromatophore vesicle was built based on AFM, electron microscopy and mass spectrometric data [29]. After performing several other simulation steps, lipids were randomly inserted to reach the asymmetric composition that mimics a real chromatophore membrane. It is important to highlight that this is the first all-atom representation of an organelle and probably the largest membrane system model to date.

In combination with experimental findings, this structural model has been used to understand further the biochemical processes that take place in the photosynthetic organelle, including light absorption, energy transfer, electron flow, the generation of the proton gradient, and an estimate of ATP production efficiency that is determined by these processes [29]. Kinetic models of the chromatophore showed that its architecture and protein arrangements are optimized for harvesting sunlight and continuously producing ATP at low illumination levels. In general, numerous studies have been focused on defining the overall shape and protein organization of chromatophores in purple bacteria as well as in the energy conversion processes that take place in this pseudo-organelle [29, 53, 54, 61, 62, 65].

Recently, the structural stability of the chromatophore model was tested through all-atom simulations carried out by Singharoy et al. (unpublished). After 500 ns of simulation time, membrane deformations were observed only in the proximity of the four bc_1 complexes

present in the model (**Figure 3-2**). Nevertheless, the molecular details of these deformations remain unclear.

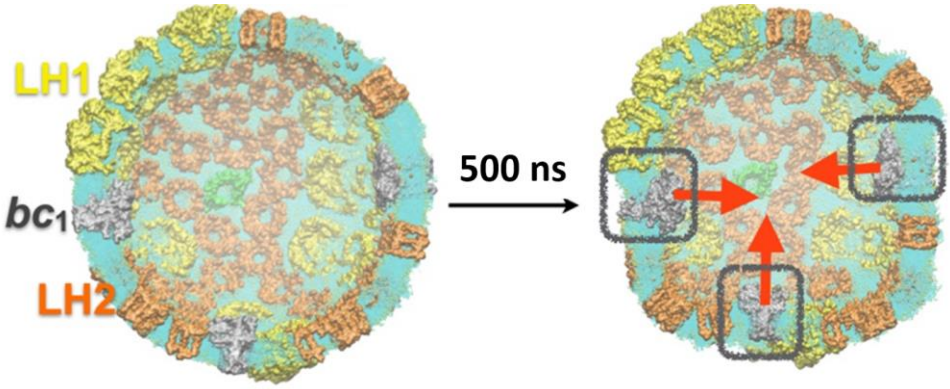

Figure 3-2 Deformation of the chromatophore membrane after 500 ns. Frames of the chromatophore simulation at the beginning and after 500 ns simulation. Proteins are shown in different colors and lipid in cyan. The co-factors, water, electron carries, and ions are not shown. The figure is used with permission of Dr. Abhishek Singharoy.

Cytochrome bc_1 complex

The cytochrome bc_1 complex is one of the fundamental components of the respiratory and photosynthetic electron transfer chains. All the bc_1 complexes, although under different names and in different organisms, share a common core structure [66]. Particularly, the cytochrome bc_1 of *R. sphaeroides* is a dimer that contains three subunits per monomer consisting of cytochrome c_1, cytochrome *b*, and the iron-sulfur protein subunit (**Figure 3-3**). The cytochrome *b* includes two heme groups (b_H and b_L), the quinol oxidation site (Q_0 – near the periplasmic side of the membrane) and the quinol reduction site (Q_i – near the cytoplasmic side of the membrane). The iron-sulfur protein contains a 2Fe-2S group, which transfers one electron onto the heme *c* group of the cytochrome c_1. Then, the electron from the heme *c* group is transferred to the soluble mobile electron carrier, cytochrome c_2.

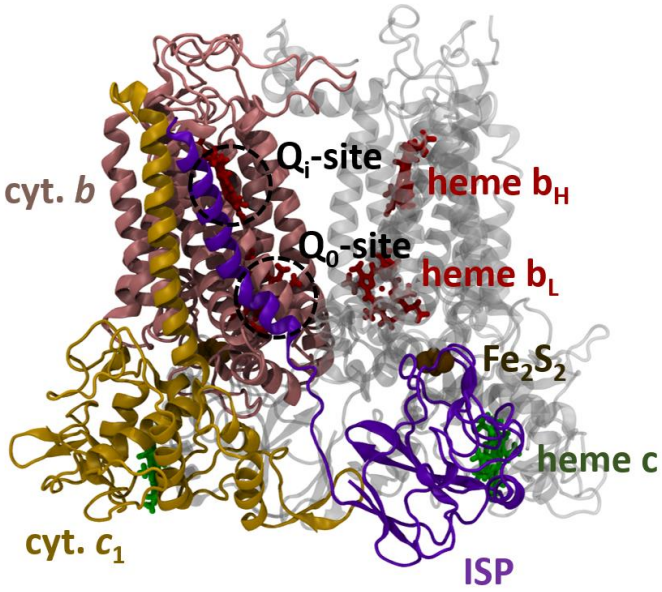

Figure 3-3 Structure of cytochrome bc_1 complex from *R. Sphaeroides*. The atomistic structure of the bc_1 dimer is shown in cartoon representation. One monomer shows the three proteins, cytochrome *b* (cyt. *b*), cytochrome c_1 (cyt. c_1) and Iron-Sulfur Protein (ISP) in pink, yellow and purple, respectively. The cofactors present in the crystal structures are shown as sticks (heme groups) and spheres (Fe_2S_2). The quinol reduction (Q_i) and oxidation (Q_0) sites are illustrated.

Protein-lipid interplay on R. sphaeroides chromatophore

The intimate relationship between photosynthetic proteins such as the cytochrome b_6f complex, the photosystem II complex, the reaction centers and the lipids of the thylakoid membrane has been well established [67-70]. However, the distribution of lipids around the photosynthetic proteins of *R. sphaeroides* remains ambiguous, despite this being a critical aspect to achieve a more realistic model of the chromatophore. The chromatophore membrane of purple bacteria is made up of different lipids asymmetrically distributed across the bilayer. The majority of the lipids are made up of phosphatidylglycerol (PG), phosphatidylethanolamine (PE), and phosphatidylcholine (PC) [10, 71, 72]. PG carries a negative charge, PC is neutral, and both are bilayer forming lipids. PE is a non-bilayer forming lipid and the most abundant lipid in the chromatophore membrane [71].

Lipid-protein interactions are more correctly studied in lipid environments that reproduce realistic protein crowding and packing, as seen in the chromatophore model

(**Figure 3-2**). However, the high number of atoms in these models makes the simulations computationally expensive, limiting the timescale of the simulation and, as a consequence, hampering an accurate investigation of lipid-protein interactions. To overcome this challenge, a common approach to studying the lipid organization around a given membrane protein is to simulate only the protein of interest embedded in a membrane composed of the most important lipids in terms of protein function, structure and lipid composition.

The work discussed in this thesis is part of an extensive project focused on describing the distribution of lipids around all the photosynthetic proteins of purple bacteria using the Martini CG model. The resulting data on lipid-protein interactions will be used to improve the atomistic chromatophore model by representing a more accurate distribution of lipids around the proteins. Specifically, this chapter concerns the study of the organization of lipids around the bc_1 complex from purple bacterium *R. sphaeroides* in a membrane containing a single bc_1 protein using Martini CG simulations.

3.2 Methods

The structure of the cytochrome bc_1 complex used in this study was retrieved from the *Protein Data Bank* (PDB), PDB ID 2QJY [73]. Cofactors such as the hemes and iron-sulfur molecules found in the bc_1 crystal structure were not included in the CG simulation as they are not expected to affect the protein-lipid interactions. Furthermore, an elastic network approach described below was used to stabilize the protein fold and compensate for the absence of those molecules [47].

The Martinize script was used to convert the atomistic bc_1 complex to a CG Martini model [74]. The ElNeDyn elastic network model was applied to particles within a cut-off value of 0.9 nm using a force constant of 500 kJ mol^{-1} nm^{-2} [47]. Afterward, the protein was embedded in a chromatophore membrane model using the *Insane* script [75], preserving the orientation that the protein adopts in the full chromatophore model (Singharoy et al., in preparation) (**Figure 3-2**). The composition of the chromatophore membrane consists of an asymmetric distribution of lipids in the cytoplasmic side (upper leaflet: 22 % POPG, 22% POPC,

56 % POPE) and periplasmic side (lower leaflet: 10% POPG, 24% POPC, 66% POPE). Protein and lipid parameters were modeled according to Martini 2.2 and Martini 2.0, respectively. The system was solvated using the standard MARTINI water model along with ions to an approximate concentration of 0.15 M NaCl. The CG system contains 57,354 particles in total.

3.2.1 Simulation Details

CG molecular dynamics simulations were carried out under NPT conditions with the GROMACS software package (version 5.1.1) [40] (http://www.gromacs.org/). Periodic boundary conditions and semi-isotropic pressure coupling were applied. Energy minimization was performed using 1000 steps of the steepest descent algorithm, followed by equilibration steps, first with position restraints applied to all protein beads, and then to the backbone beads only. The simulation was performed with a 20 fs time step. The temperature at 310 K and pressure 1 bar were maintained during the production run using v-rescale thermostat [44] and Berendsen barostat [43], respectively, with a time constant for coupling of 1.0 ps for the temperature and 4.0 ps for the pressure. The neighbor list was updated using the Verlet neighbor search algorithm. Non-bonded Coulomb interactions were treated using a reaction-field with a cut-off of 1.1 nm. Lennard-Jones interactions were calculated using a cut-off at 1.1 nm, with the potential shifted to zero at the cut-off using potential-shift-Verlet. The system was simulated for 40 μs of production run with position restraints applied to the backbone beads of the proteins using a force constant of 1 kJ mol^{-1} nm^{-2}.

3.2.2 Analysis

The last 10 μs of the 40 μs Martini simulation was used with frames saved every 500 ps for the analysis of lipid dynamics. The molecular graphics viewer VMD [76] was used to visualize the system.

Lipid distribution

To analyze the lipid distribution around the protein as a function of time, gmx_select from the GROMACS package (version 5.1.1) (http://www.gromacs.org/) was used. Lipids were

represented by their phosphate groups, and the number of lipids within a cut-off distance of 0.7 nm was plotted over the simulation time [77].

Density and membrane thickness 2D maps were calculated using g_mydensity and g_thickness [78]. In all cases, the plane of the membrane (xy) was divided in 10,000 bins, with grid cells of ca. 0.2 nm x 0.2 nm and the quantity of interest was averaged over time in each grid cell. Membrane thickness is defined as the distance between the average positions of the lipid phosphate groups (PO4 bead) in the two leaflets. The positions of the PO4 beads were used to represent the lipids in the partial density and membrane thickness plots. Density analyses were averaged over the last 5 μs, while membrane thickness was averaged over three window times of 5 μs each, (0 μs – 5 μs, 20 μs – 25 μs, 35 μs – 40 μs). The density graphs show the normalized density, per lipid type, by dividing the density in each grid cell by the average mass density of the PO4 beads.

D-E index

The depletion-enrichment (D-E) index of a given lipid type was calculated as described in [79]. The analysis was performed using different cut-off distances of 0.7 nm, 1.4 nm and 2.1 nm around the protein. For a type of lipid *L*, the ratio of lipid *L* within a given cut-off x, termed as $Ratio\ (L)_x$, and the ratio of the lipid *L* with respect to the bulk $Ratio\ (L)_{bulk}$ are described as:

$$Ratio\ (L)_x = \frac{(\#\ L)_x}{(Total\ \#\ Lipid)_x} \qquad Ratio\ (L)_{bulk} = \frac{Total\ \#\ (L)}{(Total\ \#\ Lipid)}$$

Then the enrichment of the lipid *L* for a given cut-off *x* is calculated from the following ratio:

$$Enrichment\ (L) = \frac{Ratio\ (L)_x}{Ratio\ (L)_{bulk}}$$

The enrichment index was calculated over three-time windows during the last five microseconds of the trajectory (35 μs – 36 μs, 37 μs – 38 μs, 39 μs – 40 μs). The corresponding standard deviation was also calculated.

Protein contact residues

To determine the residues that are in contact with each of the lipid types (PG, PE, PC), the script get_contacts.py written by Dr. Manuel N. Melo (http://www.itqb.unl.pt/labs/multiscale-modeling/people) was used. The script calculates the number of frames in which a residue is in contact with a specific lipid type within a given cut-off. Lipids were represented by their phosphate groups. Residues that maintained contacts with lipids within a distance less than 0.7 nm over 50% of the frames in the last 10 μs of the simulation were identified. Only residues that were found in both monomers were selected for further analysis.

3.3 Results

The bc_1 complex of *R. sphaeroides* was simulated for 40 μs embedded in a chromatophore membrane [10, 71, 72] with an asymmetric distribution of PG, PE and PC lipids in the membrane leaflets (**Figure 3-4**). In the following sections, we analyze the distribution and enrichment of lipids around the protein, discuss the pattern of the local membrane thickness and identify the primary residues that contribute to protein-lipid interactions and might be associated with the overall shape of the membrane induced by the protein.

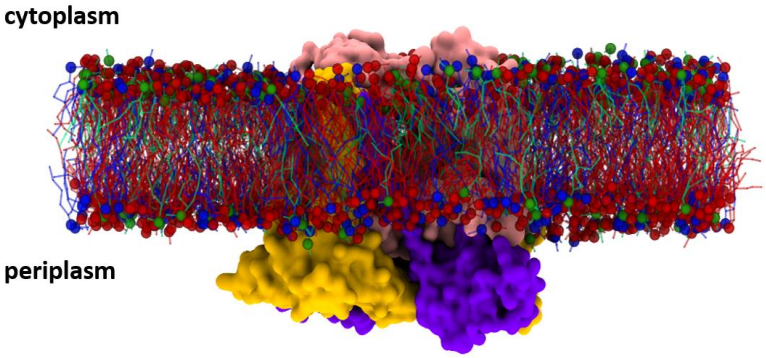

Figure 3-4 System of bc_1 complex in CG representation. The protein is inserted in a bilayer composed of lipids: POPG (green), POPE (red), POPC (blue). Lipid headgroups and tails are shown in van der Waals and lines representation, respectively. Protein subunits are shown in surface representation: cytochrome *b* (pink), cytochrome c_1 (yellow), ISP (purple). Water and ions are not shown for clarity.

3.3.1 Lipid distribution in the bc_1 complex vicinity

Lipid count

To evaluate the lipid composition of the first lipid shell, the number of PG, PE, and PC lipids within 0.7 nm cut-off from the protein was calculated as a function of time, in the upper and lower leaflets (**Figure 3-5**). Overall, the count of lipids does not show considerable variation relative to the starting point. Fluctuations around the average values are stable from early on in the simulation. PE lipids are the most abundant in the first lipid shell around the protein in both leaflets.

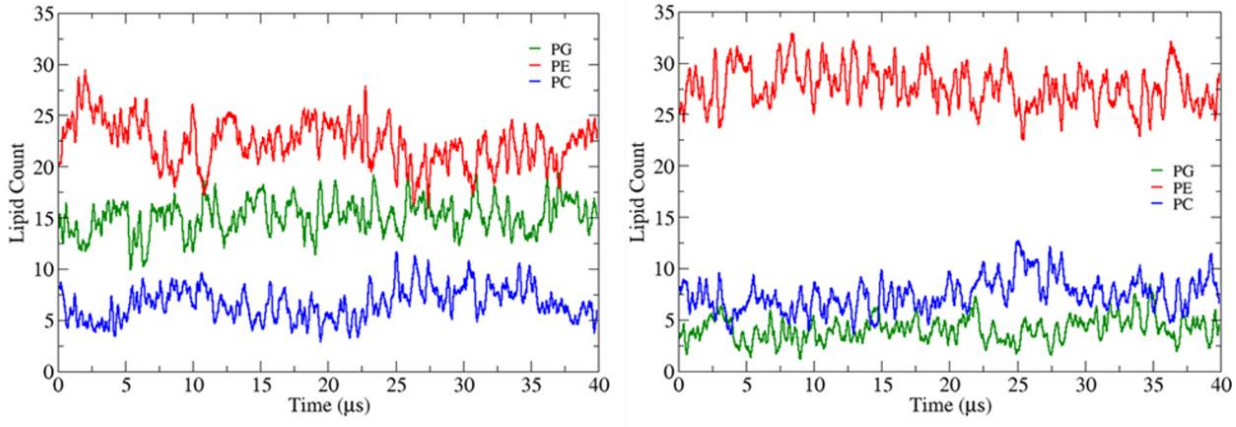

Figure 3-5 Lipid environment around bc_1 complex. Lipid count within 0.7 nm cut-off from the protein calculated as a function of the simulation time. Left panel, upper leaflet: PE (22±3), PG (16±3), PC (6±3). Right panel, lower leaflet: PE (27±3), PG (4±2), POPC (8±2). The number of lipids and the standard deviation are shown in parenthesis. Only the PO4 beads of the lipids were considered in this analysis.

2D density maps

To better characterize the lipid organization around the protein, the lateral density distribution of the lipid PO4 beads was averaged over the last 5 μs of the trajectory (**Figure 3-6**). Considering the asymmetry of both the protein structure and the lipid distribution across the bilayer, the analysis was performed independently for the upper and lower leaflets (**Figure 3-6**). A high density of PG and PE lipids around the protein is observed in the upper leaflet. PG molecules are localized surrounding the complex, while PE lipids prefer the transmembrane

inner cavities at the dimer interface. In fact, due to the localization of PG lipids, they seem to be sharing or interchanging with PE lipids in the inner protein cavities. Additionally, one PG molecule penetrates inside the protein early in the simulation, which explains the high-density spot observed in the map (**Figure 3-6**, PG-upper). A substantial density of PC lipids is not observed in the upper leaflet around the protein, and only a small region close to one of the monomers showed a high density of this lipid.

In the lower leaflet, enriched PG lipids are confined to small regions near the protein surface (**Figure 3-6**). The density of PE lipids is localized in small regions around the protein, and PE depletion is observed. A low density of PC lipids is detected near the protein in the lower leaflet.

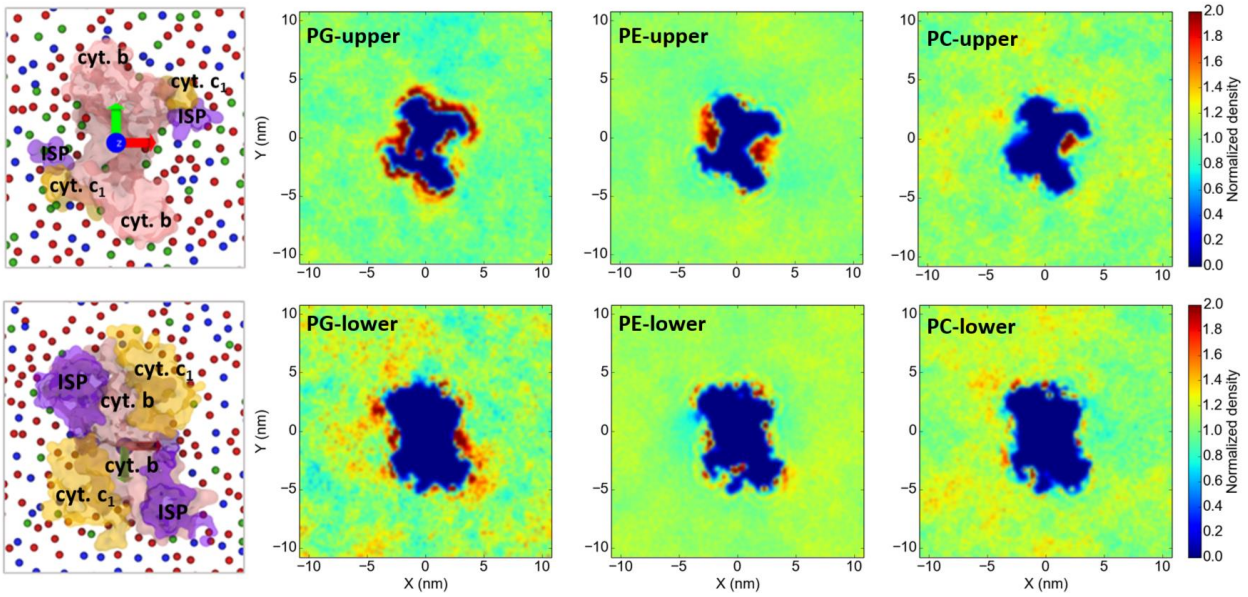

Figure 3-6. **2D density maps of the three different chromatophore lipids**. **Top panels:** cytoplasmic view of bc_1 complex and lipid density in the upper leaflet. **Lower panels**: periplasmic view of the protein and lipid density in the lower leaflet. In the snapshots of the protein: cytochrome *b* (pink), cytochrome c_1 (yellow), ISP (purple) and lipid head groups are colored as in **Figure 3-4**.

3.3.2 Depletion-Enrichment index analysis

The density map analysis outlines the lipid distribution in the simulation system. To quantify the preference of the protein for the three lipids studied, the depletion-enrichment (D-E) index analysis was performed as described in [79]. **Table 3-1** shows the enrichment index for approximately the first three lipid shells around the protein represented by 0.7 nm, 1.4 nm and 2.1 nm. The analysis was performed separately for the upper and lower leaflet in the last 5 µs of the simulation. According to the definition of the D-E index, values larger than 1 indicate enrichment of a given lipid group within a given distance cut-off, while values smaller than 1 indicate depletion.

Table 3-1 Depletion-Enrichment index analysis

	Upper			**Lower**		
	0.7 nm	1.4 nm	2.1 nm	0.7 nm	1.4 nm	2.1 nm
POPG	1.66±0.18	1.32±0.10	1.19±0.07	1.10±0.13	1.13±0.03	1.13±0.02
POPE	0.93±0.05	0.94±0.03	0.91±0.02	1.00±0.02	1.02±0.02	1.04±0.01
POPC	0.68±0.09	0.79±0.05	0.85±0.03	0.80±0.10	0.94±0.04	1.00±0.04

For PG lipids, values higher than 1 were obtained in the upper and lower leaflets for three distance cut-offs. In the upper leaflet, a little depletion of PE lipids is observed for the three distance cut-offs. In the lower leaflet, PE lipids appear to be slightly enriched around the protein in three distance cut-offs. Finally, PC lipids were depleted in both upper and lower leaflets although a gradual enrichment increase was observed in both leaflets as the cut-off distance increases.

3.3.3 Thickness of the membrane

To illustrate the membrane structure perturbation induced by bc_1 complex, the thickness of the bilayer was calculated. **Figure 3-7** shows the membrane thickness calculated as an average of the PO4 beads distances between the upper and lower leaflets averaged over three windows of 5 µs each. The thickness pattern is not uniform around the protein surface, and shows strong differences compared to the thickness of the bulk region of the membrane.

The thinner regions correspond to cytochrome c_1, ISP and the cavities between the monomers. The thinner regions that correspond to the dimer cavities match the areas of increased density of PE and PG shown in the density map (**Figure 3-6**). Regions of relatively thicker membrane are found surrounding the surfaces of cytochrome b monomers.

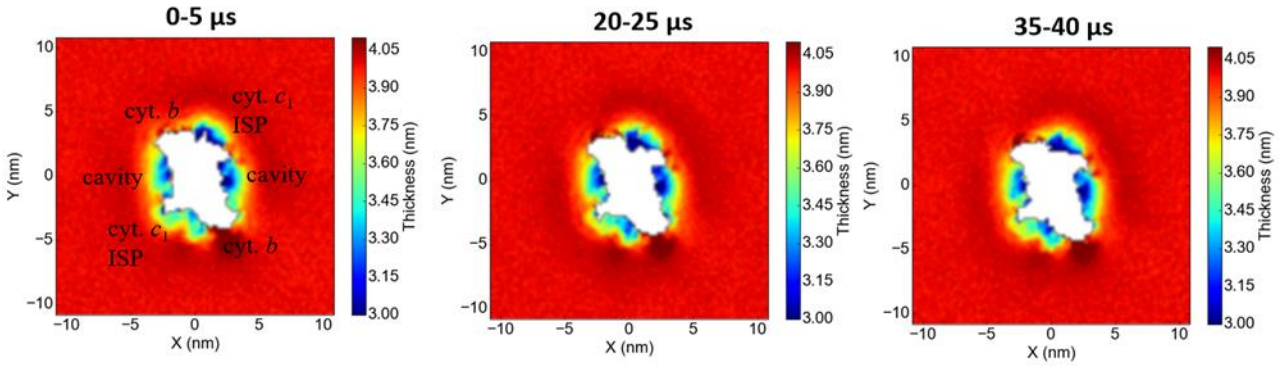

Figure 3-7 Membrane thickness. White regions illustrate the protein surface where there are no lipid head groups.

3.3.4 Protein residues interacting with PG and PE lipids

To identify the main residues that contribute to PG enrichment and high PE density around the protein, the residues that maintain contacts with the lipid phosphate groups at a distance less than 0.7 nm were identified. The analysis was performed over the last 10 µs and only the residues that keep contacts during more than 50% of the time and were present in both monomers were chosen. This analysis shows that PG lipids preferably interact with a belt of positively charged residues (Lys and Arg) localized on the cytoplasmic side of the protein (**Figure 3-8 A, B**)(**Table 3-2**). PE lipids do not have any obvious binding site and interact with a wider range of residues (**Table 3-2**). Several residues on the cytoplasmic site establish contacts with both PG and PE lipids.

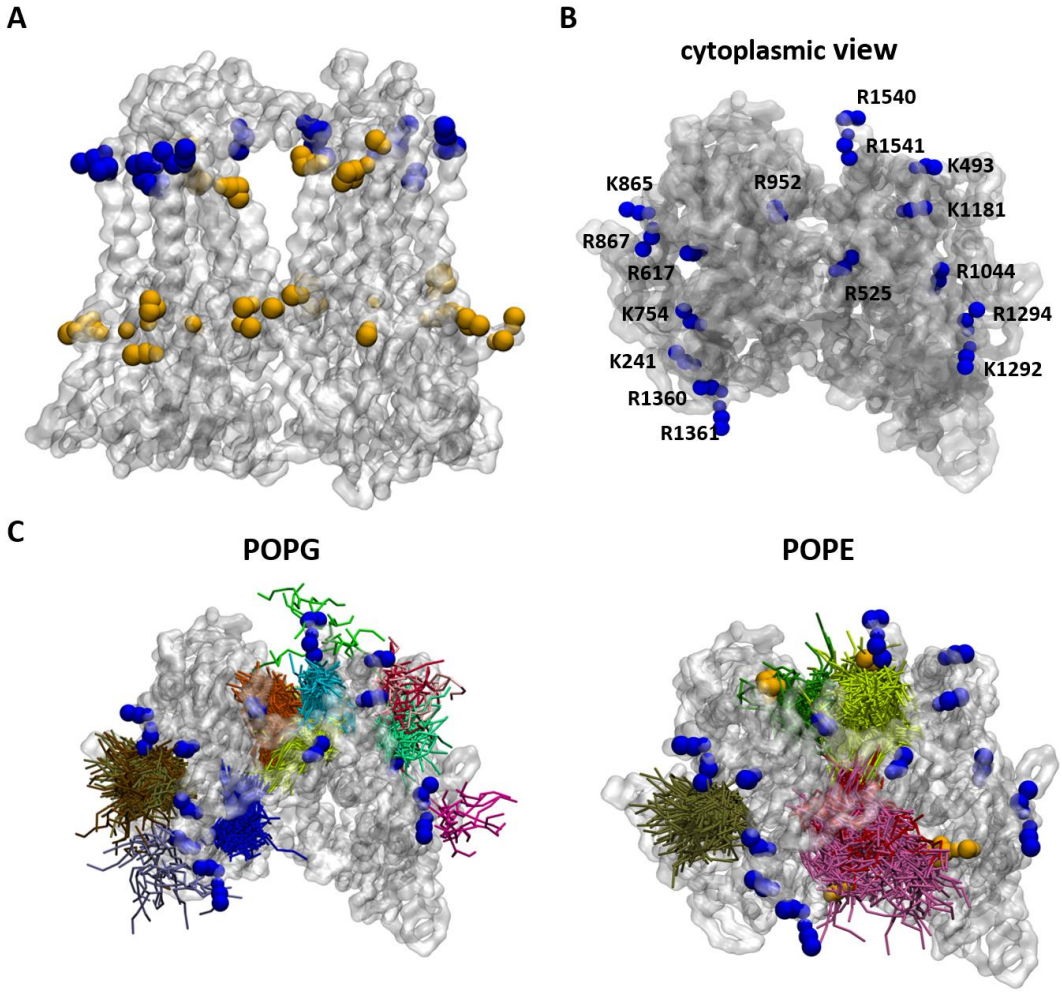

Figure 3-8 Residues contributing to bc_1 complex-lipid interactions. A) Distribution of the residues that interact with the protein over the last 10 µs of simulation time in more than 50% of the frames. Residues in blue (Arg and Lys) interact with both PE and PG lipids. Residues in orange (Asn, Ser, Trp, Phe, Gln, Tyr, Thr, Glu, His, Ala) interact only with PE lipids. Residues are shown in van der Waals representation. **B)** Top view of the protein. Arg and Lys residues represented in (**A**) are shown. **C)** Multiples frames of PE and PG lipids that interact with the protein between 35-40 µs are shown in licorice representation, in different colors.

A high-density spot corresponding to one PG molecule is manifest in the density map (**Figure 3-6**). This corresponds to one PG that penetrated the inner cavity of one monomer early in the simulation and reached the site composed of the residues Thr962, Trp975, Phe1146, Asn1151, Phe1174, His1147. This site corresponds to one of the heme b_H groups of the bc_1

complex that is not represented in the CG model. The PG molecule is shown in yellow in **Figure 3-8 C**.

Table 3-2 Residues from the bc_1-complex that interact with PG and PE lipids. Residue numbers correspond to the CG model of the bc_1 complex. The letter of each protein chain is shown as superscript. Residues that also maintain interaction with PG lipids are shown in bold.

Subunits	POPG	POPE
Cytochrome c_1	Lys241[a], Lys493[b]	Asn118[a], Asn370[b], Gly12[a], Gly264[b], Lys16[a], Lys268[b], ***Lys241[a]***, ***Lys493[b]***, Ser115[a], Ser367[b], Phe14[a], Phe266[b], Ala217[a], Ala469[b], Gln120[a], Gln371[b]
Cytochrome *b*	Arg525[c], Arg942[d], Lys865[c], Lys1292[d], Arg867[c], Arg1294[d], Lys754[c], Lys1181[d], Arg617[c], Arg1044[d]	Trp851[c], Trp1278[d], Tyr890[c], Tyr1217[d], Asn591[c], Asn1018[d], ***Lys865[c]***, ***Lys1292[d]***, ***Arg867[c]***, ***Arg1294[d]***, Ser678[c], Ser1105[d], Thr854[c], Thr1281[d], Thr825[c], Thr1252[d], ***Arg617[c]***, ***Arg1044[d]***, Glu629[c], Glu1056[d]
ISP	Arg1359[e], Arg1540[f], Arg1360[e], Arg1541[f]	Asn1384[e], Asn1565[f], ***Arg1360[e]***, ***Arg1541[e]***, His1365[e], His1546[f], ***Arg1359[e]***, ***Arg1540[f]***, Asn1387[e], Asn1568[f]

3.3.5 Membrane deformation

Visual inspection of the system shows whether the bc_1 complex would generate local curvature of the membrane (**Figure 3-9**). The snapshots show that the membrane remains flat in the time period studied. However, it is observed that the protein induces a slight remodeling of the bilayer in the region surrounding the protein at 30 μs, which persists over the rest of the simulation, as shown in the last snapshot at 40 μs. The resulting concave shape involves approximately four layers of lipids around the protein in both the upper and lower leaflets. As shown in **Figure 3-9**, the lipid phosphate groups surround the bc_1 complex guided by a belt of charged residues on the protein surface.

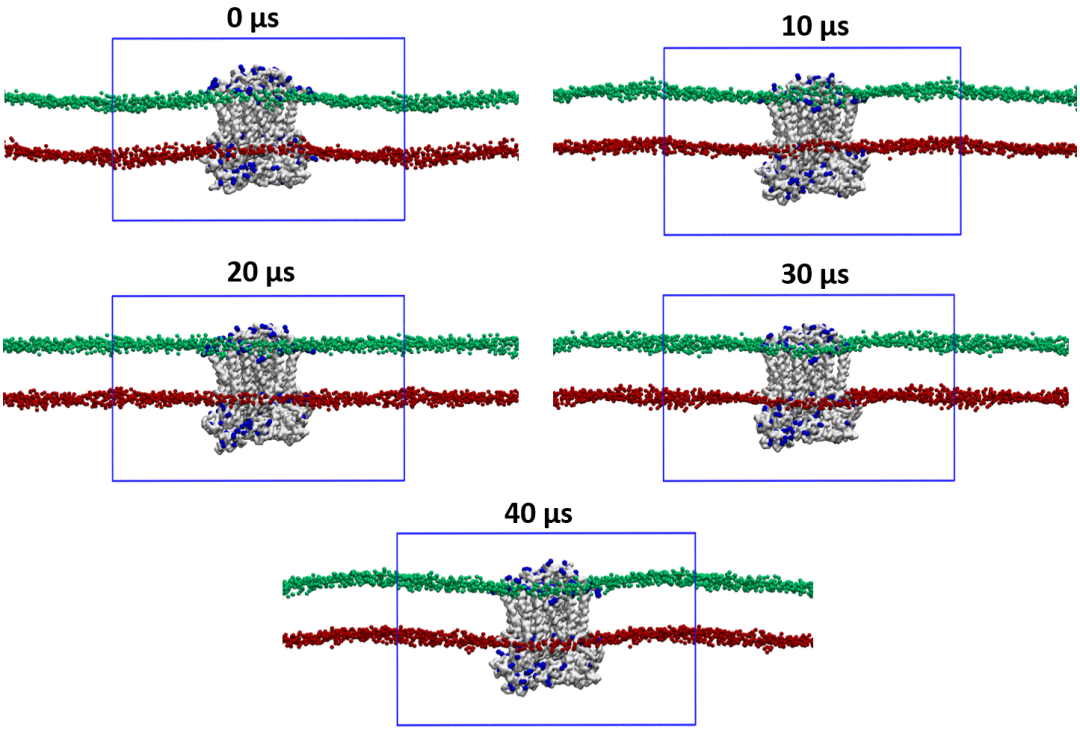

Figure 3-9 Membrane deformation around the *bc*$_1$ complex at different times. Upper leaflet in green, lower leaflet in red, protein is represented in white surface, and positively and negatively charged residues of the protein surface in blue.

3.4 Discussion

The structure and function of integral membrane proteins depends strongly on the lipid composition of the lipids surrounding them [80]. At the same time, a membrane protein usually affects the layer of lipids immediately around the proteins [81]. Here we found that the number of lipids calculated over time, corresponding approximately to the first lipid shell around the *bc*$_1$ complex, maintains a similar trend over the simulation. This indicates that adequate sampling of the lipid distribution around the protein was obtained during the 40 µs simulation. This result is consistent with recent lipid-protein interaction studies of similar systems performed using CG simulations [79].

Enrichment of PG lipid surrounding the protein surface

One of the most interesting finding of this study is the enrichment of PG lipids near bc_1 in the cytoplasmic side of the membrane. A belt of positively charged Arg and Lys residues located on the cytoplasmic region of the bc_1 complex can explain the enrichment of the anionic PG lipids. This result is consistent with results obtained by Arnarez et. al (2013) [49] who found that cardiolipin binding sites on two bc_1 complexes, bovine heart and yeast CIII, were also enriched in positively charged residues Arg and Lys, localized on the protein surfaces.

In general, the abundance of positively charged residues on the cytoplasmic side of the protein is a typical feature of inner membrane protein topology and seems to be associated with their mechanism of biogenesis [82, 83]. Van Klompenburg et al. described that the presence of positively charged residues and negatively charged lipids work together to determine the orientation of the membrane proteins in the cell [84]. In particular, for a bacterial respiratory complex, functionally comparable to a bc_1 complex, it has been reported that the absence of anionic lipids in the membrane affects the function of the protein [85]. Even though this result mainly referred to cardiolipin lipids, the authors pointed out that the lack of PG lipids, although to a lesser extent, may also result in similar effects.

To analyze the enrichment of PG in the lower leaflet, the residues corresponding to the periplasmic side of the protein were examined. The result shows that PG lipids interact randomly with positively charged and polar residues, which may explain the enrichment and the high-density mass spots observed in the lower leaflet. However, considering the lower percentage of PG in the lower leaflet, only a couple of lipids around the protein, for instance, 4 ± 2 as obtained by lipid count analysis, make a great contribution to the enrichment analysis.

High density of PE and PG lipid in the inner cavities and surrounding areas of the dimer

A high-density signal of PE lipid is observed in the inner cavity of the bc_1 complex (**Figure 3-6**). Although to a lesser extent, PG lipids were also found in this cavity of the protein. A recent study based on X-ray data reported the presence of lipid acyl tails connecting the quinol oxidase site (Q_0-site) and the inner cavity of the cytochrome b_6f [68]. The authors pointed out that the location of these lipids may enhance the quinone/quinol traffic through the protein. It was suggested that since quinone/quinol molecules have a long hydrophobic tail, they can interact with other sites of the protein beyond the quinone binding site. Therefore, the presence of lipids in the cavity, localized inside the hydrophobic core of the b_6f complex, might occupy potentially inefficient quinone binding sites, which might enhance the quinol/quinone diffusion through the protein [68]. A CG-MD simulation study also illustrated the presence of lipids (POPC and cardiolipin) in the transmembrane inner cavity of both bc_1 complexes, bovine heart and yeast CIII [49].

We found that two charged residues, Glu629 and Glu1056, located at the inner surface of each cytochrome *b* might contribute to the crowding of PE into the complex cavity. The fully protonated ammonium group of PE lipids can electrostatically interact with the negatively charged side chain of the glutamate. Kwa et al. proposed a similar interaction pattern between residue Glu20 of LH2 from *R. sphaeroides* and PE lipids [72]. The small head group of PE lipids and its tendency to induce non-bilayer curvature might explain the deformation of the bilayer observed in the inner protein cavities of the atomistic model of the chromatophore. Further analyses should be performed to evaluate the extent of the interaction between these residues of the bc_1 complex and PE lipids.

Previous experimental studies have highlighted the importance of PE lipids for the insertion of bacterial inner membrane proteins into the lipid bilayer and for the proper functioning of these proteins [83, 86]. It has been shown that both PE and PG lipids can be localized in the annular region of bacterial membrane proteins such as lactose permease (LacY), which uses the proton gradient to transport sugar into the cell [87]. In our simulation, PG and PE lipids were identified in the proximity of the bc_1 protein (**Figure 3-6, Figure 3-8 A, B**).

Lipid organization around *bc*$_1$ complex

Analysis of the thickness of the bilayer showed that lipids were not uniformly organized around the *bc*$_1$ complex. Cytochrome *b* has the greatest and hydrophobic area of the dimeric protein, which is favorable to establish hydrophobic interactions with the lipid acyl chains and may explain the thicker regions of the bilayer around these monomers (**Figure 3-7**). Although the topology of the monomers that comprise the dimeric structure of *bc*$_1$ complex is identical, the monomers are dominated by conformational changes that take place during protein function. When one pocket of cofactors (Q_0 and Q_i) binds quinone, the other pocket binds semiquinone or quinol; this difference induces different conformations in the iron-sulfur proteins of the complex [66]. This disparity explains the asymmetry in the results observed in the density and thickness maps. On the other hand, due to the coarse-graining approach used in this study, no conformational changes can occur in the protein beyond minor fluctuations around the average structure. Therefore, the outcomes of the simulation presented here strongly depend on the initial structure of the protein. **Figure 3-10** summarizes the behavior of the lipids around the protein and shows the acyl chains of the lipids covering the hydrophobic grooves of the protein surface, and binding through nonpolar interactions to the amino acids.

The dynamic process of photosynthetic membrane remodeling is challenging to observe experimentally but can be observed with computational modeling as has been demonstrated by Hsin et al. (2010) [63]. Information about the membrane architecture has been relevant to understand the functional processes that take place in the chromatophore. Therefore, knowledge about the lipid bilayer curvature around the *bc*$_1$ complex may offer potential insights into the location and function of the protein in the chromatophore. The result obtained here further support the idea that the *bc*$_1$ complex does not induce curvature of the membrane, as the membrane remains overall flat on the time scale of this simulation **(Figure 3-9, Appendix B**) [62]. However, in some regions, the PO4 groups of the lipids can be lower than in others because of interactions with charged residues on the protein surface. Other phosphate groups access the hydrophobic cavities between the monomers in the upper leaflet causing the thinner region of the bilayer and provoking a slight bending of the membrane

(**Appendix B, Figure B1, A**). The results from the CG simulation reported in this thesis are consistent with recent experimental and computational works (Singharoy et al., unpublished), supporting the idea that bc_1 complexes are found in a concave and flat region during membrane budding [29].

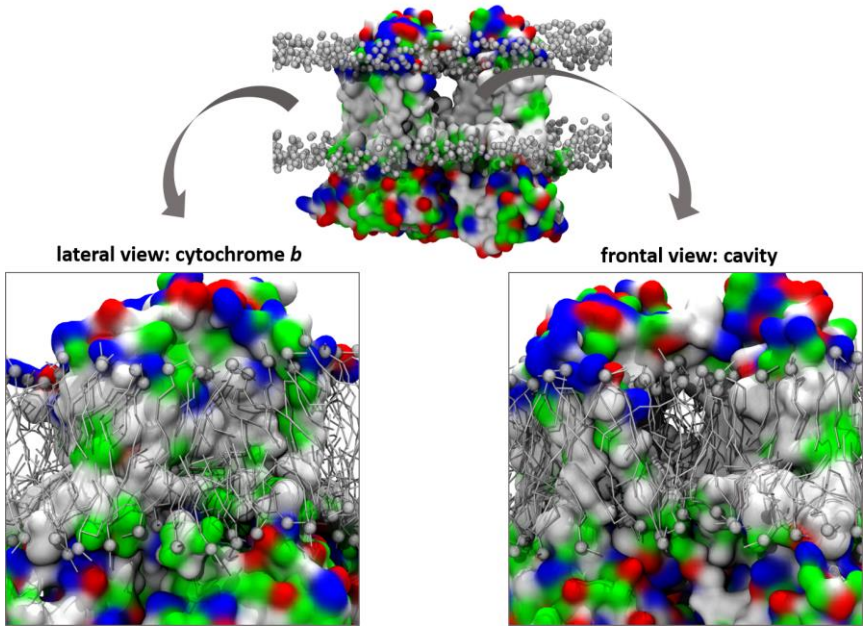

Figure 3-10 Lipid organization on the protein surface. Protein is shown in Quick Surface representation, positively charged residues (blue), negatively charged residues (red), polar residues (green) and nonpolar residues (white). In the front view (lower right panel), lipids occupy the cavity between the dimers. In the lateral view (lower left panel), corresponding to cytochrome *b*, lipid tails are straighter than in the protein cavity. In both representations lipid head groups surround charged and polar residues on the protein surface.

In addition to electrostatic interaction, a hydrophobic mismatch between the lipids and the bc_1 complex might also explain the membrane remodeling around the protein. In fact, lipid acyl chain stretched in some regions around the bc_1 complex, while in others, such as in the protein cavity, a more disorder of lipid tails is observed. This behavior might be induced by the fact that lipids accommodate and cover the hydrophobic surfaces of the protein avoiding its unfavorable exposure to the hydrophilic environment.

Based on the previous result, it could be hypothesized that a similar event takes place in the chromatophore vesicle. The deformation of the bilayer around the bc_1 complex may be a consequence of two aspects: i) the presence of charged residues in the surface of the protein which induce a certain distribution of lipids around them and ii) the difference in thickness mismatch concerning the rest of photosynthetic proteins. Therefore, the lipid bilayer remodels in such a way to go up and down to cover the hydrophobic surfaces of all the photosynthetic proteins presented in the vesicle. Interestingly, the rest of the photosynthetic proteins LH2, LH1-RC, and ATPase have hydrophobic surfaces with similar shapes and sizes respect to the normal of the membrane, which could explain the homogeneous degree of curvature around these proteins in the chromatophore vesicle, contrary to the deformation observed around the bc_1 complexes (**Figure 3-2**).

This study was limited by the absence of cofactors molecules, such as the heme groups and the 2Fe-2S group, in the bc_1 complex. Due to the lack of cofactors, at least one lipid penetrates to the position occupied by a heme b_H group in one of the cytochrome *b* proteins. Another limitation of this study is related to the CG nature of the simulation approach. The protein conformation is fixed, thus changes in the structure that could potentially be induced by lipids cannot be observed.

Therefore, further research should be focused on evaluating the lipid-protein interactions together with the cofactors at an atomistic level of detail through backmapping the CG system to an AA model and the inclusion of the cofactors in the model. Additionally, it would be interesting to check if the residues found in this study that maintain contacts with the lipids, are evolutionarily conserved in other bc_1 proteins.

Chapter Four: Molecular views of a eukaryotic plasma membrane

4.1 Introduction

To date, several studies have begun to investigate the dynamic organization of lipids and proteins in cell membranes using Martini CG simulations [7, 88]. The complexity of these models is revealed by the variety of lipid species included in the simulation systems, their asymmetric arrangement in the bilayer, and the presence of numerous copies and types of proteins. Most of the recent simulations of these complex membranes have been carried out in parallel with the study presented in this dissertation.

A pioneering study is a work performed by Ingólfsson et al. in 2014 [89]. The authors modeled an idealized mammalian plasma membrane, consisting of approximately 20 000 lipids, with 63 lipid types, including 14 different head groups and 11 different tails, asymmetrically distributed between the leaflets. The system, simulated for 80 μs, with an unprecedented complexity in lipid composition and simulation time, provides a detailed view of the lipid organization of the plasma membrane. The model gives information about lipid flip-flop, transient domain formation and coupling of lipids domains between the bilayer leaflets, events that have been experimentally observed in plasma membrane models [89]. Recently, a similar study was performed to investigate the lipid organization in a human brain plasma membrane model [90].

The addition of proteins to membranes models has further increased the complexity of large-scale simulation systems [88, 91-94]. Koldsø and Sansom (2015) [94], for instance, embedded different copies of proteins found in eukaryotic plasma membranes such as glycophorin A dimer, G-protein-coupled receptors, and α-helix transmembrane domains, in a bilayer composed of seven species of the most representative lipids found in the mammalian plasma membrane. Three systems, at length scales greater than 100 nm, were built. Each system consists of several copies of a given protein type, with proteins covering ca. 6-12% of the membrane. Simulation results after 5 μs show that the presence of proteins stabilized the

curvature and fluctuation of the membrane. Specific interactions between cholesterol (CHOL) and phosphatidylinositol (PIPs) lipids with the proteins were observed, which may have biological significance [94].

Another example is a work by Marino et al. (2016) [93]. In this study, the authors used an array of 16 copies of the μ-opioid receptor, a GPCR protein, in membrane patches of 25×25 nm^2 and 50×50 nm^2 each. The goal was to investigate the lipid distribution around the receptors in different conformations and study how this could impact receptor dimerization. The membranes represented two different degrees of protein crowding. The lipid composition of the membrane was the same as the model developed by Ingólfsson et al. [89].

Although the studies mentioned above provide evidence of recent progress in simulating complex membrane models, two aspects remain challenging: 1. building large-scale simulations systems, and 2. studying lipid-protein interactions in lipid environments that approach the level of protein crowding and packing of membranes in living cells. Another significant obstacle in the study of large-scale systems is the time required to equilibrate lipid-lipid and lipid-protein interactions before extracting reliable results from the simulation. The dynamics of the Martini model is fast due to the reduced degrees of freedom. However, complex and large systems require longer timescales. Lipids take tens of microseconds to equilibrate around proteins in systems with a box size of ca. 37 nm x 37 nm in X and Y [79]. Thus, it is expected that a more prolonged equilibration time is needed for bigger systems (X and Y greater than 120 nm).

Specifically, our group has been interested in building more representative models of a generic eukaryotic plasma membrane. Initially, the lipid environment around ten eukaryotic plasma membrane proteins using the CG Martini model was explored in detail [79]. The eukaryotic membrane proteins include: P-glycoprotein (P-gp, PDB ID 4M1M) [95]; Aquaporin-1 (AQP1, PDB ID 1J4N) [96]; dopamine transporter, (DAT, PDB ID 4M48) [97]; epidermal growth factor receptor (EGFR) [98]; transporter 1 (GluT1, PDB ID 4PYP) [99]; voltage-dependent Shaker potassium channel 1.2 (Kv1.2, PDB ID 3LUT [100], residues 32 to 4421 for each monomer); sodium, potassium pump (Na,K-ATPase, PDB ID 4HYT) [101]; AMPA-sensitive glutamate

receptor 2 (GluA2, PDB ID 3KG2) [102]; δ-opioid receptor (δ-OPR, PDB ID 4N6H) [103]; Prostaglandin H_2 synthase (COX1, PDB ID 1Q4G) [104].

The choice of these proteins was based on the availability of their three-dimensional structures (**Figure 4-1A**), their eukaryotic origins, and an attempt to represent some of the diversity of the membrane proteome. Proteins exemplify the main classes of eukaryotic plasma membrane proteins including receptors (GluA2, EGFR, δ-OPR), transporters (P-gp, GluT1, DAT, Na,K-ATPase), channels (Kv1.2, AQP1) and enzymes (COX1). Some of these proteins are largely membrane embedded (e.g. AQP1, DAT) whereas others have large extramembrane domains (e.g. EGFR, GluA2).

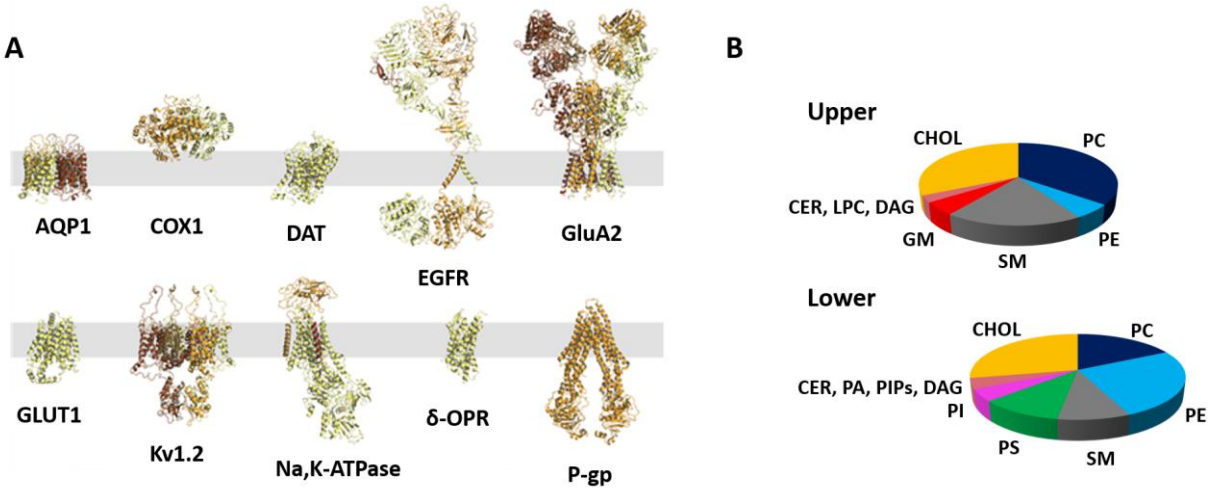

Figure 4-1 Composition of the complex biological membrane model. **A)** Cartoon representation of the membrane proteins used in the study by Corradi et al. (2017) [79]. Grey region represents the hydrophobic core of the membrane. **B)** Pie charts show the distribution of the main lipid head groups in the upper and lower leaflet: Cholesterol (CHOL), phosphatidylcholine (PC), phosphatidylethanolamine (PE), sphingomyelin (SM), ganglioside (GM), phosphatidylserine (PS), phosphatidic acid (PA), phosphatidylinositol (PI), PI-phosphate, -bisphosphate, -trisphosphate (PIPs), Ceramide (CER), diacylglycerol (DAG), lysophosphatidylcholine (LPC). Figure A adapted from Corradi et al. (2017) [79].

Four copies of each of the proteins shown in Figure 4-1A were embedded in the plasma membrane model from Ingólfsson et al. (2014) [89]. Analyses of the simulations, performed over 30 to 50 μs, show that the timescale for lipid organization and redistribution induced by the presence of the proteins, although variable from system to system, reaches a certain degree of convergence after approximately 25 μs [79]. Highlighting the complexity of lipid-protein interactions, the results show that each of the ten proteins affects the lipid environment and membrane properties in a unique way. In fact, a key result is that the composition and organization of lipids surrounding each protein were different for each protein type. Proteins prefer lipids of a certain head group type or with a given acyl chain length or degree of unsaturation. Many proteins are surrounded by a shell of annular lipids that behave differently than bulk lipids based on their exchange rates at the protein-lipid interface [105]. In the next sections we describe how we took advantage of these simulations to overcome some of the major limitations of currently available tools to build large-scale membrane systems.

Current tools for building plasma membrane models

Several tools including web-based or distributed applications have been developed to construct membrane structures, for reviews see [106, 107]. Models can be generated by solvating the protein(s) with lipid forming a bilayer and adding water molecules and ions into the system. Two widely-used tools to build initial configurations of biological membranes are the web-service CHARMM-GUI [38, 108] and the Python-based tool INSANE (INSert membrANE) [75]; the latter is specifically geared towards generating coarse-grained models. Despite the successes of these methods, they present some limitations to constructing multi-protein membrane models. For instance: CHARMM-GUI only builds membrane containing a single protein, and INSANE is not suitable for inserting several copies of proteins with different shapes in a bilayer, mainly due to the time-consuming task of generating an array of proteins with aligned hydrophobic belts before covered them with the non-polar lipid tails.

To support initial building of multi-component bilayers and to overcome some of the limitations of the current methods, Plasmabuilder, a Python-based tool, was developed in our group written by Dr. Gurpreet Singh (**Appendix A**). Given a current set of files, each containing a protein embedded in a bilayer, the plasmabuilder extracts the protein along with lipids within a given cut-off and places them in a larger system according to a specified ratio of lipids and proteins. Parameters that can be modified to build the desired system include the number, composition, and distribution of lipids in the upper and lower leaflets, the total area of the bilayer to be covered by proteins and the number of proteins inserted. The proteins together with their neighboring lipids are taken from previous simulations and the rest of the lipids to complete the bilayer from a library. Both lipids and proteins are represented using the Martini CG model.

In the initial tests carried out to assess the plasmabuilder performance in building test systems, two main issues were identified. First, the final configuration of the membrane model includes space not filled with lipids. These holes affect the initial distribution of lipids in the membrane because they allow the diffusion of lipids from one leaflet to another. A second problem associated with the first version of the plasmabuilder is that some lipids in the system show a certain degree of overlap. This affects the building process of the model and simulations because these overlaps could not be resolved during the energy minimization steps.

In this study, a new protocol based on improving the performance of the plasmabuilder and the development of MD simulation steps for building complex biological membrane models is presented. Preliminary analysis performed on a system simulated for 6 μs are shown. Finally, the discussion focuses the remaining challenges in building more realistic large-scale membrane models.

4.2 Methods

4.2.1 Setting up the plasmabuilder to build large-scale membrane models

The plasmabuilder was used to create the initial configuration of larger membranes. The building blocks consisting of proteins and their annular lipids were extracted from smaller systems simulated in our group [79]. As explained in the Introduction, after the proteins and neighboring lipids were placed in the larger membrane, a certain degree of lipid overlap took place (**Figure 4-2 A**). In principle this could be solved by manually separating the lipid, but of a large system with thousands of lipids this is not feasible. To avoid the lipid overlap, the area assigned to the lipids that are inserted in the upper and lower leaflet to fill the box was increased by 30%.

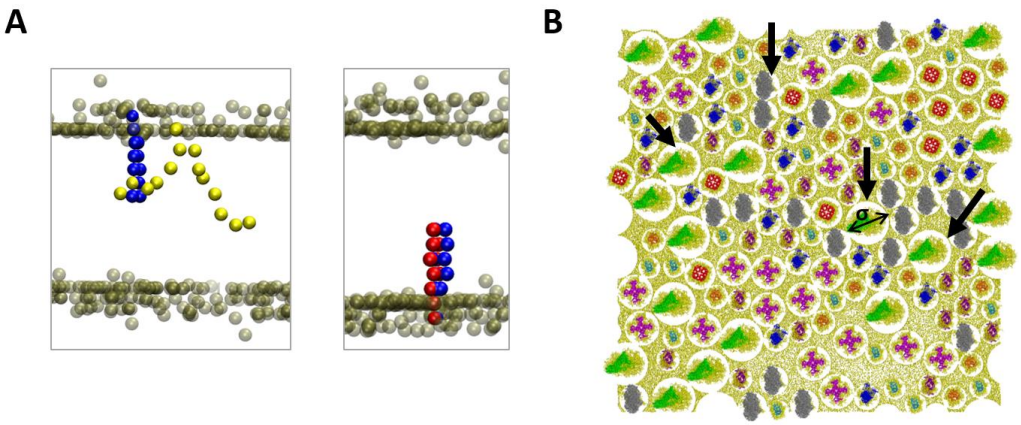

Figure 4-2 Technical difficulties of the plasmabuilder. A) Lipid overlap, between the lipid inserted with the protein (yellow) and the lipid inserted to complete the system (blue), or between lipids inserted to fill the membrane space (blue and red). **B)** Holes around EGFR (green) and GLUA2 (gray) are shown with the arrow. Lipids are in yellow and proteins in different colors.

The holes present in the initial configuration of the system, described in previous paragraphs (**Figure 4-2 B**), are due to how the plasmabuilder inserts each protein with its annular lipids (extracted from the small-scale simulations) in a larger system. The plasma builder uses a parameter sigma, the distance between the two farthest particles of the group, which becomes the diameter of circle surrounding the protein and lipids. If the protein is

symmetric with a cylindrical shape as AQP1 (red) (**Figure 4-2 B**), all lipids are surrounding the protein at the same distance from the circumference edge, and the free lipid space is quickly closed after some equilibration steps. However, if the transmembrane region of the protein is not symmetric, like in EGFR (green) and GLUA2 (gray), there is empty space of molecules corresponding to the narrow region of the protein, which takes longer to be closed, enabling undesirable effects such as the movement of lipids between the lipid leaflets. Several approaches to avoid this problem were tested, including several unsuccessful results described below.

First, the model generated with the plasmabuilder was solvated with water and ions, and position restraints were applied to all proteins. The position restraints on proteins did not allow the box to shrink, an event that is needed in this case. Therefore, after a few steps of molecular dynamics simulation, although some holes were closed, most of them were filled with water (**Figure 4-3 A**). Then, the system without water and ions was simulated without applying position restraints on the proteins. Both molecular and stochastic dynamics simulations were performed. The holes closed after approximately 4ns of simulation, but the arrangement of lipids in the bilayer was compromised, with lipids oriented parallel to the bilayer surface (**Figure 4-3 B**). Therefore, we concluded that water was needed to keep the arrangement of the bilayer, and simulations were carried out without applying position restraints on the proteins. After 10 ns, undulations of the membranes started to appear (**Figure 4-3 C**). Following previously published strategies to suppress bilayer undulations [89], position restraints with a force constant 2 of kJ mol^{-1} nm^{-2} were applied, on the Z direction only, to the PO4 beads of the predominant PC lipid species in the outer leaflet. After 30 ns of molecular dynamics simulations, the holes were closed. However, several lipids diffused between the leaflets, and lipids that are supposed to be only in the lower leaflet, such as negatively charged lipids, were found in the upper leaflet (**Figure 4-3 D**).

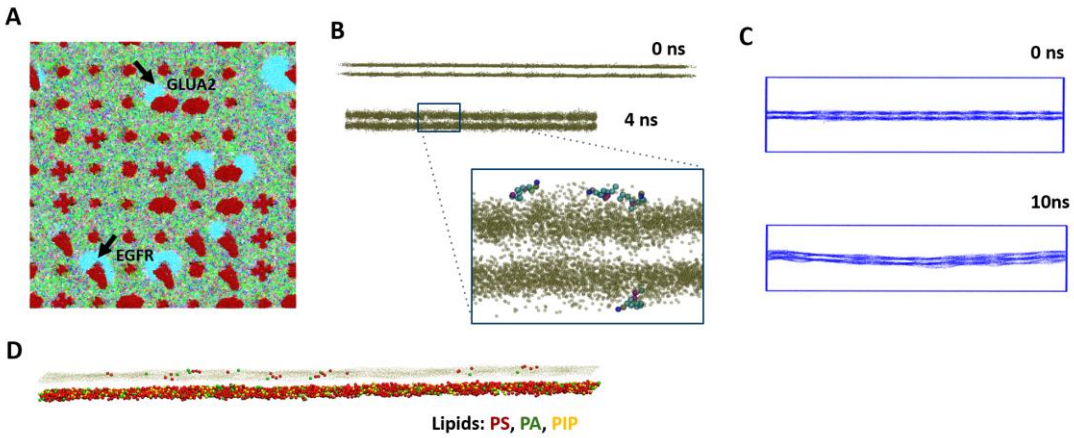

Figure 4-3 Challenges with standard approaches to close the holes around proteins. A) Examples of holes filled with water molecules are shown with arrows; protein in red, lipids in green, water in cyan. **B)** Lipids (in cyan) found parallel to the surface bilayer; lipid head groups in green. **C)** Undulations of the bilayer observed after 10 ns simulation. **D)** Examples of some lipids that are supposed to be in the lower leaflet (negatively charged lipids) in the lipid composition that mimics a plasma membrane go across the hydrophobic region of the bilayer and are found in the upper leaflet after simulation steps. Lipids are represented by their phosphate groups. The upper bilayer is depicted in gray.

To overcome the holes-related problems, position restraints with respect to the Z-direction were applied to the head groups of all lipids with force constant of 20 kJ mol^{-1} nm^{-2} until the holes around the proteins were closed. With this approach, all lipids can move in X and Y dimensions to closing the holes but not in Z, avoiding the "flip-flop". Next section describes the steps followed to build and simulate the model shown in this dissertation.

4.2.2 Model setup

Using the improved plasmabuilder, copies of eight eukaryotic membrane proteins with the equilibrated lipids were extracted from smaller membrane patches and placed in a system with an initial box size of 150 x 150 nm^2. A cut-off of 0.7 nm was used to select the neighboring lipids, accounting for approximately the first shell of lipids around each protein. The final lipid composition of the larger systems built in this thesis and the smaller systems generated in our group has the same composition and distribution of lipids between leaflets as the plasma membrane model described by Ingólfsson et al. [79, 89]. Therefore, the area per lipid obtained in the paper of a plasma membrane for the upper and lower leaflets [89], was assigned

independently to the upper and lower leaflets during the generation of our larger plasma membrane models.

Assuming a transmembrane α-helix occupies 1.5 nm^2 as described in [109], 14% of the area of the bilayer was covered by proteins. Table 4-1 shows the membrane proteins and the number of copies inserted in the model, with a total of 150 proteins. The system has over 7 million particles and 60,679 lipids. At this stage of the building process, the model still contains holes around the proteins, and a few steps of MD simulations were performed to close the holes as described in the following section (**Appendix A**, **Figure A1 D**).

Table 4-1 Protein composition of the plasma membrane model. The number of transmembrane helices (TMH) is indicated as subunits x TMHs per subunits.

Protein name	PDB ID	# copies	# TMH
P-glycoprotein (P-gp)	4M1M	20	2x6
Aquaporin-1 (AQP1)	1J4N	10	4x6
Dopamine transporter (DAT)	4M48	20	2x6
Epidermal growth factor receptor (EGFR)	-	20	2x1
transporter 1 (GLUT1)	4PYP	20	1x12
Voltage-dependent Shaker potassium channel 1.2 (Kv1.2)	3LUT	20	4x6
sodium, potassium pump (Na,K-ATPase)	4HYT	20	1x12
AMPA-sensitive glutamate receptor 2 (GluA2)	3KG2	20	4x4

4.2.3 Molecular dynamics simulations

All the simulations were performed using GROMACS 4.6 (www.gromacs.org) [40]. The force field used was Martini 2.2 [74] for proteins and Martini 2.0 for lipids [9]. The ElNEDyn approach was used to fix the protein structure [47] as described in [79]. The membrane model created with the plasmabuilder was minimized in two cycles of 2000 steps each, using steepest descent energy minimization. Then, the box was solvated with standard MARTINI water molecules (using genbox). Water molecules placed inside the hydrophobic regions of the bilayer were removed. Ion molecules were added to the simulation box (using genion) to neutralize the net charge of the system. After solvating the simulation box, another run of energy minimization was performed to relax the system and avoid any steric clashes between water, lipids, and proteins.

Two equilibration simulations of 500 ps each, with a time step of 2 fs and 10 fs respectively were performed. Then, the time step was increased to 20 fs, and another simulation was performed until the holes were closed (30 ns for a system of this size). During these equilibration runs, position restraints, with force constant of 20 kJ mol^{-1} nm^{-2}, were applied to all PO4 groups of the lipids in the upper and lower leaflets.

All molecular dynamics simulations were performed using the isothermal-isobaric (NPT) ensemble and with a 20 fs time step. The temperature was controlled using a velocity-rescaling thermostat [44] with a coupling constant of 1.0 ps (0.5 ps for equilibrium runs). A semi-isotropic pressure of 1.0 bar was maintained using the Berendsen barostat [43] with a coupling constant of 3 ps (1.0 ps for equilibrium runs) and compressibility of $3x10^{-4}$ bar^{-1}. The electrostatic interactions were calculated using a shifted potential with a cut-off of 1.2 nm and a dielectric constant of 15. The Van der Waals interactions were also calculated using a shifted potential with a cut-off of 1.2 nm and a switch at 0.9 nm.

For production runs, two systems were simulated: one system (PR-on) with position restraints (force constant of 2 kJ mol^{-1} nm^2) applied to the PO4 beads of PC lipids in the outer leaflet; a second system (PR-off) with no restraints on the PO4 beads. Both systems have been simulated for 6 μs, during which the proteins were free to move in the membrane plane.

Results were analyzed using GROMACS tools and in-house scripts. VMD software [76] was used to visualize the simulation trajectory and make the images. The radial distribution function (RDF) based on the center of mass (COM) of each protein was calculated for three-time windows (1 μs – 2 μs, 3 μs – 4 μs, 5 μs – 6 μs) with a bin width of 0.4 nm using the GROMACS command g_rdf. To represent the position of the proteins on the curved bilayer for the PR-off system, the deformation of the membrane was estimated for three snapshots of the systems at 0 μs, 3 μs, 6 μs and the minimum and maximum value of the z coordinates of the PO4 beads of the lipids respect to the initial position was calculated using a Python script (stream_Vs_z_2D.py) [110]. The number of GM and PIP lipids in contact with AQP1, Kv1.2, DAT and EGFR proteins, as a function of time, was approximated by the number of GM1 (for GM

lipids), CP (for PIP lipids) beads within a 0.7 nm and 2.8 nm cut-off from each protein, as done previously in [79]. The calculation was performed using the g_select GROMACS tool [40].

In addition to the previous model, three models were built with the same lipid composition as the plasma membrane model described by Ingólfsson et al. [89]. In addition to the proteins described in (**Table 4-1**), two membrane proteins - prostaglandin H_2 synthase (COX1, peripheral membrane protein) and δ-opioid receptor (δ-OPR, integral membrane protein) - were added. Consisting of a different distribution of proteins in the membrane, the models have box sizes of 152 nm x 152 nm x 60 nm, and contain 148 proteins, over 12 million of particles in total. Two of these models have been simulated for 1 μs and the third for 40 μs. One of them is currently used for force-field development and testing in Dr. Marrink's group, Groningen. The simulations are being performed using GROMACS software package (version 5.1.1) [40].

4.3 Results

4.3.1 Protocol to build membrane models

A new protocol to build and simulate complex plasma membrane models was developed and consists of the following steps: (1) use the plasmabuilder to extract the proteins with the equilibrated neighboring lipids from small membrane patches (before this step, the size of the box, cut-off for the neighboring lipids, area of the membrane covered with proteins, and the lipid composition of the upper and lower leaflets need to be defined); (2) insertion of additional lipids to fill the box and match the membrane composition; (3) solvate the system with water and ions; (4) perform short equilibration simulations with position restraints in the z-direction on lipid headgroups, in both the upper and lower leaflets; (5) to continue with production run, remove the position restraints applied on all lipid head groups. With this procedure, proteins with annular lipids extracted from other simulations are combined to build large membrane patches.

A model that mimics a prototypical eukaryotic plasma membrane was built consisting of 150 proteins and more than 60,000 types of lipids (**Figure 4-4 A**). The model contains multiple

copies of eight representative eukaryotic plasma membrane proteins inserted in a bilayer with their pre-equilibrated neighboring lipids within a cut-off of 0.7 nm. These proteins with lipids were taken from previous simulations performed in our group where the lipid distribution reached a reasonable degree of convergence after approximately 25 µs. A benchmark analysis shows the performance in nanoseconds per day (ns/day) for a system with a box size of 136 nm x 136 nm x 42 nm box size and 7,102,116 particles simulated using GROMACS 4.6 [40] (**Figure 4-4 B**). The simulations presented in this chapter were performed using 2400 cores.

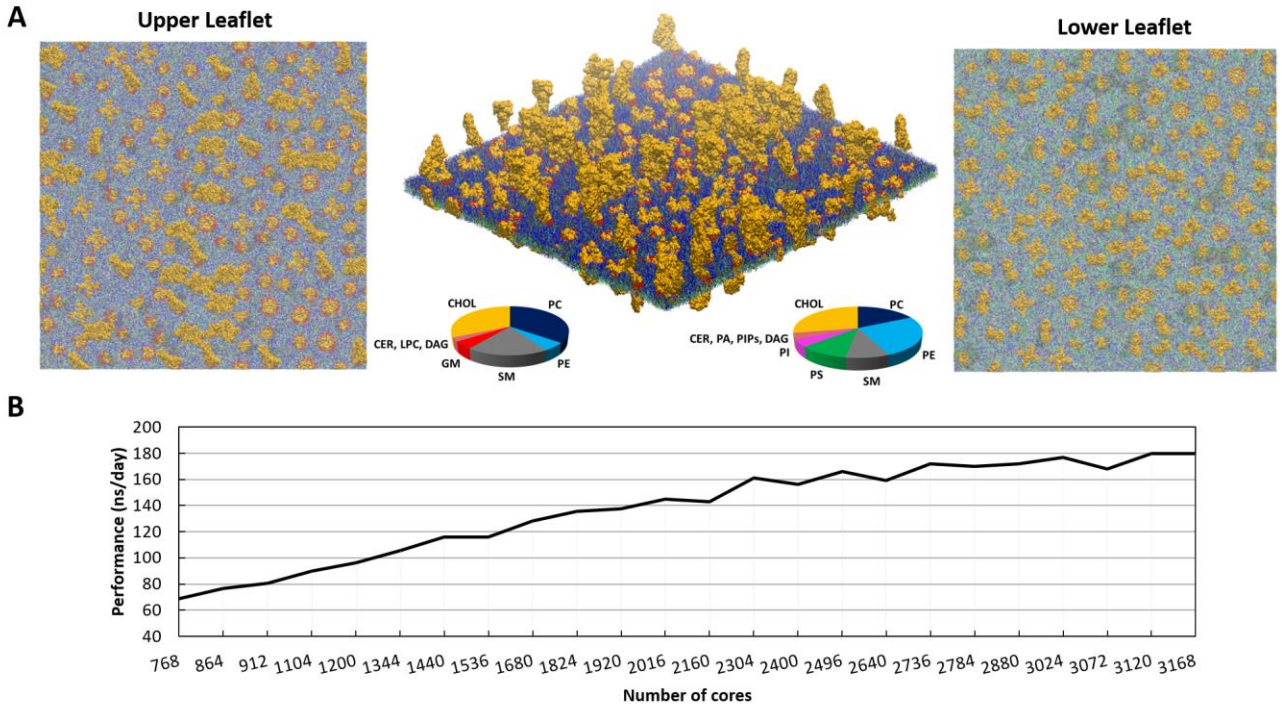

Figure 4-4 Plasma membrane model. A) Proteins are represented in orange while lipids are shown in different colors depending on their head groups. Pie charts show the distribution of the main lipid head groups in the upper and lower leaflets. Box size (136 nm x 136 nm x 45 nm). **B)** Benchmark of the system composed of over 7 million of particles. Position restraints on lipid head groups were applied.

4.3.2 Preliminary analyses of a plasma membrane model

Using the protocol described in section 4.3.1, a plasma membrane model was built and simulated for 6 μs. To investigate the early effect of undulations on the membrane structure, we compared two approaches: i) a membrane with position restraints (PR-on) applied on phosphate groups of the predominant PC lipid species on the upper leaflet with respect to the Z-position only - this strategy was used in [89] to suppress bilayer undulations - and ii) a membrane without position restraints (PR-off). **Figure 4-5** shows the effect of these two approaches on the dimensions of the box. In the system PR-on, the size of the box in X, Y and Z fluctuates around average values that stabilize early in the simulation. In contrast, in the system PR-off a decrease in X, Y box sizes is accompanied by an increase in the Z-dimension of the box, which is a consequence of the bilayer undulations (**Movie 1**) and an almost constant box volume. It is important to highlight that in these graphs the simulations performed to close the holes around proteins, which took approximately 30 ns, are not included. The time required to close the holes depends on the size of the systems and type of protein inserted. The distribution of lipids around the proteins or the stable value of the box sizes in the plane of the membrane does indicate that the structure is suitable for performing production run simulations. The following sections show the initial behavior of this complex system under the two different conditions.

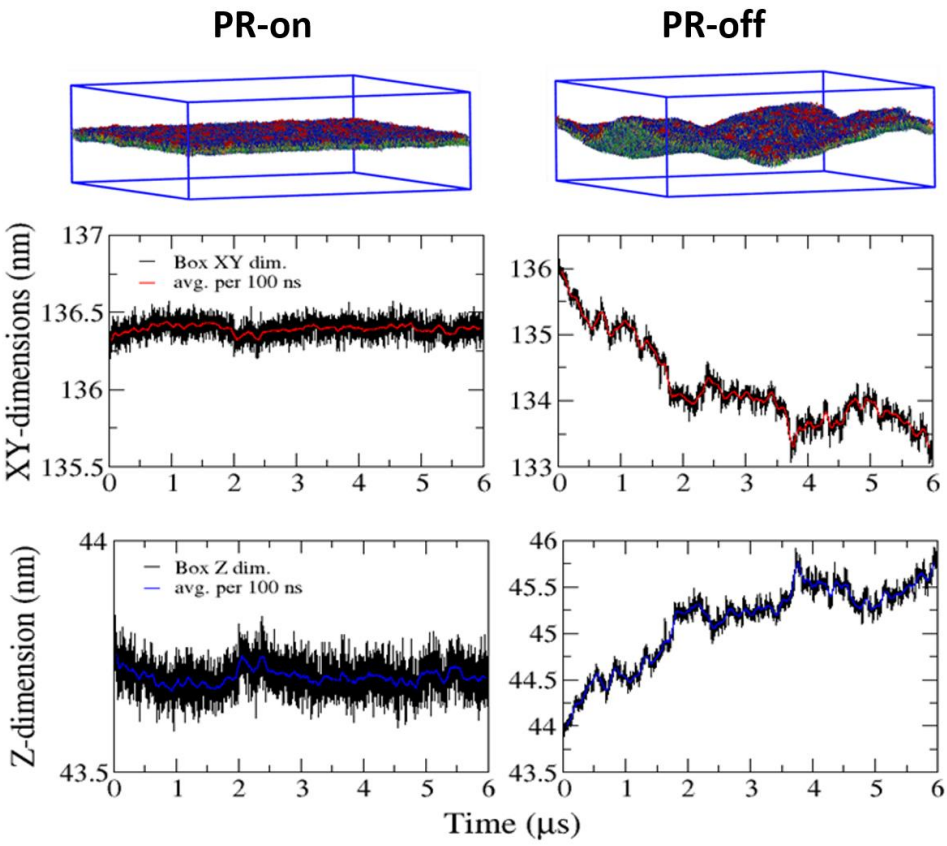

Figure 4-5 Approaches to studying the lateral organization of lipid-protein interactions. A membrane with position restraints (PR-on) and a membrane without position restraints (PR-off). The position restraints were applied to phosphate groups of selected lipids headgroups in the upper leaflet. Lower panels show the effect of the position restraints on the box sizes.

Protein packing in the membrane

After 6 μs simulation, a similar arrangement of integral membrane proteins is observed in both PR-on and PR-off systems (**Figure 4-6 A**). However, slightly more aggregation of the proteins with large extramembrane regions is observed in the PR-off systems compared with the initial distribution of proteins in the membrane. The extramembrane regions of these two proteins started to interact early in the simulations. Small proteins or proteins with minimal extramembrane domains such as DAT (yellow) and AQP1 (red) diffuse in the plane of the membrane and do not aggregate after 6μs (**Figure 4-6 A**).

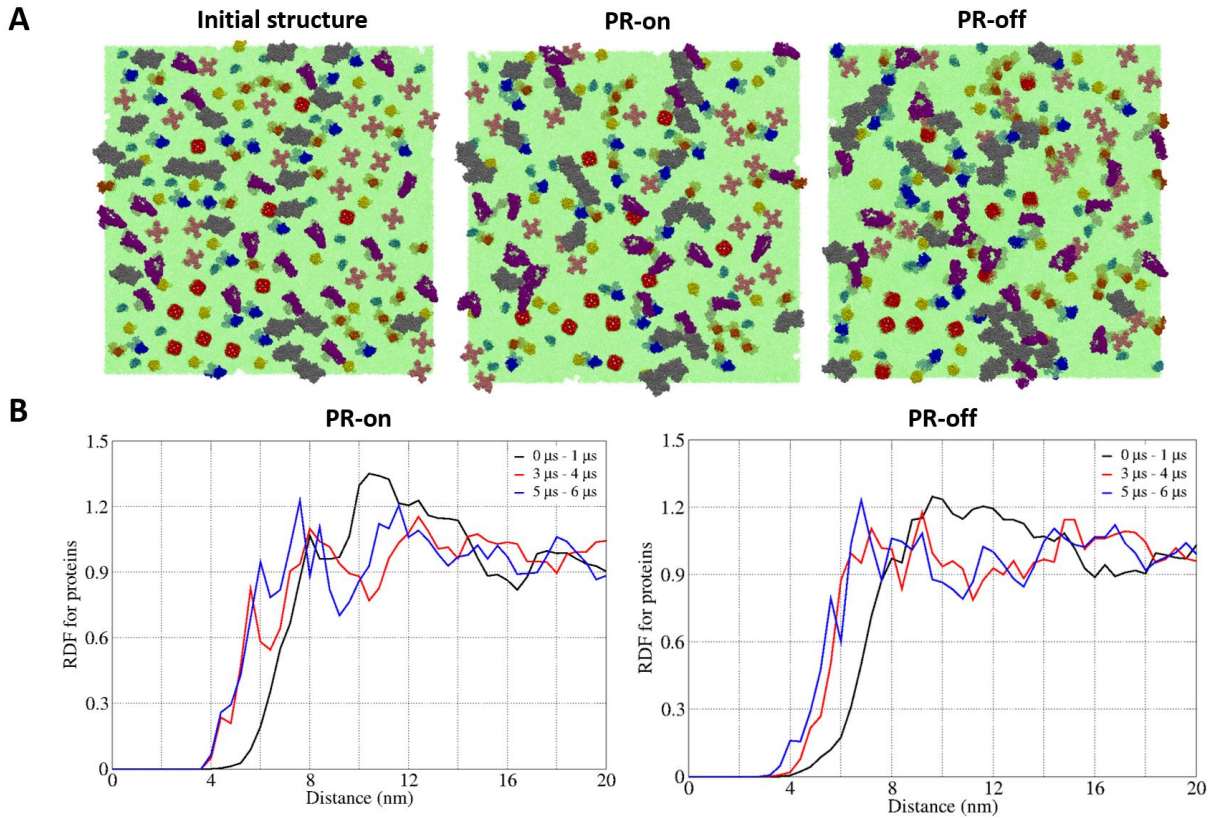

Figure 4-6 Protein packing in the membrane. Initial (0 µs) and final configuration (6 µs) of the simulation systems. **A)** Different protein families are shown in different colors : AQP1 (red), Kv1.2 (pink), P-gp (orange), GLUT1 (cyan), DAT (yellow), Na,K-ATPase (blue); proteins with large extramembrane regions: EGFR (purple), GluA2 (gray) ; lipids are in green. **B)** Radial Distribution Function (RDF) of the center of mass (COM) of the proteins, calculated for over three time intervals.

To quantify the protein packing on the membrane, the RDF of the center of mass (COMs) of the proteins was calculated over three time windows (**Figure 4-6 B**). At the beginning of the simulations, proteins were predominantly found around 10 nm from each other. A gradual shift of the peaks to the left in both graphs during 3 to 4 µs corresponds to a closer packing of the proteins (**Figure 4-6 B**). Small changes in the organization of proteins between the last time windows (3 µs – 4 µs, 5 µs – 6 µs) were observed in the systems.

Location of the proteins regarding the undulation of the membrane

To investigate the distribution of the proteins with respect to the membrane deformation in the PR-off system, changes in the z positions of the lipid head groups were calculated and visualized as described in [94, 106]. Snapshots of the system at 0 µs, 3 µs, and 6 µs illustrate how membrane deformation evolves as a function of time (**Figure 4-7 A**). Starting from an initial z value of 25 nm, the membrane was deformed up to approximately 10 nm on each side (**Figure 4-7 A**).

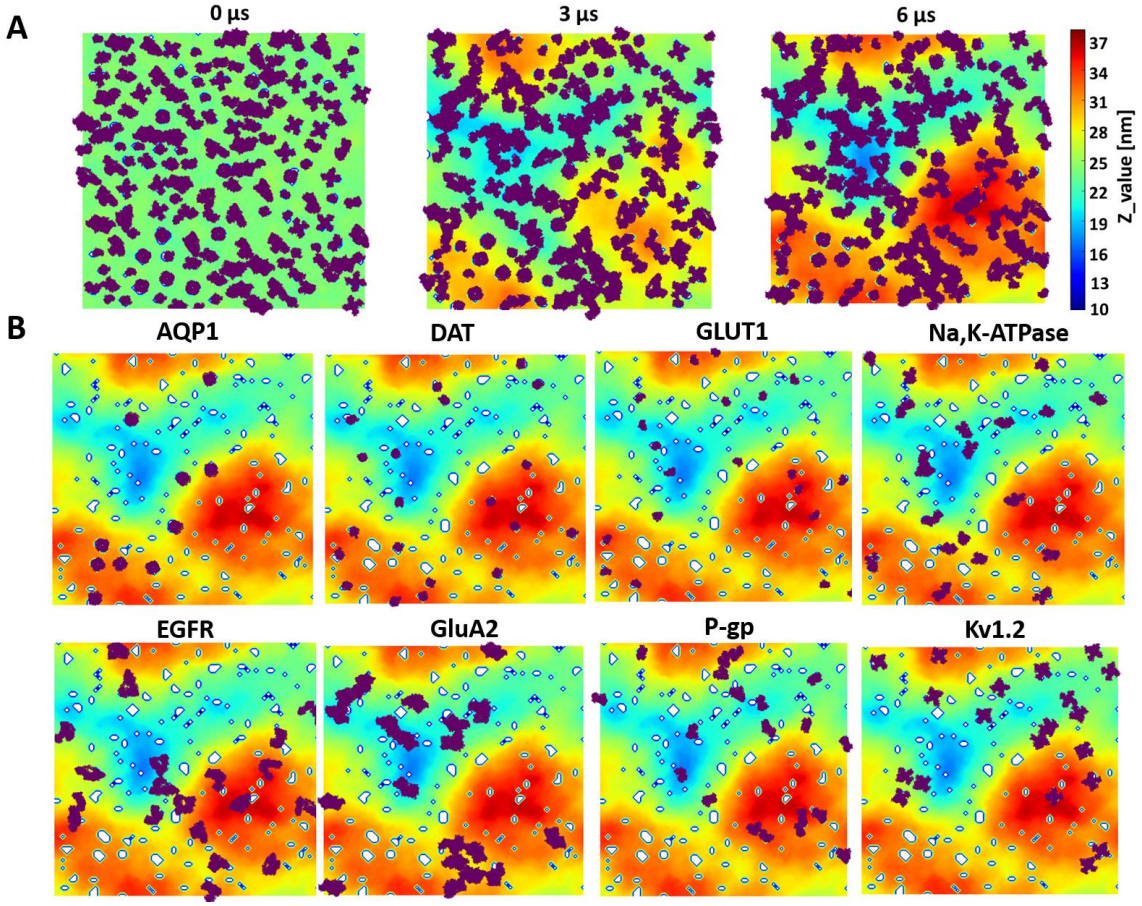

Figure 4-7 Snapshots of the membrane undulation of the PR-off system. The color bar represents the z values of the PO4 beads of lipids in the upper leaflet. Initial values are at ca. 24 nm. **A)** Snapshots show the evolution of the membrane undulation over time. **B)** Graphs represent the distribution of each protein family after 6 µs of simulation; white spots depict the rest of the proteins that are not shown explicitly.

The distribution of protein copies of the same protein in the last frame of the simulation is shown separately (**Figure 4-7 B**). AQP1 molecules, with few exceptions, distribute along the boundaries between high and low z values. DAT, GLUT1, EGFR, and P-gp are distributed across the simulation box independently from the magnitude of membrane deformation. Kv1.2, GluA2 and Na,K-ATPase mainly prefer the lower z values of the bilayer.

Local lipid environment around proteins

To test whether the neighboring lipids extracted with the proteins from the small systems were still attached to the protein after 6 μs, the number of lipids of selected species around four selected proteins (AQP1, DAT, EGFR, Kv1.2) were counted as function of time. This analysis was performed on the PR-on system, for the lipid species GM and PIP lipids that were within 0.7 nm and 2.8 nm from the proteins. These cut-off values were chosen to evaluate approximately the first and four layers of lipids around the proteins.

Overall, the fluctuations of the number of lipid around an average value from the beginning of the simulation at 0.7 nm cut-off shows that the lipid composition of the first shell of lipids from the proteins is preserved over the simulated period. However, it does not mean that the lipids are always in the same spot, as lipids exchanges are observed over the course of the simulation. For the 2.8 nm cut-off, a slight increase in the number of GM lipids around the EGFR and Kv1.2 proteins over time was observed, while the number of PIP lipids around the protein did not increase at the higher cut-off.

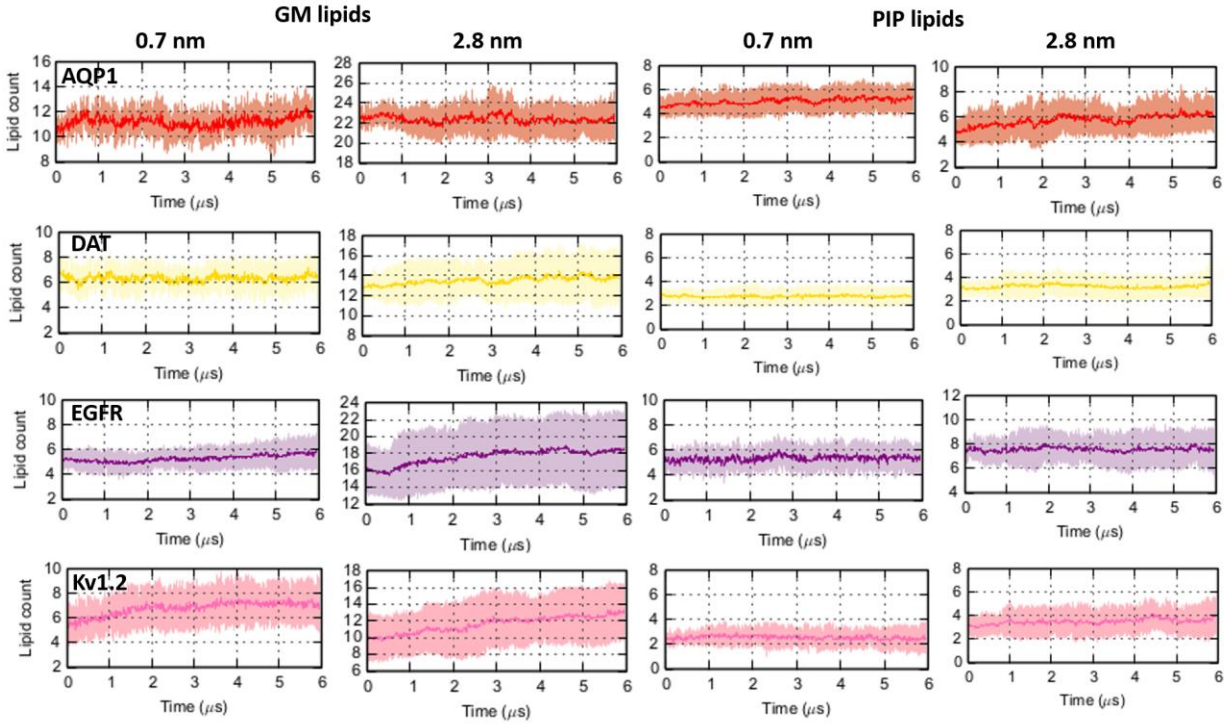

Figure 4-8 Lipid count. The number of GM and PIP lipids in contact with the AQP1, DAT, EGFR, Kv1.2 proteins, as a function of time, was approximated by the number of GM1 (for GM lipids) and CP (for PIP lipids) beads found within a 0.7 nm and 2.8 nm cut-off from the proteins. The color of the graphs corresponds to the protein coloring in **Figure 4-6**. Averages over multiple copies of each protein are illustrated with the darker lines; the lighter lines indicate the standard error in the average.

4.4 Discussion

Protocol to build large-scale biological membrane models

The most relevant outcome of this study is the development of a protocol to generate initial structures of large-scale membrane models suitable for MD simulations. Consisting of receptors, transporters, channels and an enzyme, all embedded in a heterogeneous and asymmetric lipid bilayer, our plasma membrane models are an improvement over recently published models [91, 92, 94] and demonstrate the effectiveness of the protocol to generate

them. A detailed explanation of how the plasmabuilder works and the folders, files, and steps required to generate these models are described in **Appendix A**.

In principle, the methods used to prepare models containing multiple proteins and lipids for simulation should not matter, if an extensive sampling is performed and the findings can reproduce experimental data. However, reaching adequate sampling of large-scale systems is still a major challenge [89, 91, 94]. One goal of the approach presented here was to reduce the equilibration time needed to equilibrate the interactions between lipids and proteins. For this reason, proteins with neighboring lipids are extracted from small systems that have achieved better sampling [79]. Neighboring lipids within a 0.7 nm cut-off from each protein were selected, although the user can choose any other cut-off value. Our approach based on extracting the proteins with the surrounding lipids is supported by several studies, which highlight that the dynamics of proteins and lipids in the membrane should be considered together [92, 111]. Niemelä et al. [111] demonstrated by atomistic and coarse-grained simulations that membrane proteins diffuse with various lipids around them. Although it was observed that the lipid perturbation due to interaction with the protein extends up to 2 nm from the outermost surface of the protein, the first layer of lipids in contact with the protein was the most affected.

The orientation and insertion of multiple membrane proteins in lipid bilayers are always challenging [107, 112]. It can be even more complicated if membrane proteins have transmembrane regions with different shapes, as it is the case of the proteins inserted in the models built here (**Figure 4-1 A**). One advantage of using the plasmabuilder tool is that proteins do not need to be placed in the lipid bilayer manually. The surrounding lipids of the proteins define the protein position and orientation in the membrane. Therefore, with the plasmabuilder, it is possible to insert multiple types of proteins into a membrane at once. Another common problem that the plasmabuilder solves is that lipids of the membranes are not deleted while inserting the proteins. On the contrary, lipids are placed around the inserted proteins until reaching a defined lipid composition for the upper and lower leaflets. Thus, the desired heterogeneity and lipid asymmetry across the bilayer is not affected.

One aspect that needs to be improved is how the plasmabuilder calculates the area occupied by proteins. The current version defines the area of the bilayer occupied by proteins considering that a single α-helix in a membrane occupies approximately 1.5 nm^2, as estimated from the three-dimensional structures of the helical membrane proteins aquaporin-1 and the ClC chloride channel [109]. We know that all proteins do not satisfy this approach and it might result in underestimating the total area covered by a given protein. For instance, α-helices in EGFR dimer occupy a bigger area because its transmembrane helices are crossed and not stretched as the case of aquaporin. Another issue regarding the determination of the area coverage by proteins is that the extramembrane regions are typically not symmetric with respect to the plane of the membrane. This is the case, for instance, for GluA2 and P-gp, in which the cross-sectional area occupied by these proteins are different in the upper and lower leaflets.

One solution to estimate the protein crowding on the membranes is to consider the lipid/protein ratio instead of the cross-sectional area occupied by the proteins. It has been reported that a lipid/protein ratio of ca. 200 per leaflet mimics the protein crowding found in a eukaryotic cell [113, 114]. In our models, the lipid/protein ratio is ca. 200, therefore, the models generated in this study allow the characterization of the dynamics of protein-lipid and protein-protein complexes under the crowded conditions of a realistic membrane.

Simulation of a eukaryotic plasma membrane model

The simulations presented in this chapter illustrate both the approach to build membrane models and the stability of the large-scale systems over the simulation time studied. It is known that simulations on the order of 10 microseconds are by far not enough to equilibrate systems of this size (greater than 100 nm in X and Y), or even to observe an adequate amount of protein reorganization [91, 92, 94, 115]. Therefore, given the short simulation time of 6 μs, our intention is not to make a relevant biological conclusion about the lipid-protein interplay in these systems, but starting to appreciate the complexity of the lateral dynamics of the membrane components and set the stage for future analyses. The main point

of the following section is to get a glance at the performance of our models. The preliminary results of our simulations are compared with similar systems simulated using the Martini CG force-field.

In both models, PR-on and PR-off, the integral proteins tend to oligomerize early in the simulation (**Figure 4-6**). No apparent changes in the displacement of the proteins are observed after the last 3 μs of the simulations. Interactions between the large domains outside the membrane of EGFR and GluA2 may be responsible for this behavior (**Figure 4-1 A**). Once they come together, no dissociation event is observed during the time studied. The diffusion of such oligomers in the bilayer takes a longer time compared with smaller or individual proteins. Therefore, despite the short simulation time, the inability of these oligomers to freely diffuse might act as a picket fence and affect the displacement of the rest of the proteins in the membrane. This hypothesis is in line with a previous study carried out using Martini model and designed to evaluate the effect of the cytoskeletal on membrane protein diffusion [91]. To represent proteins anchored to the cytoskeleton, the authors applied restrictions to some selected proteins, inhibiting their displacement in the lateral plane of the membrane. As a result, a reduction of the diffusional mobility of both proteins and lipids was observed, in addition to the compartmentalization of the membrane by the restrained proteins [91].

A recent work by Javanainen et al. (2017) [92] shows that the diffusion of proteins in protein-poor conditions are entirely different from a membrane crowded with proteins. The study was carried out using the Martini model in a single lipid membrane consisting of DPPC lipids. Proteins were chosen to have a cylindric shape and minimal extramembrane domains. Calculation of the diffusion coefficient in simulations of 100 μs allowed the authors to conclude that the protein diffusion coefficient has an inverse relationship to the lateral radius of the protein. However, due to the approximation of taking proteins with minimal extramembrane domains embedded in bilayers composed of DPPC only, unanswered questions remain. For instance, what would happen if proteins with large extramembrane regions were added to the models, or the lipid composition complexity increased? The models generated in this thesis better represent a membrane of a living cell and therefore are a better choice to answer these

questions. Further research about this topic using our models will allow us to investigate the effect of the extramembrane domains in protein diffusion under more realistic conditions of protein crowding and lipid composition.

Undulations of the membrane have been reported in other large-scale simulations, in both membranes crowded with proteins and protein-free membranes [91, 94]. While the membranes free of proteins undergo larger undulations, the presence of proteins tends to decrease these membrane fluctuations. We do not have a value of the deformation along the bilayer normal of our membrane free of proteins. However, the initial membrane deformation of our models (ca. ± 10 nm along the bilayer normal) is in accordance with recent simulation studies in membranes with similar lipid composition and protein crowding [91, 115]. Overall, this result agrees with the general view that in membrane models, an increase in the degree of protein crowding and clustering flattens the lipid bilayer.

Qualitatively bigger aggregation of proteins, mainly EGFR and GluA2, are observed in the PR-off system compared to the PR-on system (**Figure 4-6 A, Figure 4-7**). We speculated that undulation of the membrane triggers the oligomerization of proteins by their transmembrane regions. However, longer simulation times are needed to make further conclusions about this topic, especially to observe whether these protein-protein interactions are transient over the simulation time. One interesting analysis to pursue further is the evaluation of the distribution of proteins in the membrane. We tried to find a pattern regarding the preference of the proteins per family for regions of specific curvature, as mentioned above, the simulation time is not enough to proceed with this analysis.

Visual inspection of the 6 μs trajectories shows that lipids are shielding protein-protein interactions and in several cases act as bridges between proteins (**Movie 2**). The number of GM and PIP lipids corresponding to the first shells of lipids around four selected proteins (APQ1, DAT, EGFR, Kv1.2) fluctuates around initial average values, which are stable over the simulation (**Figure 4-8**). It is known that lipids such as GM in the upper leaflet and PIP in the lower leaflet tend to aggregate both by themselves and also near proteins [79, 89]. Analyses of the small simulations, from which the proteins with the neighboring lipids were extracted, show that GM

lipids form clusters around all proteins. Consequently, it is expected that proteins inserted in the larger systems are surrounded by a high number of GM lipids. In the simulation presented here, an increase in the GM lipids around EGFR and Kv1.2 at 2.8 nm cut-off is explained by their association with other proteins (**Figure 4-8**). However, the glycolipid-protein interactions should be carefully evaluated for these models. Recently, Gu et al. [116] reported that glycolipid-protein interactions are more prevalent in CG simulations than in the atomistic simulations. The authors proposed modified Martini parameters for gangliosides that better represent the sizes and structures of ganglioside clusters observed at the atomistic level, which may affect future results on GM-protein interactions simulated with Martini.

Challenges, limitations and assumptions for generating membrane model

One of the main challenges in building realistic membrane models is the selection of the lipid and protein composition. Experimental information about the exact lipid and protein composition of eukaryotic plasma membranes is still lacking. In our models, we use almost the same number of protein copies for all species, despite knowing that certain types of membrane proteins have higher copies numbers than others. We also know that these proteins (receptors, transporters, enzymes, and channels) do not occur together in the plasma membrane of any particular cell. Thus, the study of protein aggregation on the models generated here must be interpreted with caution, at least until the model consists of different proteins that are known to aggregate together. Due to the elastic network, protein conformation cannot be studied using CG Martini simulations (**Chapter 1**), which in turn may affect their interactions with other proteins as well as lipids [47]. Our models do not have cytoskeleton elements, and the possible glycosylation of proteins is ignored.

Another limitation for MD simulations in general, even those employing CG methods, is that the timescales achievable for large-scale simulations are too short to reach an adequate sampling of the phase space. We know that achieving equilibrium or steady state is unlikely for large-scale systems. One common approach to obtaining more sampling of a simulation is generating multiple replicates of the system and simulate them for a shorter period of times. These shorter simulations might sample other regions of the phase space, providing useful

information of the lipid-protein interplay. Even when full equilibration will not be achieved, trends in protein reorganization might become more obvious. Therefore, several replicates of the system studied should be performed to increase the sampling. A balance between the timescale of the simulation and the number of replicates will need to be defined.

4.5 Future directions

Despite the assumptions and limitations, the plasmabuilder methodology will accelerate the building process of membrane models, which will be useful to characterize the dynamic nature of protein-lipid and protein-protein complexes. To further improve the scope of these models, next simulations should be performed with the proposed Martini parameters for glycolipids. In the current version of the plasmabuilder, the last snapshot of each small simulated system is used to extract the proteins with the neighboring lipids. Future work should be focused on considering different patches of the protein+neighboring lipids to avoid the same lipid organization around the proteins. This can be done by extracting different frames of the protein with the surrounding lipids, instead of only considering the last frame. INSANE is a very well-known tool to generate CG models [75]. The final goal is to merge the capabilities of the plasmabuilder and INSANE, as having both tools together will be the best for users of the Martini CG models.

Chapter Five: General Conclusions

Chapter 3 extends our knowledge of the interaction between the bc_1 complex from *R. sphaeroides* and the lipids constituting the chromatophore membrane. The lipid reorganization around the bc_1 complex of purple bacteria was studied and characterized on a multi-microsecond time scale. A high density of PE and PG lipids was found in the proximity of the protein and the dimer inner cavity. The enrichment of PG lipids on the cytoplasmic side of the protein is driven by the presence of Arg and Lys residues on the protein surface. The lipid bilayer around the bc_1 complex remains flat over the simulation time, and in some regions, the bilayer thickness is smaller. Overall, these results will be used to improve the atomistic chromatophore model by representing a more accurate distribution of lipids around the bc_1 complex.

Chapter 4 summarizes the progress towards simulating the lateral organization of biological membranes. We demonstrated the feasibility of a protocol to build arbitrary large-scale plasma membrane models from pre-equilibrated lipid patches around proteins. This outcome eliminates some limitations of current tools and allows the construction of complex membrane models that mimic the protein crowdedness and lipid diversity found in living cells. Although the models described here remain a significant simplification of a living cell membrane, they are more realistic than previous models. The models generated by the plasmabuilder reproduce some elements of the complexity of real membranes, such as protein crowding, the presence of proteins with extramembrane domains, and lipid asymmetry. These models will be a starting point for future simulation studies to investigate how proteins and lipid diffusion are affected in complex environments and will facilitate the modification of lipid mixtures to mimic specific membranes.

5.1 Outlook

The similarity of membrane models to living cell membranes determines how well molecular dynamics simulations imitate the dynamic behavior of lipids and proteins in cell membranes. We are still far from being able to represent an entire cell in full physiological details. This is not only because of the challenges encountered to simulate biological processes that take place at a different time and length scales but also due to the lack of information about real cell components, which vary over time and change in ways that can affect the fate of the cell. However, science keeps moving forward, and the next goal is to represent the complexity of real cell membranes. The research presented in this thesis moved towards achieving this goal.

The first model studied in this thesis represents the heterogeneous and asymmetric lipid composition found in chromatophore membranes. The CG simulation provides a detailed insight into the preferential distribution of lipids around the bc_1 complex, which is valuable information for ongoing research focused on simulating a simple organelle whose components are all known.

Membrane models created with the protocol described in Chapter 4 will open a large number of potential avenues for future research. These models (box size ~ 150 nm x 150 nm in X and Y) together with their extended simulation times will be useful to complement techniques such as super-resolution microscopy, aiming to close the gap between experiments and simulations in complexity and length scales. We will be able to represent specific eukaryotic membranes when information about their membrane components becomes available. They will provide a platform for further protein-lipid interactions studies in different cell types, enabling the identification of new pharmacological targets. They will also be relevant for biomedical and biotechnological applications in the study of drug and gene delivery into eukaryotic cells.

Bibliography

1. Escribá, P.V. and G.L. Nicolson, *Membrane structure and function: relevance of lipid and protein structures in cellular physiology, pathology and therapy.* Biochimica et biophysica acta, 2014. **1838**(6): p. 1449-1450.
2. Chabanon, M., J.C. Stachowiak, and P. Rangamani, *Systems biology of cellular membranes: a convergence with biophysics.* Wiley Interdisciplinary Reviews: Systems Biology and Medicine, 2017.
3. Holthuis, J.C.M. and A.K. Menon, *Lipid landscapes and pipelines in membrane homeostasis.* Nature, 2014. **510**(7503): p. 48-57.
4. Escriba, P.V., et al., *Membrane lipid therapy: Modulation of the cell membrane composition and structure as a molecular base for drug discovery and new disease treatment.* Progress in Lipid Research, 2015. **59**: p. 38-53.
5. Marrink, S.J. and D.P. Tieleman, *Perspective on the Martini model.* Chemical Society Reviews, 2013. **42**(16): p. 6801-6822.
6. Ingolfsson, H.I., P. Tieleman, and S. Marrink, *Lipid Organization of the Plasma Membrane.* Biophysical Journal, 2015. **108**(2): p. 358A-358A.
7. Ingolfsson, H.I., et al., *Computational 'microscopy' of cellular membranes.* J Cell Sci, 2016. **129**(2): p. 257-68.
8. Marrink, S.J. and D.P. Tieleman, *Perspective on the Martini model.* Chem Soc Rev, 2013. **42**(16): p. 6801-22.
9. Marrink, S.J., et al., *The MARTINI force field: Coarse grained model for biomolecular simulations.* Journal of Physical Chemistry B, 2007. **111**(27): p. 7812-7824.
10. Marinetti, G. and K. Cattieu, *Lipid analysis of cells and chromatophores of Rhodopseudomonas sphaeroides.* Chemistry and Physics of Lipids, 1981. **28**(3): p. 241-251.
11. Hedger, G. and M.S. Sansom, *Lipid interaction sites on channels, transporters and receptors: recent insights from molecular dynamics simulations.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2016. **1858**(10): p. 2390-2400.
12. Engelman, D.M., *Membranes are more mosaic than fluid.* Nature, 2005. **438**(7068): p. 578.
13. Fagerberg, L., et al., *Prediction of the human membrane proteome.* Proteomics, 2010. **10**(6): p. 1141-1149.
14. Yildirim, M.A., et al., *Drug--target network.* Nature biotechnology, 2007. **25**(10): p. 1119.
15. Nelson, D.L., A.L. Lehninger, and M.M. Cox, *Lehninger principles of biochemistry*. 2008: Macmillan.
16. van Meer, G., *Membrane lipids, where they are and how they behave: Sphingolipids on the move.* Faseb Journal, 2010. **24**.
17. Nicolson, G.L., *Cell membrane fluid–mosaic structure and cancer metastasis.* Cancer research, 2015. **75**(7): p. 1169-1176.

18. Dallner, G., P. Siekevitz, and G.E. Palade, *Biogenesis of endoplasmic reticulum membranes.* The Journal of cell biology, 1966. **30**(1): p. 97-117.
19. Kolega, J., M. Manabe, and T.-T. Sun, *Basement membrane heterogeneity and variation in corneal epithelial differentiation.* Differentiation, 1989. **42**(1): p. 54-63.
20. Klose, C., M.A. Surma, and K. Simons, *Organellar lipidomics—background and perspectives.* Current opinion in cell biology, 2013. **25**(4): p. 406-413.
21. Coskun, Ü., et al., *Regulation of human EGF receptor by lipids.* Proceedings of the National Academy of Sciences, 2011. **108**(22): p. 9044-9048.
22. Lee, A.G., *Biological membranes: the importance of molecular detail.* Trends in biochemical sciences, 2011. **36**(9): p. 493-500.
23. Dawaliby, R., et al., *ALLOSTERIC REGULATION OF GPCR ACTIVITY BY PHOSPHOLIPIDS.* Nature chemical biology, 2016. **12**(1): p. 35.
24. Paila, Y.D. and A. Chattopadhyay, *The function of G-protein coupled receptors and membrane cholesterol: specific or general interaction?* Glycoconjugate journal, 2009. **26**(6): p. 711.
25. Goni, F.M., *The basic structure and dynamics of cell membranes: An update of the Singer-Nicolson model.* Biochimica Et Biophysica Acta-Biomembranes, 2014. **1838**(6): p. 1467-1476.
26. Eggeling, C. and A. Honigmann, *Closing the gap: The approach of optical and computational microscopy to uncover biomembrane organization.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2016. **1858**(10): p. 2558-2568.
27. Ivanova, P.T., et al., *Lipidomics: a mass spectrometry based systems level analysis of cellular lipids.* Current opinion in chemical biology, 2009. **13**(5): p. 526-531.
28. Smolders, K., et al., *An effective plasma membrane proteomics approach for small tissue samples.* Scientific Reports, 2015. **5**.
29. Cartron, M.L., et al., *Integration of energy and electron transfer processes in the photosynthetic membrane of Rhodobacter sphaeroides.* Biochimica Et Biophysica Acta-Bioenergetics, 2014. **1837**(10): p. 1769-1780.
30. De Zorzi, R., et al., *Single-particle electron microscopy in the study of membrane protein structure.* Journal of Electron Microscopy, 2015. **65**(1): p. 81-96.
31. Reichow, S.L. and T. Gonen, *Lipid–protein interactions probed by electron crystallography.* Current opinion in structural biology, 2009. **19**(5): p. 560-565.
32. Godin, A.G., B. Lounis, and L. Cognet, *Super-resolution microscopy approaches for live cell imaging.* Biophysical journal, 2014. **107**(8): p. 1777-1784.
33. Ingolfsson, H.I., et al., *Computational 'microscopy of cellular membranes.* Journal of cell science, 2016. **129**(2): p. 257-268.
34. Schlick, T., et al., *Biomolecular modeling and simulation: a field coming of age.* Quarterly Reviews of Biophysics, 2011. **44**(2): p. 191-228.
35. Pluhackova, K. and R.A. Böckmann, *Biomembranes in atomistic and coarse-grained simulations.* Journal of Physics: Condensed Matter, 2015. **27**(32): p. 323103.

36. Lindahl, E., B. Hess, and D. Van Der Spoel, *GROMACS 3.0: a package for molecular simulation and trajectory analysis.* Journal of molecular modeling, 2001. **7**(8): p. 306-317.
37. Case, D.A., et al., *The Amber biomolecular simulation programs.* Journal of computational chemistry, 2005. **26**(16): p. 1668-1688.
38. Brooks, B.R., et al., *CHARMM: a program for macromolecular energy, minimization, and dynamics calculations.* Journal of computational chemistry, 1983. **4**(2): p. 187-217.
39. Phillips, J.C., et al., *Scalable molecular dynamics with NAMD.* Journal of computational chemistry, 2005. **26**(16): p. 1781-1802.
40. Hess, B., et al., *GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation.* Journal of Chemical Theory and Computation, 2008. **4**(3): p. 435-447.
41. Van Der Spoel, D., et al., *GROMACS: fast, flexible, and free.* Journal of computational chemistry, 2005. **26**(16): p. 1701-1718.
42. Guvench, O. and A.D. MacKerell, *Comparison of protein force fields for molecular dynamics simulations.* Molecular modeling of proteins, 2008: p. 63-88.
43. Berendsen, H.J.C., et al., *MOLECULAR-DYNAMICS WITH COUPLING TO AN EXTERNAL BATH.* Journal of Chemical Physics, 1984. **81**(8): p. 3684-3690.
44. Bussi, G., D. Donadio, and M. Parrinello, *Canonical sampling through velocity rescaling.* Journal of Chemical Physics, 2007. **126**(1).
45. Grossfield, A. and D.M. Zuckerman, *Quantifying uncertainty and sampling quality in biomolecular simulations.* Annual reports in computational chemistry, 2009. **5**: p. 23-48.
46. Ingolfsson, H.I., et al., *Computational 'microscopy' of cellular membranes.* Journal of cell science, 2016. **129**(2): p. 257-68.
47. Periole, X., et al., *Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition.* Journal of Chemical Theory and Computation, 2009. **5**(9): p. 2531-2543.
48. Chavent, M., A.L. Duncan, and M.S. Sansom, *Molecular dynamics simulations of membrane proteins and their interactions: from nanoscale to mesoscale.* Current opinion in structural biology, 2016. **40**: p. 8-16.
49. Arnarez, C., et al., *Evidence for cardiolipin binding sites on the membrane-exposed surface of the cytochrome bc 1.* Journal of the American Chemical Society, 2013. **135**(8): p. 3112-3120.
50. Scholes, G.D., et al., *Lessons from nature about solar light harvesting.* Nature chemistry, 2011. **3**(10): p. 763-774.
51. Blankenship, R.E., et al., *Comparing photosynthetic and photovoltaic efficiencies and recognizing the potential for improvement.* science, 2011. **332**(6031): p. 805-809.
52. Hu, X., et al., *Photosynthetic apparatus of purple bacteria.* Quarterly reviews of biophysics, 2002. **35**(1): p. 1-62.
53. Sener, M., et al. *Visualization of energy conversion processes in a light harvesting organelle at atomic detail*. in *Proceedings of the International Conference on High Performance Computing, Networking, Storage and Analysis*. 2014.

54. Sener, M.K., et al., *Atomic-level structural and functional model of a bacterial photosynthetic membrane vesicle.* Proceedings of the National Academy of Sciences, 2007. **104**(40): p. 15723-15728.
55. Fetisova, Z.G., A.M. Freiberg, and K.E. Timpmann, *Long-range molecular order as an efficient strategy for light harvesting in photosynthesis.* Nature, 1988. **334**(6183): p. 633-634.
56. Cogdell, R.J., A. Gall, and J. Köhler, *The architecture and function of the light-harvesting apparatus of purple bacteria: from single molecules to in vivo membranes.* Quarterly reviews of biophysics, 2006. **39**(3): p. 227-324.
57. Liu, L.-N. and S. Scheuring, *Investigation of photosynthetic membrane structure using atomic force microscopy.* Trends in plant science, 2013. **18**(5): p. 277-286.
58. Gonçalves, R.P., et al., *Architecture of the native photosynthetic apparatus of Phaeospirillum molischianum.* Journal of structural biology, 2005. **152**(3): p. 221-228.
59. Bahatyrova, S., et al., *The native architecture of a photosynthetic membrane.* Nature, 2004. **430**(7003): p. 1058-1062.
60. Kiley, P.J., A. Varga, and S. Kaplan, *Physiological and structural analysis of light-harvesting mutants of Rhodobacter sphaeroides.* Journal of bacteriology, 1988. **170**(3): p. 1103-1115.
61. Scheuring, S., et al., *The architecture of Rhodobacter sphaeroides chromatophores.* Biochimica et Biophysica Acta (BBA)-Bioenergetics, 2014. **1837**(8): p. 1263-1270.
62. Chandler, D.E., et al., *Intrinsic curvature properties of photosynthetic proteins in chromatophores.* Biophysical journal, 2008. **95**(6): p. 2822-2836.
63. Hsin, J., et al., *Self-Assembly of Photosynthetic Membranes.* ChemPhysChem, 2010. **11**(6): p. 1154-1159.
64. Chandler, D.E., et al., *Membrane curvature induced by aggregates of LH2s and monomeric LH1s.* Biophysical journal, 2009. **97**(11): p. 2978-2984.
65. Chandler, D.E., et al., *Light Harvesting by Lamellar Chromatophores in Rhodospirillum photometricum.* Biophysical Journal, 2014. **106**(11): p. 2503-2510.
66. Cape, J.L., M.K. Bowman, and D.M. Kramer, *Understanding the cytochrome bc complexes by what they don't do. The Q-cycle at 30.* Trends in plant science, 2006. **11**(1): p. 46-55.
67. Jones, M.R., *Lipids in photosynthetic reaction centres: structural roles and functional holes.* Progress in lipid research, 2007. **46**(1): p. 56-87.
68. Hasan, S.S. and W.A. Cramer, *Internal lipid architecture of the hetero-oligomeric cytochrome b 6 f complex.* Structure, 2014. **22**(7): p. 1008-1015.
69. van Eerden, F.J., et al., *Molecular dynamics of photosystem II embedded in the thylakoid membrane.* The Journal of Physical Chemistry B, 2016. **121**(15): p. 3237-3249.
70. Wada, H. and N. Murata, *The essential role of phosphatidylglycerol in photosynthesis.* Photosynthesis research, 2007. **92**(2): p. 205-215.
71. Al-Bayatti, K.K. and J.Y. Takemoto, *Phospholipid topography of the photosynthetic membrane of Rhodopseudomonas sphaeroides.* Biochemistry, 1981. **20**(19): p. 5489-5495.

72. Kwa, L.G., et al., *Mutation of a single residue, β-glutamate-20, alters protein–lipid interactions of light harvesting complex II.* Molecular microbiology, 2008. **67**(1): p. 63-77.
73. Esser, L., et al., *Inhibitor-complexed Structures of the Cytochrome bc1 from the Photosynthetic Bacterium Rhodobacter sphaeroides.* Journal of Biological Chemistry, 2007. **283**(5): p. 2846-2857.
74. de Jong, D.H., et al., *Improved Parameters for the Martini Coarse-Grained Protein Force Field.* Journal of Chemical Theory and Computation, 2013. **9**(1): p. 687-697.
75. Wassenaar, T.A., et al., *Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations.* Journal of Chemical Theory and Computation, 2015. **11**(5): p. 2144-2155.
76. Humphrey, W., Andrew Dalke, and Klaus Schulten, *VMD: visual molecular dynamics*, in *Journal of Molecular Graphics & Modelling*. 1996. p. 33-38.
77. Mehmood, S., et al., *Structural and functional basis for lipid synergy on the activity of the antibacterial peptide ABC transporter McjD.* Journal of Biological Chemistry, 2016. **291**(41): p. 21656-21668.
78. Castillo, N., et al., *Free energy of WALP23 dimer association in DMPC, DPPC, and DOPC bilayers.* Chemistry and physics of lipids, 2013. **169**: p. 95-105.
79. Corradi, V., et al., *Lipid-protein interactions are unique fingerprints for membrane proteins.* bioRxiv, 2017.
80. Lee, A.G., *How lipids affect the activities of integral membrane proteins.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2004. **1666**(1): p. 62-87.
81. Lee, A., *Lipid–protein interactions in biological membranes: a structural perspective.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2003. **1612**(1): p. 1-40.
82. von Heijne, G., *The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology.* The EMBO journal, 1986. **5**(11): p. 3021.
83. Dowhan, W. and M. Bogdanov, *Lipid-dependent membrane protein topogenesis.* Annual review of biochemistry, 2009. **78**: p. 515-540.
84. van Klompenburg, W., et al., *Anionic phospholipids are determinants of membrane protein topology.* The EMBO Journal, 1997. **16**(14): p. 4261-4266.
85. Arias-Cartin, R., et al., *Cardiolipin-based respiratory complex activation in bacteria.* Proceedings of the National Academy of Sciences, 2011. **108**(19): p. 7781-7786.
86. Vitrac, H., M. Bogdanov, and W. Dowhan, *Proper fatty acid composition rather than an ionizable lipid amine is required for full transport function of lactose permease from Escherichia coli.* Journal of Biological Chemistry, 2013. **288**(8): p. 5873-5885.
87. Picas, L., et al., *Evidence of phosphatidylethanolamine and phosphatidylglycerol presence at the annular region of lactose permease of Escherichia coli.* Biochimica et Biophysica Acta (BBA)-Biomembranes, 2010. **1798**(2): p. 291-296.
88. Arnarez, C., S. Marrink, and X. Periole, *Molecular mechanism of cardiolipin-mediated assembly of respiratory chain supercomplexes.* Chemical Science, 2016. **7**(7): p. 4435-4443.

89. Ingólfsson, H.I., et al., *Lipid Organization of the Plasma Membrane.* Journal of the American Chemical Society, 2014. **136**(41): p. 14554-14559.
90. Ingólfsson, H.I., et al., *Computational Lipidomics of the Neuronal Plasma Membrane.* Biophysical Journal, 2017. **113**(10): p. 2271-2280.
91. Koldsø, H., et al., *Membrane Compartmentalization Reducing the Mobility of Lipids and Proteins within a Model Plasma Membrane.* The Journal of Physical Chemistry B, 2016. **120**(34): p. 8873-8881.
92. Javanainen, M., et al., *Diffusion of Integral Membrane Proteins in Protein-Rich Membranes.* The Journal of Physical Chemistry Letters, 2017. **8**(17): p. 4308-4313.
93. Wade, R.C., et al., *Impact of Lipid Composition and Receptor Conformation on the Spatio-temporal Organization of μ-Opioid Receptors in a Multi-component Plasma Membrane Model.* PLOS Computational Biology, 2016. **12**(12): p. e1005240.
94. Koldsø, H. and M.S.P. Sansom, *Organization and Dynamics of Receptor Proteins in a Plasma Membrane.* Journal of the American Chemical Society, 2015. **137**(46): p. 14694-14704.
95. Li, J., K.F. Jaimes, and S.G. Aller, *Refined structures of mouse P-glycoprotein.* Protein Science, 2014. **23**(1): p. 34-46.
96. Sui, H., et al., *Structural basis of water-specific transport through the AQP1 water channel.* Nature, 2001. **414**(6866): p. 872-878.
97. Penmatsa, A., K.H. Wang, and E. Gouaux, *X-ray structure of dopamine transporter elucidates antidepressant mechanism.* Nature, 2013. **503**(7474): p. 85-90.
98. Arkhipov, A., et al., *Architecture and membrane interactions of the EGF receptor.* Cell, 2013. **152**(3): p. 557-569.
99. Deng, D., et al., *Crystal structure of the human glucose transporter GLUT1.* Nature, 2014. **510**(7503): p. 121-125.
100. Chen, X., et al., *Structure of the full-length Shaker potassium channel Kv1. 2 by normal-mode-based X-ray crystallographic refinement.* Proceedings of the National Academy of Sciences, 2010. **107**(25): p. 11352-11357.
101. Laursen, M., et al., *Crystal structure of the high-affinity Na+, K+-ATPase–ouabain complex with Mg2+ bound in the cation binding site.* Proceedings of the National Academy of Sciences, 2013. **110**(27): p. 10958-10963.
102. Sobolevsky, A.I., M.P. Rosconi, and E. Gouaux, *X-ray structure, symmetry and mechanism of an AMPA-subtype glutamate receptor.* Nature, 2009. **462**(7274): p. 745-756.
103. Fenalti, G., et al., *Molecular control of [dgr]-opioid receptor signalling.* Nature, 2014. **506**(7487): p. 191-196.
104. Gupta, K., et al., *The 2.0 Å resolution crystal structure of prostaglandin H 2 synthase-1: structural insights into an unusual peroxidase.* Journal of molecular biology, 2004. **335**(2): p. 503-518.
105. Anderson, R.G.W. and K. Jacobson, *Cell biology - A role for lipid shells in targeting proteins to caveolae, rafts, and other lipid domains.* Science, 2002. **296**(5574): p. 1821-1825.

106. Javanainen, M. and H. Martinez-Seara, *Efficient preparation and analysis of membrane and membrane protein systems.* Biochimica et Biophysica Acta (BBA) - Biomembranes, 2016. **1858**(10): p. 2468-2482.
107. Kandt, C., W.L. Ash, and D.P. Tieleman, *Setting up and running molecular dynamics simulations of membrane proteins.* Methods, 2007. **41**(4): p. 475-488.
108. Qi, Y., et al., *CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field.* Journal of Chemical Theory and Computation, 2015. **11**(9): p. 4486-4494.
109. Takamori, S., et al., *Molecular anatomy of a trafficking organelle.* Cell, 2006. **127**(4): p. 831-846.
110. Chavent, M., et al., *Methodologies for the analysis of instantaneous lipid diffusion in md simulations of large membrane systems.* Faraday discussions, 2014. **169**: p. 455-475.
111. Niemelä, P.S., et al., *Membrane proteins diffuse as dynamic complexes with lipids.* Journal of the American Chemical Society, 2010. **132**(22): p. 7574-7575.
112. Jefferys, E., et al., *Alchembed: a computational method for incorporating multiple proteins into complex lipid geometries.* Journal of chemical theory and computation, 2015. **11**(6): p. 2743-2754.
113. Dupuy, A.D. and D.M. Engelman, *Protein area occupancy at the center of the red blood cell membrane.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(8): p. 2848-2852.
114. Jacobson, K., O.G. Mouritsen, and R.G. Anderson, *Lipid rafts: at a crossroad between cell biology and physics.* Nature cell biology, 2007. **9**(1): p. 7-14.
115. Duncan, A.L., et al., *Protein crowding and lipid complexity influence the nanoscale dynamic organization of ion channels in cell membranes.* Scientific reports, 2017. **7**(1): p. 16647.
116. Gu, R.-X., et al., *Ganglioside-Lipid and Ganglioside-Protein Interactions Revealed by Coarse-Grained and Atomistic Molecular Dynamics Simulations.* The Journal of Physical Chemistry B, 2016.
117. Eastman, P., et al., *OpenMM 4: A Reusable, Extensible, Hardware Independent Library for High Performance Molecular Simulation.* Journal of Chemical Theory and Computation, 2013. **9**(1): p. 461-469.
118. Hsin, J., et al., *Protein-induced membrane curvature investigated through molecular dynamics flexible fitting.* Biophysical journal, 2009. **97**(1): p. 321-329.
119. Racine, J., *gnuplot 4.0: a portable interactive plotting utility.* Journal of Applied Econometrics, 2006. **21**(1): p. 133-141.

Appendix A

Plasmabuilder

1. Overview

The primary goal of the plasmabuilder is to provide an initial configuration of membrane models suitable for MD simulation (**Figure A1**). A set of files of simulated systems containing protein+lipids are needed for the plasmabuilder to extract the protein along with lipids within a given cut-off and place them in a larger system.

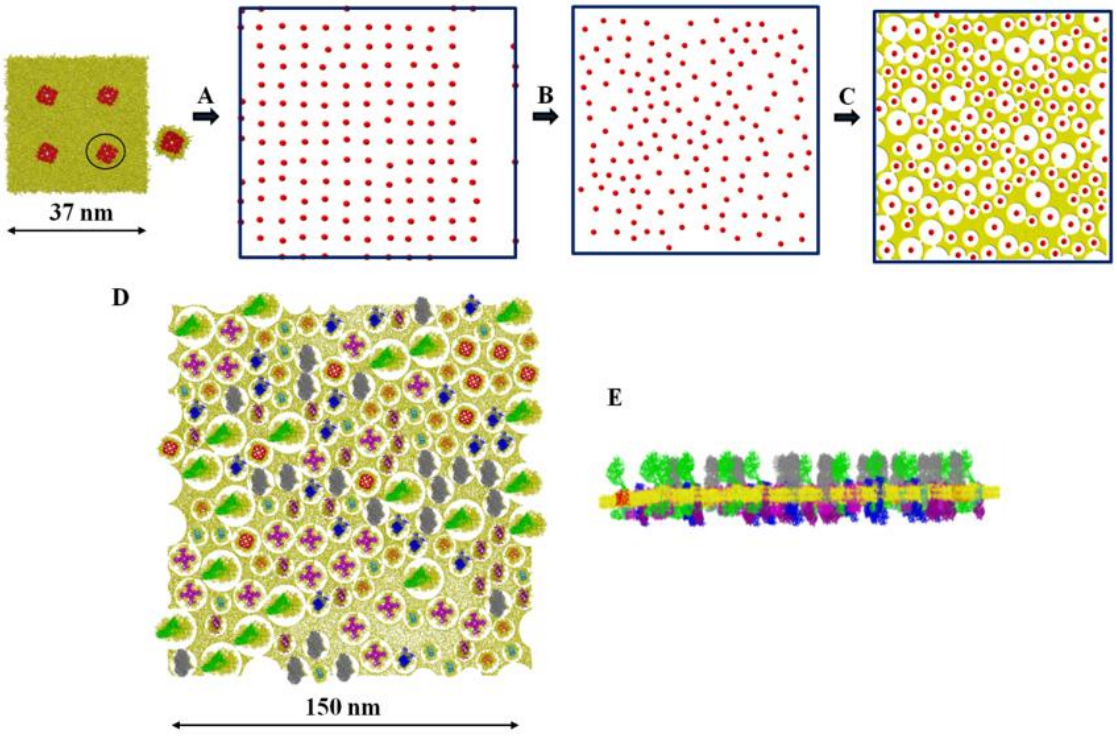

Figure A1 Diagram showing the plasmabuilder workflow. A) The plasmabuilder extracts proteins with lipids within a given cut-off from previously simulated systems and place them in a larger simulation box, according to a specified ratio of proteins as well as lipids. Proteins are represented in red and lipids in yellow. **B)** Spheres that represent the position of proteins and neighboring lipids are distributed randomly over a given space by using 2D Langevin dynamics simulations. **C)** Spheres representing lipids (yellow) are added to the system, and the simulation mentioned above is performed. **D)** Top and side view **E)** of the model. Each sphere was replaced by the respective proteins and lipids. Lipids are represented in yellow and each protein in different colors. Some holes around proteins are observed (**D**).

1. Requirements

Plasmabuilder is written in Python 2.7. GROMACS software version 4.6.7 [40] and OpenMM 6.0.1 [117] are also required to perform the plasmabuilder steps.

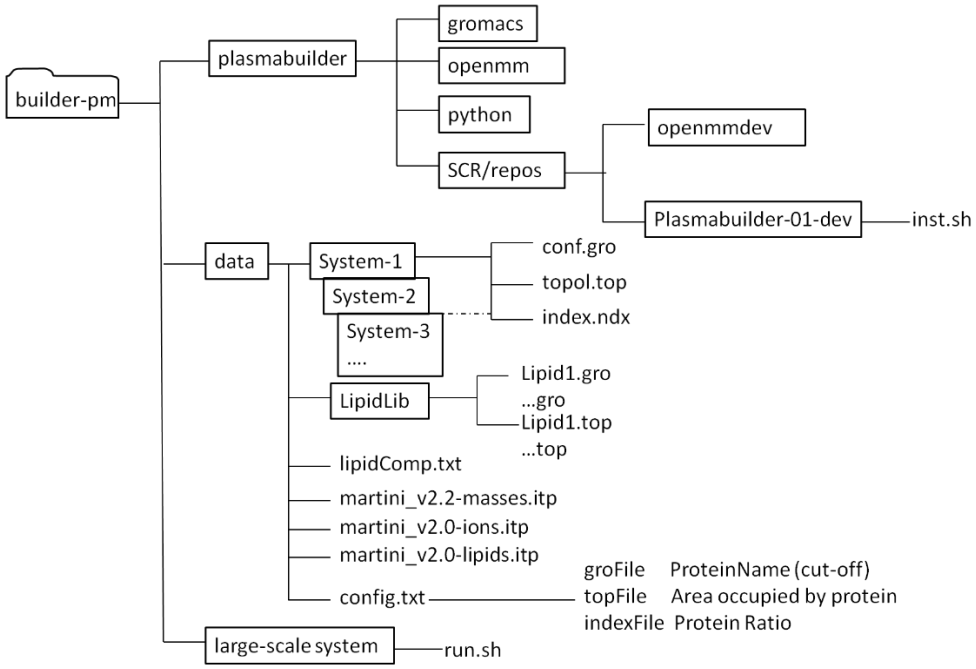

Figure A2. Organization of the folder to execute the plasmabuilder on Linux

2. Installation

The folder "plasmabuilder" contains the plasmabuilder scripts required for the extraction of the proteins and neighboring lipids, and their placement in a larger box.

A helper bash script (inst.sh) is provided for installation. Python 2.7, Gromacs 4.6.7 and OpenMM must be pre-installed and configured. They need to be in the $PATH for the installation to be successful.

3. Input files

The plasmabuilder requires several parameters before building the initial configuration of large-scale systems (**Figure A1**). All these parameters and the initial files are located in the folder "data" (**Figure A2**). Each of these parameters is listed below with example files/values used to build the models presented in this chapter:

1- Coordinates of the systems containing a protein in a bilayer (conf.gro files). These files refer to each small system from which the protein and lipids are extracted. In this study, up to ten different systems were used to generate more complex models.

2- Index files (index.ndx) for each system in which the protein with neighboring lipids at a desired cut-off have been joined as a group. In the models presented here, groups containing one protein and lipids within a cut-off of 0.7 nm around that protein were created.

3- Folder containing the lipid library (lipidLib) and the lipid composition (lipidComp.txt) for the membrane that is generated. Files containing the topology information of each lipid type, and the definition of the particles in the CG force-field (*.itp).

4- In the file config.txt, the following information needs to be provided:

- The area assigned to each protein (float). In this work, it was assumed that a helix occupies 1.5 nm^2 [109]. Therefore, the area for each protein is 1.5 nm^2 times the number of transmembrane helices (TMH) in the protein.
- The relative number of copies to be inserted for each protein (integer). An equal number of copies was used for most proteins.
- Relative number and type of lipids in the upper and lower leaflet of the bilayer. Here, the same lipid composition as previously used to mimic a eukaryotic plasma membrane was chosen [79, 89].
- The percentage of the membrane area to be covered by proteins (float). A value of 14% was chosen, however, as explained in the Discussion, this value is underestimated.

5- To generate the initial system, the file run.sh in the folder “large-scale system” is used.

Available options are as follow:

Example:

```
addplasma.py --outGro system.gro –outTop system.top --lipConfig ../data/lipidComp1.txt \
--protConfig ../data/config4.txt --lipLibPath ../data/lipidLib \ --box 170 170 60 --
percentProtein 14 --APLLower 0.6636 --APLUpper 0.6156 --protSteps 250000 --lipSteps
100000
```

--outGro, --outTop: Output, coordinates and topology files.

--lipConfig, --protConfig: Define the lipid and protein composition.

--lipLibPath: Lipid library from which the rest of lipids are extracted to fill the box.

--box: Define the box size

--percentProtein: Define the percentage of the membrane to be covered by proteins.

--APLLower, --APLUpper: Area assigned to the lipids in the upper and lower leaflets. The values were extracted from the study performed by Ingólfsson et al. [89] and increased by 20%.

--protSteps, --lipSteps: Steps to perform the simulation to distribute the spheres in the box. After these steps the spheres are converted to proteins and lipids.

It is important to emphasize that after the generation of a model with the plasmabuilder some steps of MD simulation, as described in the Methods section, need to be performed to close the holes around the proteins.

Underlying Theory

The procedure performs by the plasmabuilder is as follows:

1- Determine the desired number of copies *c* of each protein:

$$c_i = b_i \left\lfloor \frac{xyp}{\sum_{i=1}^{n} a_i b_i} \right\rfloor$$

where *b* is the relative number of proteins; x,y indicates the box dimensions; *p* is the desired area of the membrane covered by proteins (e.g. 14 %); *a* is the area assigned to each protein (1.5 nm^2 x #TMH).

2- Determine the number of lipids for upper and lower leaflets:

$$d_j = \left\lceil \frac{\lceil ((xy - xyp)/e) \rceil \, f_j}{\sum_{j=1}^{J} f_j} \right\rceil$$

The number of lipids *d* in upper of lower leaflets for a given lipid type is determined by using the desired area per lipid *e*, and the relative ratio *f* of each lipid type; box dimensions x,y; desired protein area *p*.

3- Determine the position of protein and neighboring lipids

Proteins along with the neighboring lipids are extracted using the provided index group. The geometric center of mass (gcom) of the group is removed. The X and Y width of the group is determined. Each molecule is replicated to the desired number. The gcom of each group is placed on a grid covering the system. Each gcom is treated as a sphere with some mass and sigma, and a unit charge. A stiff harmonic potential is applied on Z coordinates, and few steps of Langevin dynamics are performed using OpenMM software package [117]. After simulations, each protein group with its neighboring lipids is translated to the location of the spheres.

4- Determine the position of lipids

The rest of the lipid molecules to fill the two-dimensional area of the membrane are loaded from a library and placed on a grid. By considering the number of lipid molecules that are already inserted with the protein, the number of lipid copies to be replicated is adjusted. In the case where the number of lipids of a given lipid type exceeds the desired number, a warning is issued, and the required number of lipid molecules from the lipid type is deleted from the protein neighborhood. At this point, few steps of Langevin dynamics mentioned above are repeated.

Appendix B

The curvature of the *bc*$_1$ membrane patch was analyzed as described by Hsin et al. (2009) [118]. The authors fitted the coordinates of atoms that represent the shape of a membrane surface using a three-dimensional quadratic equation such as

$$Z(x,y) = a + bx + cx^2 + dy + exy + fy^2 \qquad (1.1)$$

The Z-axis in the equation 1.1 corresponds to the direction perpendicular to the plane of the membrane. From this fitting, the coefficients a, b, c, d, e, and f were obtained and substituted in equations 1.2 and 1.3 to calculate the radii of curvatures the in X and Y axes of the membrane.

$$r_x = \frac{[1+(b+ey+2cx)^2]^{3/2}}{2c} \qquad (1.2)$$

$$r_y = \frac{[1+(d+ex+2fy)^2]^{3/2}}{2f} \qquad (1.3)$$

In this thesis, to quantitatively determine the curvature of the membrane in the *bc*$_1$ system, the coordinates of lipid phosphate groups of the upper leaflet were selected to represent the approximate shape of the membrane. The fitting was carried out using Gnuplot software [119].

The radii of curvature of each point represented by a phosphate group were computed both along the X-axis (R_x) and the Y-axis (R_y) using equations 1.2 and 1.3, considering the coefficients computed from equation 1.1. The curvatures at each point k_x and k_y are the reciprocal of the radii of curvature r_x and r_y.

The radii of curvature of the *bc*$_1$ membrane patch were measured at the last frame of the simulation. Values of 73-76 nm along the X-axis and 231-233 nm along the Y-axis were obtained and correspond to curvatures of approximately 0.01 nm^{-1} and 0.004 nm^{-1}, respectively. These values show that the curvature of the membrane patch is close to zero which supports that the membrane is overall flat as observed in the time frames of the system (**Figure 3-9**). The result agrees with the analysis carried out with the RC-LH1-PufX, which is expected to bend the membrane and the curvature values were calculated using the same approach as presented here [118]. The analysis of the *bc*$_1$ system shows that the curvature

along the Y-axis is smaller than the curvature along the X-axis as can be observed in **Figure B1**.

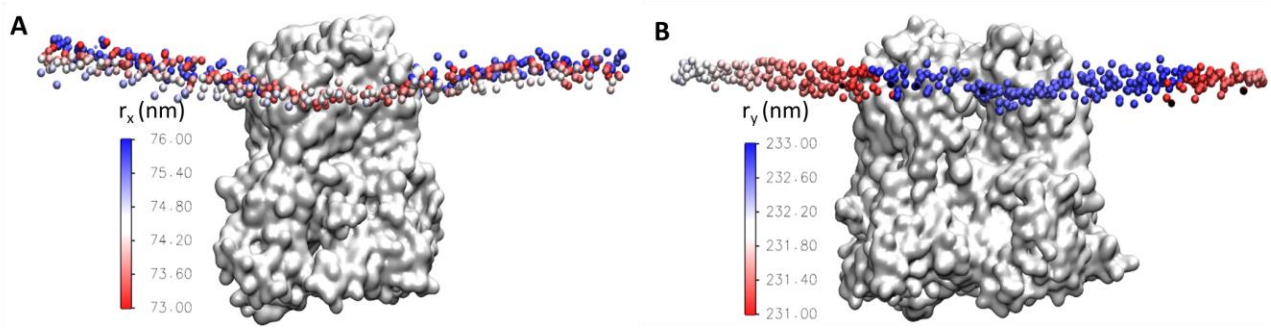

Figure B1 Analysis of membrane curvature shape by the *bc*$_1$ complex. Protein is shown in white surface. Phosphate lipid groups of the upper leaflet are colored to represent the radii of curvature. Color scale bars indicate the range in the radius of curvature along X-axis (**A**) and Y-axis (**B**). The maximum curvature is colored red.

Appendix C

Copyright Permissions

Figure 1.2, left image:

This Agreement between University of Calgary -- Karelia Delgado Magnero ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	4293890759957
License date	Feb 21, 2018
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature
Licensed Content Title	Membranes are more mosaic than fluid
Licensed Content Author	Donald M. Engelman
Licensed Content Date	Nov 30, 2005
Licensed Content Volume	438
Licensed Content Issue	7068
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
High-res required	no
Will you be translating?	no
Circulation/distribution	<501
Author of this Springer Nature content	no
Title	Modeling of Biological Membranes by A Coarse-grained Approach
Instructor name	n/a
Institution name	n/a
Expected presentation date	Mar 2018
Portions	Figure 1.b
Requestor Location	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Department of Biological Science
Billing Type	Invoice
Billing Address	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Department of Biological Science
Total	0.00 CAD

Figure 2.1:

This Agreement between University of Calgary -- Karelia Delgado Magnero ("You") and John Wiley and Sons ("John Wiley and Sons") consists of your license details and the terms and conditions provided by John Wiley and Sons and Copyright Clearance Center.

License Number	4293861048989
License date	Feb 21, 2018
Licensed Content Publisher	John Wiley and Sons
Licensed Content Publication	Wiley Interdisciplinary Reviews: Systems Biology and Medicine
Licensed Content Title	Systems biology of cellular membranes: a convergence with biophysics
Licensed Content Author	Morgan Chabanon,Jeanne C. Stachowiak,Padmini Rangamani
Licensed Content Date	May 5, 2017
Licensed Content Pages	1
Type of use	Dissertation/Thesis
Requestor type	University/Academic
Format	Print and electronic
Portion	Figure/table
Number of figures/tables	1
Original Wiley figure/table number(s)	Figure 2
Will you be translating?	No
Title of your thesis / dissertation	Modeling of Biological Membranes by A Coarse-grained Approach
Expected completion date	Mar 2018
Expected size (number of pages)	86
Requestor Location	University of Calgary 2500 University Dr NW Calgary, AB T2N 1N4 Canada Attn: Department of Biological Science
Publisher Tax ID	EU826007151
Total	0.00 CAD

Figure 3.1 and 3.2:

Dear Karelia,

I am very happy to give you permission for the use of Fig. 3.1 and Fig. 3.2 for your thesis and otherwise in presentations etc.

Best wishes,
Abhishek Singharoy