TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

MARJUT VÄHÄHEIKKILÄ

G PROTEIN-COUPLED RECEPTORS INDUCE DOMAIN FORMATION IN PLASMA MEMBRANE MODELS PROBED BY MOLECULAR DYNAMICS SIMULATIONS

Master of Science thesis

Examiner: Prof. Ilpo Vattulainen
Examiner and topic approved by the Faculty Council of the Faculty of Natural Sciences on 9th September 2015

ABSTRACT

MARJUT VÄHÄHEIKKILÄ: G Protein-Coupled Receptors Induce Domain Formation in Plasma Membrane Models Probed by Molecular Dynamics Simulations
Tampere University of Technology
Master of Science thesis, 70 pages, 2 Appendix pages
April 2016
Master's Degree Program in Biotechnology
Major: Tissue Engineering
Supervisors: Dr. Waldemar Kulig and Prof. Ilpo Vattulainen
Examiner: Prof. Ilpo Vattulainen
Keywords: lipid raft, protein-lipid interactions, lipid-lipid interactions, self-assembly

Lipid domains in plasma membranes are typically nanosized regions with a distinct lipid content. They have been suggested to have multiple functions. For example, the lipid environment can affect the activity of membrane proteins through direct lipid binding or by changing membrane properties. However, the detailed plasma membrane lipid organization and forces driving domain formation have remained elusive, since lipid domains are difficult to study due to their nanoscale size.

In this thesis, we study the effects of three different G protein-coupled receptors (GPCRs) on the lateral and transmembrane lipid distributions. To this end, we use self-assembly through molecular dynamics (MD) simulations to study the lipid environment of the β_2-adrenergic receptor, the μ-opioid receptor, and rhodopsin proteins in plasma membrane models that account for the effect of 16 abundant lipid species.

Our study reveals that GPCRs affect the lateral organization but not the transmembrane distribution of lipids. Polyunsaturated lipids and one monounsaturated lipid, phosphatidylinositol, are preferred to reside at the protein surface over other studied lipids. Thus, we conclude that GPCR proteins can induce domain formation in plasma membrane models. The results provide important insights into domain formation, which is important to better understand the functioning of GPCR proteins that are essential for human health.

TIIVISTELMÄ

MARJUT VÄHÄHEIKKILÄ: G-proteiinikytkentäiset reseptorit saavat aikaan lipidikoostumukseltaan toisistaan poikkeavia solukalvoalueita molekyylidynamiikkasimulaatioiden mukaan
Tampereen teknillinen yliopisto
Diplomityö, 70 sivua, 2 liitesivua
Huhtikuu 2016
Biotekniikan tutkinto-ohjelma
Pääaine: Kudosteknologia
Ohjaajat: Dr. Waldemar Kulig and Prof. Ilpo Vattulainen
Tarkastajat: Prof. Ilpo Vattulainen
Avainsanat: lipidilautta, proteiini-lipidi -vuorovaikutukset, lipidi-lipidi -vuorovaikutukset, itsejärjestyminen

Lipidilautat ovat solukalvojen osia, joilla on muusta solukalvosta poikkeava lipidikoostumus. Lipidilautoilla on useita tehtäviä. Ne voivat esimerkiksi vaikuttaa kalvoproteiinien aktivoitumiseen suoran proteiini-lipidi -vuorovaikutuksen kautta tai epäsuorasti vaikuttamalla solukalvon ominaisuuksiin. Lipidilauttojen tärkeydestä huolimatta niiden tarkka rakenne on tuntematon, sillä lipidilauttojen tutkiminen on haastavaa niiden pienen koon vuoksi. Tässä opinnäytetyössä tutkitaan, kuinka kolme G-proteiinikytkentäistä reseptoria (GPCR) vaikuttaa solukalvon rakenteeseen.

Molekyylidynamiikan (MD) simulaatiot ovat tutkimusmenetelmä, jolla voidaan saada atomistinen tai lähes atomistinen kuva tutkittavasta kohteesta. Tässä työssä tutkitaan β_2-adrenergisen reseptorin, μ-opioidireseptorin ja rodopsiinin lipidiympäristöä MD-simulaatioiden avulla. Kukin itsejärjestyvä solukalvomalli sisältää yhden proteiinin ja 16 tärkeää ja yleistä lipidityyppiä niiden luonnollisessa pitoisuudessa.

Tässä työssä osoitettiin, että tutkitut GPCR-proteiinit vaikuttavat lipidien lateraaliseen järjestymiseen, mutta eivät lipidien jakaantumiseen kalvon eri puolille. Monityydyttymättömät lipidit ja yksi kertatyydyttymätön lipidi, fosfatidyyli-inositoli, olivat keskimäärin lähempänä proteiinien pintaa kuin muut lipidit. GPCR-proteiinit saavat siten aikaan alueita, joilla on toisistaan poikkeava lipidikoostumus. Tämä työ tuo esille näkemystä lipidialueiden syntymisestä, mikä on tärkeää ihmisten terveyteen vaikuttavien GPCR-proteiinien toiminnan ymmärtämiselle.

PREFACE

This Master of Science thesis was conducted in the Biological Physics and Soft Matter Group of the Department of Physics at Tampere University of Technology. The research work was carried out between June 2015 and January 2016. The objective of this thesis was to study the effects of G protein-coupled receptors on the lateral and transmembrane distributions of the most abundant lipids in the plasma membranes. The study was done computationally using Molecular Dynamics simulations. The computational resources were provided by Tampere Center for Scientific Computing (TCSC).

I wish to thank my employer and thesis examiner professor Ilpo Vattulainen for supervising my work and for the opportunity of working in the group since June 2014. I also want to thank Doctor of Philosophy Waldemar Kulig for supervising my work.

I greatly appreciate the company of my coworkers and all the good advice they have given. I also want to thank my family and friends for their support during the years of my studies and during the whole process of this thesis.

Tampere, April 2016

Marjut Vähäheikkilä

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

2D	Two-dimensional
A	Area
APL	Area per lipid
cAMP	Cyclic adenosine monophosphate
CG	Coarse-grained
CHOL	Cholesterol
Cl^-	Chloride ion
C_n	Constant for the Ryckaert-Bellemans potential
DNA	Deoxyribonucleic acid
DOPC	1-(9Z-hexadecenoyl)-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine
DOPE	1,2-di-(9Z-hexadecenoyl)-sn-glycero-3-phosphoethanolamine
DPCE	N-(octadecanoyl)-sphing-4-enine
DPSM	N-(octadecanoyl)-sphing-4-enine-1-phosphocholine
DXSM	N-(tetracosanoyl)-tetracosasphing-4-enine-1-phosphocholine
F	Force
GDP	Guanosine diphosphate
GPCR	G protein-coupled receptor
G protein	Guanine nucleotide-binding protein
GSL	Glycosphingolipid
GTP	Guanosine triphosphate
H	Helix
I	Moment of inertia
k_B	The Boltzmann constant
k_b	The force constant for stretching of bonds
k_θ	The angular force constant
k_ξ	Force constant for improper dihedral harmonic potential
k_ϕ	Force constant for periodic type proper dihedral potential
m	Mass
MD	Molecular dynamics
N	Number of particles
Na^+	Sodium ion
NpT	Normal pressure and temperature

P	Pressure
PAPC	1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphocholine
PAPE	1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphoethanolamine
PAPS	1-hexadecanoyl-2-(5Z,8Z,11Z,14Z-eicosatetraenoyl)-sn-glycero-3-phosphoserine
PBC	Periodic boundary conditions
PC	Phosphatidylcholine
PDB	Protein data bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PODG	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycerol
POPA	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphate
POPC	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphocholine
POPE	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoethanolamine
POPI	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phospho-(1'-myo-inositol)
POPS	1-hexadecanoyl-2-(9Z-octadecenoyl)-sn-glycero-3-phosphoserine
PS	Phosphatidylserine
PUPS	1-hexadecanoyl-2-(4Z,7Z,10Z,13Z,16Z,19Z-docosahexaenoyl)-sn-glycero-3-phosphoserine
$p(v_i)$	The probability function of a velocity v_i
q	Charge
R	Radius
r	Distance
RB	Ryckaert-Bellemans
r_0	Reference bond length
RDF	Radial distribution function
R_g	Radius of gyration
Rho	Rhodopsin
r_{ij}	Distance between particles i and j
SM	Sphingomyelin
T	Temperature
t	Time

T_0	Reference temperature
TCSC	Tampere Center for Scientific Computing
TM	Transmembrane
TUT	Tampere University of Technology
V	Potential energy
v	Velocity
V_C	Coulomb potential
V_{LJ}	Lennard-Jones potential
x	Position

β_2AR	β_2-adrenergic receptor
Δt	Time step
ϵ_0	Vacuum permittivity constant
θ_0	The reference angle
θ_{ijk}	Angle between particles i, j, and k
μ-OR	μ-opioid receptor
ξ_0	Reference improper dihedral angle
ξ_{ijkl}	Improper dihedral angle
τ_p	Time constant for pressure coupling
τ_t	Time constant for temperature coupling
ϕ	Torsion angle
ϕ_0	Reference torsion angle

1 INTRODUCTION

Biological membranes are vital for all living organisms. Plasma membranes surround all eukaryotic cells, separate the cell contents from their surroundings, and regulate everything that goes into or out of the cells [1, 2]. Biological membranes are lipid bilayers hosting membrane proteins. The basic structure of a lipid membrane is simple: lipids are packed into two arrays with lipid tails pointing inwards toward each other, and polar headgroups facing the water phase.

However, natural plasma membranes are astonishingly complex structures composed of an immense amount of different molecules. There can be more than 1 000 different lipid species [3] in any eukaryotic cell and the biological reason for this variety is a major question of biomembrane research. Nowadays, we are beginning to understand the vast amount of different roles and functions of plasma membrane lipids [3]. Lipids are more than only structural building blocks of cell membranes. On the contrary, lipids have multiple functional roles. For example, lipids can affect protein activity [4] through specific and direct binding or by changing bilayer properties.

Experiments have shown that plasma membranes have nanosized domains with distinctive lipid content and membrane properties [5]. These domains have been suggested to be functional and influence, for instance, membrane fluidity, membrane protein trafficking, and protein activity [6, 7]. Lipid domains also affect the bilayer properties [8] such as membrane potential, permeability, shape, and surface charge that can affect the activity of membrane properties. Therefore, the lipid environment of plasma membrane proteins is essential for the proper functioning of the proteins.

Even though there already is evidence for the existence and importance of different lipid domains within plasma membranes, many questions still remain. The exact properties, structures, and biological functions of the domains and the factors that induce compartmentalization are yet to be resolved. Lipid domains have been difficult to study due to their small size that cannot be resolved with conventional

methods, such as light or fluorescence microscopy [5, 9].

G protein-coupled receptors (GPCRs) are the largest family of membrane proteins [10, 11]. The lipid environment around GPCRs is important since lipids can affect the function of these proteins [12]. The proper function of GPCRs is crucial since they have multiple important roles in a human body. The malfunction of different GPCRs can cause multiple diseases, such as heart failure, hypertension, or nephrogenic diabetes insipidus [13]. The understanding of the functioning of GPCR proteins would help to develop new drugs to cure these diseases.

This work studies the lipid environment of G protein-coupled receptors to answer some of the relevant unsolved questions related to lipid domains. The thesis focuses on three GPCR proteins: β_2-adrenergic receptor (β_2AR), μ-opioid receptor (μ-OR), and rhodopsin (Rho). The research questions are the following: Can GPCRs induce domain formation in plasma membrane models, and how GPCRs affect the lateral and transmembrane distributions of different lipid species abundant in the plasma membranes?

To address the research questions, a computational modelling approach called molecular dynamics (MD) is applied. Molecular dynamics simulations are a powerful tool to obtain atomistic or near atomistic information of the studied systems to reveal the properties of biological systems. Therefore, our approach overcomes the previous difficulties related to the small size of the lipid domains. We have built 3×50 systems modeling plasma membranes, each containing one of the three GPCR proteins and 16 abundant plasma membrane lipid species in natural concentrations. Further, to improve sampling and thus the reliability of our simulations, we employ self-assembly starting from a random initial configuration as the method of choice to consider the long-term arrangement of lipids around the studied receptors. The author of this thesis has built the model systems, performed the MD simulations, and conducted the analyses.

This thesis is structured as follows. The biological background information relevant to this thesis is provided in Chapter 2. The features of molecular dynamics simulations are reviewed in Chapter 3. Chapter 4 introduces the used models and analysis methods employed in this study. Chapter 5 presents the most relevant results and Chapter 6 discusses the conclusions of this thesis.

2 BIOLOGICAL BACKGROUND

This Chapter presents an overview of the biological background relevant to this thesis. Firstly, the components of plasma membranes are presented: all proteins and lipids studied in this thesis are introduced. Then, the importance of molecular interactions is reviewed. Finally, the lateral and transmembrane distributions of lipids in plasma membranes are discussed.

2.1 Cells: the basic structures of life

All living things, including plants, animals, and bacteria, are composed of cells [1, p. 1]. Cells are small membrane-enclosed units filled with an aqueous solution. The cells and their aqueous solution are packed with an immense amount of different chemical compounds and molecular structures, such as cell organelles, proteins, lipids, carbohydrates, and deoxyribonucleic acid (DNA). These molecules and compounds enable the fundamental functions of life: growth, energy production, adaptation to the environment, and the ability to create copies of themselves, amongst other things.

The plasma membrane separates the contents of cells from their surroundings [14, p. 3]. It is mainly composed of protein and lipid molecules that constitute a thin and hydrophobic barrier around all cells. The plasma membranes are about 5 nm thick bilayers where lipids are arranged into arrays of two leaflets residing on top of each other. The plasma membrane has many functions. For instance, it is a selective barrier to the free passage of ions and charged molecules, it governs signaling proteins to transfer signals into cells, and it hosts proteins to participate in several reaction pathways acting as enzymes.

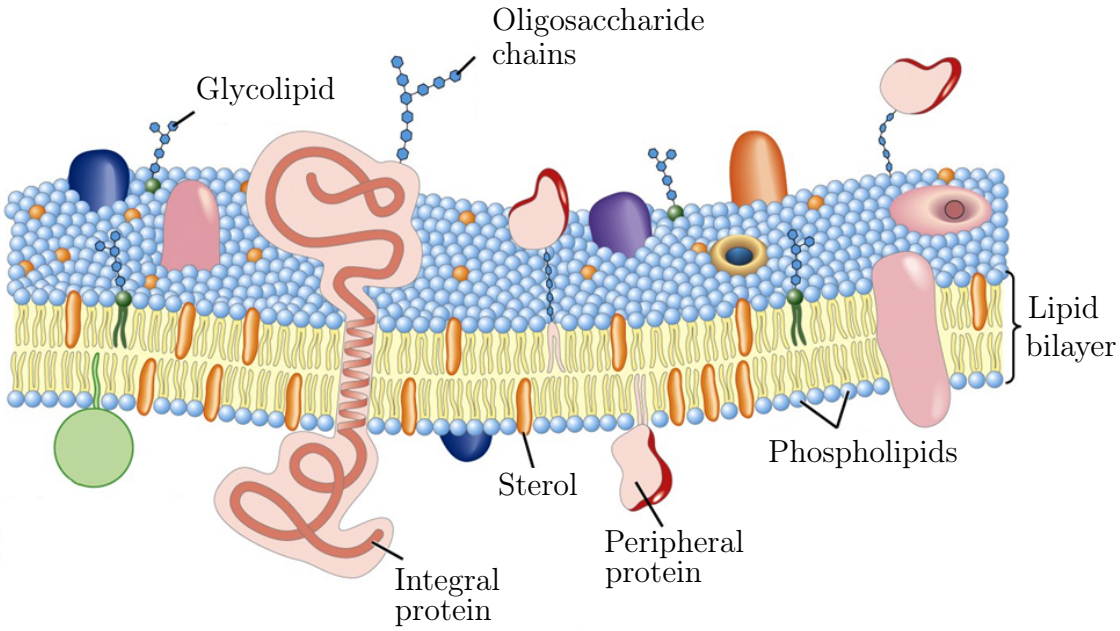


Figure 2.1 *A schematic figure of a plasma membrane with its basic components illustrated. Figure is adapted from [14].*

2.2 Proteins and lipids as the main components of plasma membranes

Proteins and lipids are the main components of the plasma membranes [1, 15], see Figure 2.1. The lipids are arranged into two leaflets that compose a lipid bilayer, where the non-polar lipid tails are pointing inward toward each other and polar heads face the water phase. The plasma membrane proteins can be incorporated into the bilayer (as integral proteins) or they can reside outside the bilayer (as peripheral proteins) linked to the membrane lipids or proteins. There are also carbohydrates attached to the membrane proteins and lipids in the extracellular bilayer surface.

Plasma membranes are extremely crowded [16] with different molecular species. More than circa 23 % of membrane area is occupied by membrane proteins and less than circa 77 % by lipids. The crowding affects many properties of membranes [17], such as lateral mobility of lipids and proteins in the plasma membrane.

2.2.1 Lipids

Lipids are a loosely defined class of biological molecules with a common feature that the molecules are insoluble in water but soluble in fat and organic solvents [1]. Lipids typically contain long hydrocarbon chains or multiple linked hydrocarbon rings.

The most important biological function of lipids is their ability to form membranes. Lipids are generally amphipathic molecules that have a hydrophilic head and two hydrophobic tails [1, 15] but there are also other types of lipids. The hydrophilic head includes a headgroup that is linked to the rest of the lipid molecule. The tails are hydrocarbon chains. The amphipathic property of lipids is crucial in driving lipids to assemble into bilayers in an aqueous solution.

There can be more than 1 000 different lipid species in any eukaryotic cell [3] due to the variation in lipid headgroups and aliphatic chains. The headgroup of a lipid can be for example a choline or an amine group. The charges of the different headgroups vary: some headgroups are neutral whereas some are negative or positive in charge. However, positively charged lipids are not found in nature [18, p. 173]. The number of carbon atoms and therefore the lengths of the lipid tails can differ greatly, and the tails can be saturated or unsaturated. Saturation means that the carbon atoms within the tails are saturated with hydrogen atoms. Thus, saturated tails do not contain any double bonds. Unsaturated lipids have one or more double bonds within the tails. Since the structures of lipid species differ, also the physical and chemical properties of lipid species are different.

Eukaryotic cells invest an eminent amount of resources to generate thousands of different lipids [19]. This suggests that there is an evolutionary advantage [3] to the use of such a complex repertoire of lipids. Indeed, lipids have multiple general functions. Previously, it was thought that lipids would merely be the structural components of biological membranes and energy storages. Nowadays, the more dynamic roles of lipids are being revealed and there is evidence that lipids also have specific functions within plasma membranes. To give a few examples, lipids can act as first and second messengers in signal transduction and molecular recognition processes, and lipids can affect protein activity by directly binding to a protein or by changing the properties of plasma membranes. Additionally, some lipids define functional membrane domains discussed more precisely in Section 2.6. We already understand the functions of numerous lipids but the full utility of the lipid repertoire remains unsolved.

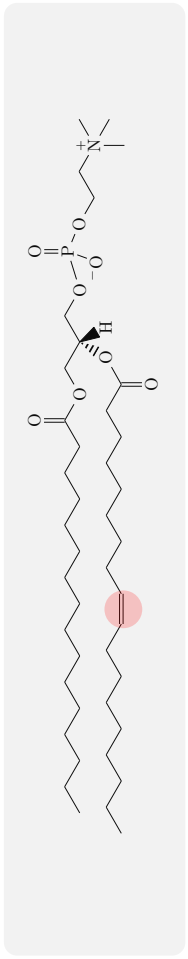
The four major lipids in the plasma membranes are phosphatidylcholine, phosphatidylethanolamine, phosphatidylserine, and sphingomyelin. Additionally, plasma membranes include many other lipids, such as cholesterol, phosphatidic acids, glycerols, and ceramides. The most abundant lipid species in the plasma membranes studied in this thesis are illustrated in Figure 2.2. The lipids include monounsaturated lipids POPC, DPSM, DXSM, POPE, POPI, POPS, POPA, PODG, and DPCE (gray background color in Fig. 2.2). DOPC, DOPE, PAPC, PAPE, PAPS, and PUPS lipids are polyunsaturated (red background color in Fig. 2.2). Additionally, the studied lipids include a sterol: cholesterol. Out of all studied lipids, POPI, POPS, POPA, PAPS, and PUPS are negatively charged, while all other lipids are neutral in charge.

2.2.2 Membrane proteins

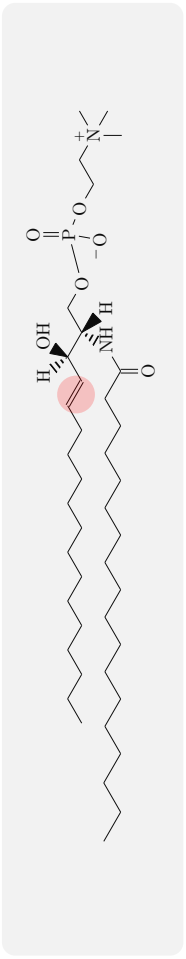
Proteins are macromolecules that consist of one or more chains of amino acids [1, p. 119–121]. Plasma membrane proteins are structures embedded in or attached to lipid bilayers. They constitute approximately 50 % of the total mass of the plasma membranes [1, p. 372] and perform a myriad of different functions within cells. The functions of different proteins arise from the huge number of shapes they can adopt. Proteins can act as enzymes, carry messages, provide mechanical support, transport molecules, regulate genes, act as motor proteins, and act as receptor proteins that detect and transfer signals, among other things.

Membrane proteins can be linked to a lipid bilayer in various ways, and each protein has a unique orientation in a lipid bilayer [1, p. 374]. The lipid-protein interactions affect the orientation and the tertiary structure of the protein within plasma membranes, which in turn can affect the function of the protein. The amino acid segments that run through the hydrophobic environment of a lipid bilayer are mainly composed of amino acids with hydrophobic side chains, whereas the hydrophilic side chains largely face the water phase.

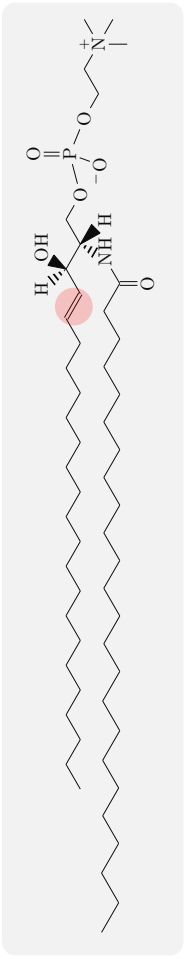
This study concentrates on GPCR proteins. GPCRs are signaling molecules that have versatile functions in cells [11]. These proteins are discussed more precisely in Section 2.3.



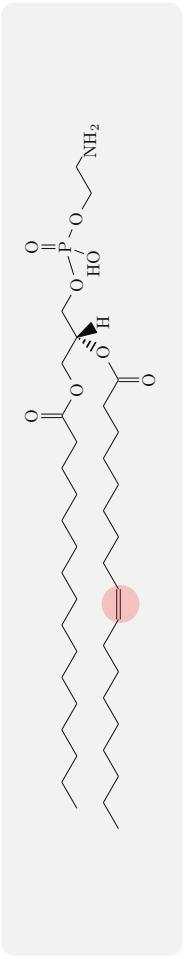
(a) POPC



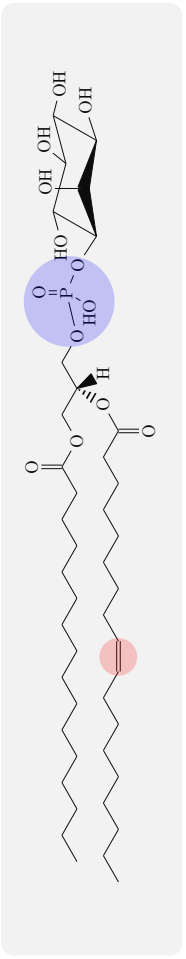
(b) DPSM



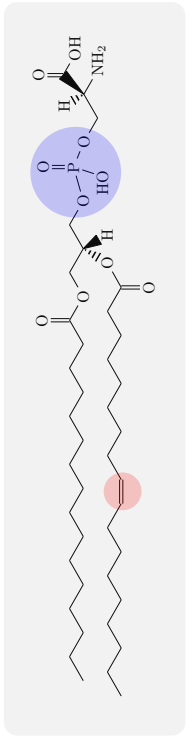
(c) DXSM



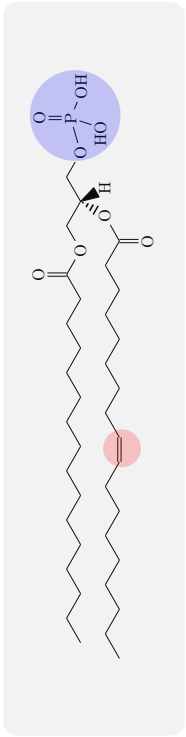
(d) POPE



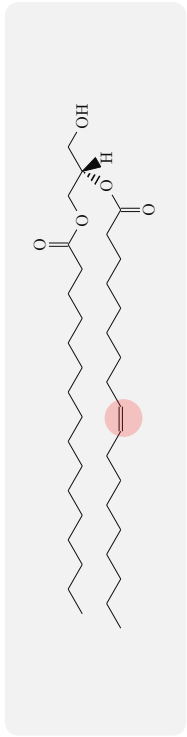
(e) POPI



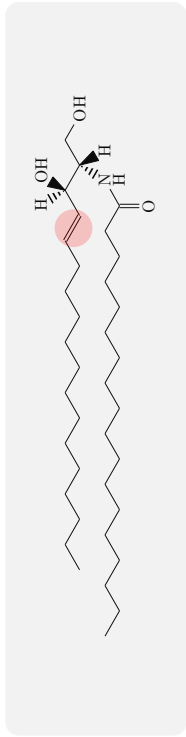
(f) POPS



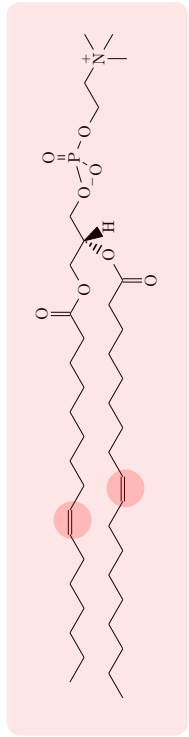
(g) POPA



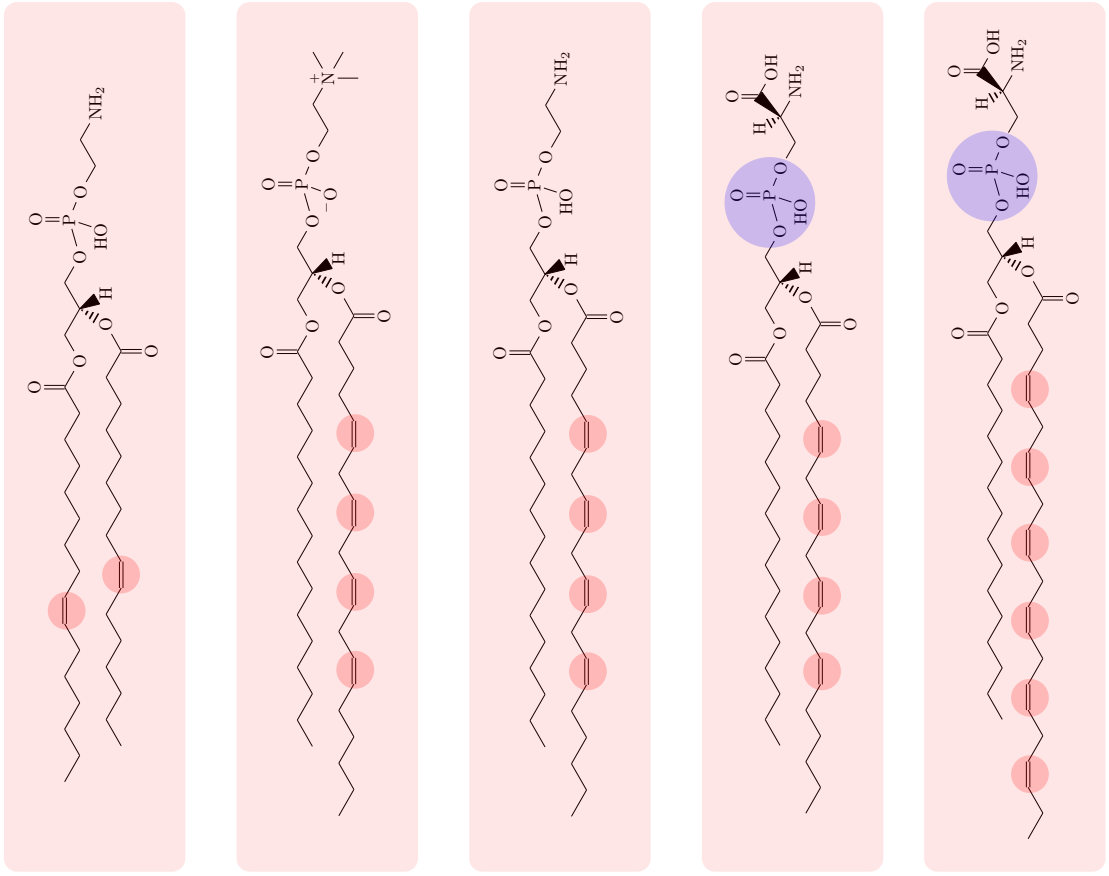
(h) PODG



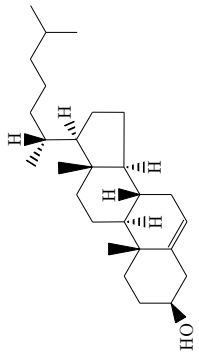
(i) DPCE



(j) DOPC



(k) DOPE (l) PAPC (m) PAPE (n) PAPS (o) PUPS



(p) CHOL

Figure 2.2 *The chemical structures [20] of all lipid species studied in this thesis. The double bonds within the lipid tails and the charged groups of negatively charged lipids are highlighted with red and blue circles, respectively.*

2.3 G protein-coupled receptors

G protein-coupled receptors are the largest family of membrane proteins [10, 11]. They mediate most of our physiological responses to neurotransmitters, hormones, and light among other environmental stimulants, and also transfer the signals across the plasma membrane. The malfunction of GPCRs can cause a wide spectrum of diseases, for example heart failure, hypertension, and nephrogenic diabetes insipidus [13]. Therefore, GPCRs are also the largest class of targets for modern drugs [21]. The knowledge of GPCR function and structure is important for developing new drugs to treat the different diseases.

All GPCRs share a common core structure and have conserved amino acids at key positions along the protein sequence. They are composed of 7 membrane spanning α-helices that are separated by three intracellular and extracellular loops [10]. Additionally, there are the carboxyl and amino termini in the common GPCR structure. The most variable structures of GPCRs are the carboxyl terminus, the intracellular loop between helices 5 and 6, and the amino terminus [11].

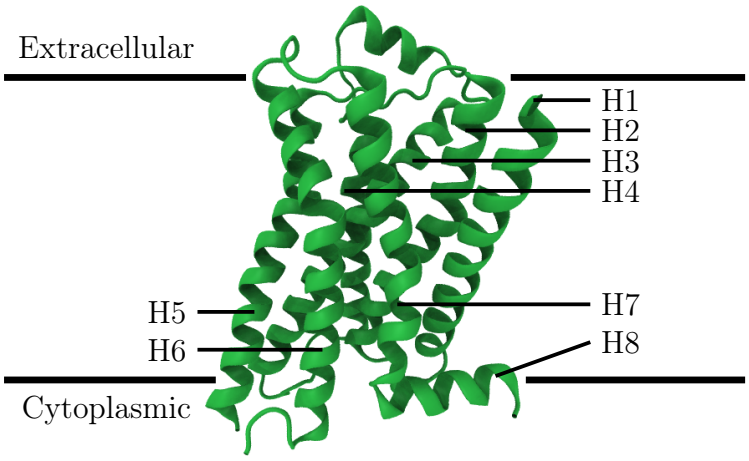
GPCRs' functional versatility is due to their structural plasticity [12]. They can exist in a combination of different conformations that can be influenced by at least different ligands, signaling and regulatory proteins, membrane properties, lipids, pH, and ions. For example, cholesterol is very important for the function of GPCRs [22].

GPCRs function as signal transducers [14, 23, p. 422] between the internal and external environments of cells. When an agonist binds to a GPCR in the extracellular side, a Guanine nucleotide-binding protein (G protein) becomes activated in the intracellular side of the membrane. The G protein then regulates an enzyme in the intracellular department, which generates an intracellular second messenger that conveys the signal further.

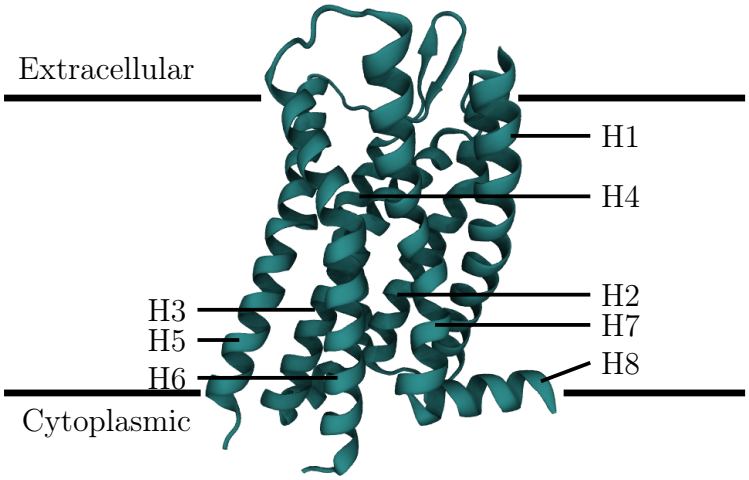
This study concentrates on three proteins that belong to the GPCR family: β_2AR, μ-OR, and Rho that are illustrated in Figure 2.3. These proteins are introduced in the following Sections.

2.3.1 β_2-adrenergic receptor

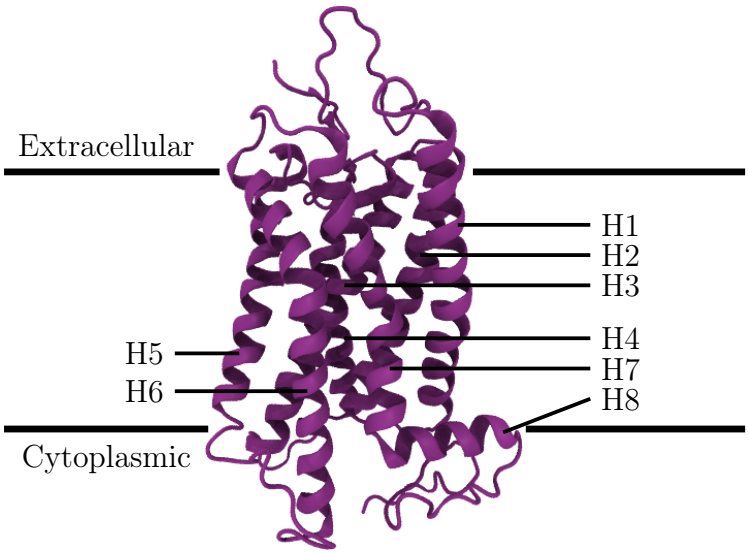
β_2-adrenergic receptor is a plasma membrane-bound protein that belongs to the family of G protein-coupled receptors [23, 25]. It is responsible for maintaining



(a) β_2AR



(b) μ-OR



(c) Rho

Figure 2.3 *The structures of β_2AR, μ-OR and Rho proteins where the helices 1-8 (H1-H8) are illustrated. The structures of proteins are taken from the Protein Data Bank [24].*

homeostasis in the human body. β_2AR is expressed in multiple cell types of the human body but especially in the airway smooth muscle cells [26]. The malfunction of β_2AR plays a central role in a number of diseases including anxiety, migraine, respiratory diseases, glaucoma, and cardiovascular diseases like hypertension [27, 28].

Like all GPCRs, β_2AR consists of seven transmembrane-spanning α-helices, an extracellular amino terminus, an intracellular carboxyl terminus, three extracellular loops, and three intracellular loops [26] (Figure 2.3 a). There is also one additional α-helix in the intracellular part that is sometimes denoted the eighth α-helix. The helices form a three-dimensional binding site for the receptor agonist in the extracellular part of the protein.

β_2AR can oscillate between at least two forms: activated and inactivated [26]. The binding of epinephrine, norepinephrine or other β_2AR agonists induce the activation of β_2AR. In the activated form, β_2AR is most often associated with the α-subunit of G_s-type G protein with a molecule of *guanosine triphosphate* (GTP). This mediates the increase of the level of an intracellular second messenger, such as *cyclic adenosine monophosphate* (cAMP). The replacement of GTP by *guanosine diphosphate* (GDP) reduces the affinity of the α-subunit for the receptor. This causes dissociation of the receptor from the α-subunit and induces the receptor to return to its inactive form. Different membrane lipids play a role in regulating the function of G protein-coupled receptors. MD simulations have already revealed functional binding sites of cholesterol in β_2AR [29].

2.3.2 μ-opioid receptor

Opium is one of the oldest drugs and its derivatives, morphine and codeine, are powerful clinical drugs to relieve severe pain [30]. These opioids produce analgesia by binding to the μ-opioid receptor in the central nervous system but they also cause different side effects. The μ-OR is expressed in multiple areas of the human brain and spinal cord [31]. The knowledge of opioid receptor structure and function enables the development of better drugs for the management of pain and addiction.

The μ-OR belongs to the GPCR family and consists of seven transmembrane (TM) α-helices [30], the eighth intracellular α-helix, three intracellular and extracellular loops, and of amino and carboxyl termini (Figure 2.3 b). The agonist binding site is in the extracellular side of the receptor in the middle of the TM helices.

The regulation and signaling of the μ-OR are strongly dependent on agonists [32]. Endogenous opioid peptides and exogenous alkaloid opiates can stimulate μ-OR, which leads to suppression of neuronal activities that causes the analgesic function [30, 31]. At the molecular level, μ-OR couples to the G_i or G_0 -type G proteins, which mediates the inhibition of adenylyl cyclase activity, closure of voltage-gated calcium ion channels, and the opening of inwardly rectifying K^+ channels. The μ-OR activation results in the release of GDP from the α-subunits of G_i and G_0 proteins, which allows GTP to bind and activate the α-subunits. The activated subunits dissociate from the receptor and act on the downstream effector systems. The hydrolysis of GTP to GDP inactivates the subunit, which terminates the signal.

2.3.3 Rhodopsin

Rhodopsin is a photoreceptor protein in retinal rod cells [33]. Retinal rod cells are differentiated neurons that are also known as photoreceptor cells [34]. These cells are responsible for detecting light signals in the form of photons in the eye. The rod outer segment, which is a specialized part of a rod cell, contains rhodopsin and auxiliary proteins that convert and amplify the light signals. The outer rod segment is composed of stacks of distinct disks, where the main component of the bilayer disk membranes is rhodopsin. Proper expression of rhodopsin is essential for the formation of the outer rod segment and therefore also important for vision. Several diseases like *Retinis pigmentosa* and congenital night blindness are linked to the malfunction of rhodopsin [35].

The cylindrical shape of rhodopsin is due to the arrangement of the seven transmembrane and one peripheral helices connected with three intracellular and extracellular loops [34] (Figure 2.3 c). Rhodopsin includes a cofactor, 11-*cis*-retinal, which is covalently linked to the helix 7 in the extracellular part of the protein [36]. The molecule 11-*cis*-retinal is essential for the detection of photons.

The absorption of the photon activates the protein [34, 37]. The absorption of the photon triggers the conversion of the 11-*cis*-retinal into its *trans* isomer. This initiates a visual cascade that leads to the closure of ligand-gated calcium channels. The closure of the channels leads to the excitation of the visual nerve that conveys an electrical impulse to the visual cortex of the brain.

2.4 Molecular interactions

The interactions between the molecules within plasma membranes are important for many functions of biological cells [38]. The lipid-lipid and protein-lipid interactions are discussed in this Section.

2.4.1 Lipid-lipid interactions

Lipid species have distinct properties: they have different charge distributions, shapes, and sizes [1], among other things. These properties of lipids affect the interactions between molecular species within plasma membranes, and the interactions between lipids cause multiple different phenomena in plasma membranes.

Firstly, lipid-lipid interactions drive the lipids to assemble spontaneously into bilayers and other well-defined structures in aqueous environments [39]. This self-assembly is governed by three key phenomena: thermodynamics, interaction forces, and molecular geometry. The assembly of lipids derives mostly from the hydrophobic interactions between the hydrocarbon tails and the hydrophilic nature of the headgroups. Therefore, the headgroups are drawn to be in contact with water and, oppositely, lipid tails repel water. This leads to the assembly of the lipids into thermodynamically favored structures where the lipid tails are packed together and the polar headgroups are facing the water phase. The geometry of lipids affects the structure that is formed. If the lipids are approximately cylindrical, then a bilayer is usually the favored structure.

Secondly, lipid-lipid interactions affect the lateral and transmembrane distributions of lipids in the plasma membranes [40]. Depending on the structures, sizes, charges, and other properties of lipids, some lipid-lipid interactions are more favorable than others, which affects the mixing of lipids in the plasma membranes. The heterogeneous distribution of lipids affects the physical and chemical properties of plasma membranes, which has important effects on cellular phenomena such as the activity of plasma membrane proteins. The lateral and transmembrane distributions of lipids in the plasma membranes are discussed in Sections 2.5 and 2.6.

2.4.2 Protein-lipid interactions

The lipids within plasma membranes can be distinguished into three groups according to their interactions with membrane proteins [4, 41]. Annular lipids are the lipids that are in contact with the membrane proteins but are not bound to them in a specific manner. Bulk lipids form the bulk of the membrane without a direct contact with the membrane proteins. Non-annular lipids interact specifically with a membrane protein and are usually buried inside the protein structure. Many mechanisms rule the protein-lipid interactions [41], such as hydrophobic thickness of the molecules, lateral pressure field of the membrane, the charge distribution, and the amino acid side chains of the proteins.

Lipid-protein interactions are essential for the cells. The interactions between proteins and lipids facilitate processes such as respiration, protein and solute transport, signal transduction, and motility [38]. Numerous studies have shown that specific lipids affect the structural stability of proteins, control protein insertion and folding, influence the assembly of multisubunit complexes, and can directly affect the function of membrane proteins [4]. A portion of membrane proteins is known to have specific lipid binding sites [29, 42, 43]. At these sites, the lipid binding is stabilized by interactions between lipid headgroup and amino acids within protein and by a number of non-polar interactions between hydrophobic lipid tails and the protein. The increasing number of known membrane protein structures paves the way to analyze the role of specific lipid-binding sites and to reveal the detailed structure and function of the membrane proteins.

2.5 Plasma membranes have an asymmetric transmembrane lipid distribution

Plasma membranes have an asymmetric transmembrane distribution of lipids [3, 44]. Lipid asymmetry is a consequence of many factors: The main site for lipid synthesis is the endoplasmic reticulum [3]. From there, the lipids are transported to the intracellular leaflet of the plasma membrane. There are many mechanisms that then distribute the lipids to both leaflets or limit their movement. These mechanisms include the biophysical properties of lipids that dictate the ability of a lipid to cross the bilayer spontaneously, mechanisms trapping lipids in one leaflet, and the activity of proteins called transporters that assist lipid transmembrane movement.

The spontaneous movement of lipids between membrane leaflets, called flip-flops, are slow processes in plasma membranes [3]. A lipid flip-flop can take place in a time scale ranging from hours to days for phosphatidylcholine lipids, and in seconds for sterols [3]. However, the transporter proteins can increase the rate of flip-flops.

Some lipid species are enriched in the extracellular and some in the intracellular leaflet [3] of the plasma membrane. Sphingomyelin (SM), phosphatidylcholine (PC), phosphatidylglycerol (PG), and glycosphingolipids (GSLs) are enriched in the extracellular leaflet whereas phosphatidylserine (PS), phosphatidylethanolamine (PE), and phosphatidylinositol (PI) are enriched in the intracellular side of the membrane [3, 8, 45]. Cholesterol is found in both leaflets but the enrichment of cholesterol in intra- or extracellular leaflets is a controversial issue [3] at the moment.

The transmembrane lipid asymmetry has important functional consequences. For example, the exposure of PS lipids on the cell surface acts as a signal for phagocytosis and blood coagulation. Membrane asymmetry also affects several bilayer properties such as membrane potential, permeability, shape, and surface charge [8], which can affect the activity of plasma membrane proteins.

2.6 Plasma membranes are heterogeneous in their lateral organization

In 1972, Singer and Nicolson [46] introduced the classical fluid mosaic model regarding the lateral distribution of lipids in biological membranes. The fluid mosaic model predicts that membranes are constituted of lipids in a fluid state and that proteins are freely floating in this lipid "sea". They also supposed a random lateral distribution of molecular components in the plasma membrane due to the free lateral and rotational movement of the molecules in a fluid state.

After the introduction of the fluid mosaic model, the understanding of the structure of cellular membranes has developed substantially [47]. Today, there is a large number of experimental evidence suggesting that stable lateral domains exist within plasma membranes. These functional nanosized domains are called lipid rafts illustrated schematically in Figure 2.4. Lipid rafts are parts of biological membranes and they are considered to be enriched in sphingomyelin (SM), cholesterol, and saturated lipids [5]. Lipid rafts are also considered as more ordered and tightly packed [48] than other parts of biological membranes. They have been suggested to

influence membrane fluidity, membrane protein trafficking, protein activity, and to regulate cellular processes such as neurotransmission [6, 7]. Membrane proteins can be assigned to different categories according to their position with relation to lipid rafts.

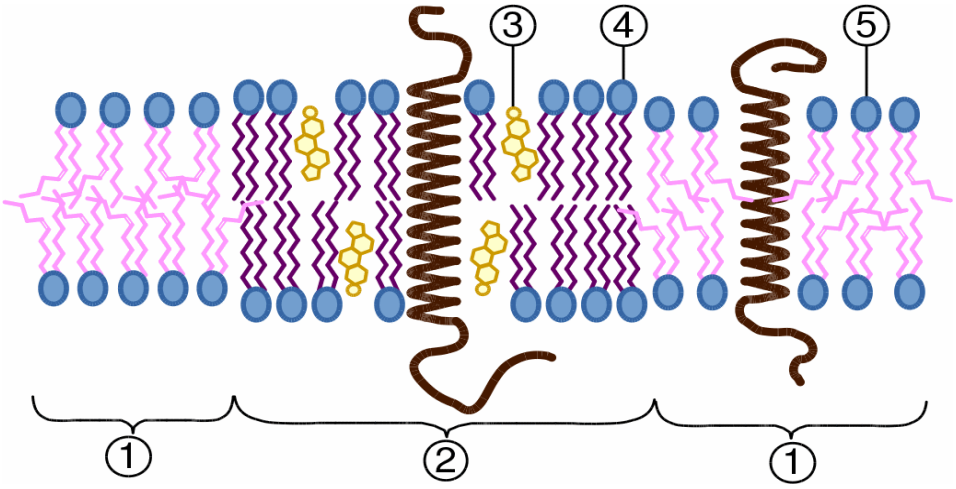


Figure 2.4 *A schematic figure of a lipid raft: 1. non-raft membrane 2. lipid raft 3. cholesterol 4. saturated lipid 5. unsaturated lipid.*

The movement of molecules is not entirely free in the membranes. The lateral lipid diffusion is a slow process in MD time scales [3]. Furthermore, the increase of protein concentration in a plasma membrane results in a decrease of lipid diffusion [17]. The lipid diffusion coefficient in biological membranes is approximately 1 $\frac{\mu m^2}{s}$ [49], which means that a lipid diffuses on average 2 µm in 1 s. In comparison, the water self-diffusion coefficient is circa 3 000 $\frac{\mu m^2}{s}$ [50] at body temperature.

Lipid rafts are functional [9, 51, 52] since they have been suggested to take part in multiple cellular processes. Among other things, these cellular processes include membrane trafficking, signal transduction, and the regulation of membrane protein activity. Lipid rafts may regulate protein activity through direct lipid-protein interactions or by changing the physicochemical properties of bilayers. Lipid domains have distinct physical properties, and lipid composition can affect protein conformation within plasma membranes and thus alter the activity of the protein. For example, Niemelä *et al.* [47] proposed that certain classes of membrane proteins are regulated by the change in the membrane pressure profile, which is altered by a change in the lipid content. Lipid rafts could also regulate protein activity by facilitating selective protein-protein interactions by concentrating proteins into a single

type of domain or by segregating proteins into separate domains [53]. It is already known that the presence of certain lipids affect the activity of some membrane proteins. For example, it has been shown that binding of the human epidermal growth factor to a specific raft lipid changes its conformation [54].

The analysis of biological membrane structures provides a framework to understand lipid-protein interactions [4]. Integral proteins are surrounded by a shell of lipid molecules. Usually, an individual lipid molecule remains within this shell only for a short period of time. However, some lipid molecules are found to interact with some proteins for substantially long times. Lipids having specific, long-lasting, and recurring contacts with proteins may have vital functional roles. Therefore, it is important to study the spatial distribution of lipids near proteins to find possible functional interactions. The obtained information is needed to understand the normal function of cells and to cure abnormalities in the case of a disease.

Despite the proposed importance of lipid domains, their exact roles in biological processes are not well understood [47, 53]. The forces driving domain formation are also still enigmatic [7]. This is because lipid domains are difficult to study due to their small size, which is not to be resolved by conventional experimental methods, such as light or fluorescence microscopy [5, 9]. Therefore, it is extremely difficult to perform experiments studying lipid composition, size, and lifetime of lipid domains using living cells [53, 55]. Molecular dynamics simulations can be used to obtain atomistic or near-atomistic information about systems modeling biological membranes. MD simulations have already been successfully applied to study specific interactions between different lipid and molecular species [2, 56, 57].

We research whether the studied proteins can induce domain formation in plasma membrane models and whether there are favored interactions between lipids and proteins. We focus on how GPCRs affect the lateral and transmembrane distributions of the most abundant membrane lipid species. The resolution of MD simulations is indispensable for elucidating the characteristics of lipid domains. Therefore, this thesis is based on data found using molecular dynamics simulations.

3 MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics simulations are tools for studying molecular structures and functions through computational methods [58, p. 3]. It indicates processes of describing complex systems with realistic models. The model consists of a representation of a biological system and of rules that describe its behavior [59, p. 3]. The objective of molecular dynamics simulations is to predict and understand macroscopic properties based on the knowledge on a microscopic scale.

When the size of the studied system approaches the nanoscale, experimental research gets very complicated. The studying of biological systems at the nanoscale is almost impossible by using existing experimental methods alone [1]. Only an electron microscope can reveal details up to a few nanometers but the preparation of the sample is extremely convoluted. The sample preparation can also affect the results of the experimental studies. What is more, the preparation of the sample usually damages the sample so that the studying of living cells with electron microscopy is impossible.

Scientific supercomputing has provided a way to study cells in unprecedented detail [60]. With MD simulations, even atomic-level information of the structures of cells and different phenomena occurring in them can be obtained. Molecular dynamics simulations do not have the drawbacks of sample preparation like staining, cutting to sections, or fixing the sample to preserve it. MD simulations can also be used to study systems in the extremes of environments, such as in high temperatures or pressures, that are impossible to achieve in a laboratory [61, p. 1]. However, also computational methods have some limitations discussed in Section 3.8, but nevertheless molecular dynamics methods serve as a powerful tool to resolve the unsolved questions of biology.

In this thesis, the lipid environment of GPCR proteins is studied using MD simulations. There are different program packages to perform the MD simulations, but here we use the GROMACS version 5.0.4 with the MARTINI force field. The consecutive

steps of performing an MD simulation and the limitations of it are discussed in this Chapter, especially from the perspective of GROMACS. The Chapter is based on GROMACS User Manual version 5.0.4 [62], if not mentioned otherwise.

3.1 System structure

In order to begin an MD simulation one has to prepare the system first. (a) Adequate molecules are chosen to model the biological system under investigation. (b) The number of each molecule and the natural solution are chosen. The model should be prepared so that it mimics the properties of the real physical system. Only relevant features of the system are usually taken into consideration to keep the system simple enough to examine it computationally. The structure is usually built according to experimental findings and verified by theoretical models.

The molecules are placed in a simulation box. The simulation box is determined by three vectors and angles which determine its shape and size. The molecules are placed randomly in the box so that every particle gets a position determined by three coordinates in a three-dimensional space. Then, velocities are assigned to each particle. The velocities of particles can be set by the user or they can be generated by a computer program according to the Maxwell–Boltzmann distribution function at a given temperature, T, according to equation

$$p(v_i) = \sqrt{\frac{m_i}{2\pi k_B T}} exp(-\frac{m_i v_i^2}{2k_B T}), \tag{3.1}$$

where $p(v_i)$ is the probability function of the velocity v_i for particle i, m_i is the mass of that particle, and k_B is the Boltzmann constant.

After the molecules have been given the positions and velocities, molecular topology has to be provided. The topology includes the physical properties of all atoms that are simulated as well as the description of the interactions between the atoms. For example, the masses and charges of atoms as well as the bonds and angles between atoms are included in a topology file. This description is often static in an MD simulation, which means that chemical reactions cannot occur during the simulation. Therefore, no bonds can break or form during an MD simulation if the molecular description is static.

3.2 Force fields

A force field describes the physical laws and parameters that govern the system. It is a set of potential functions that defines the interactions between the particles in the system. The interactions between particles can be either repulsive or attractive. Force fields include the mathematical formulas that are used to calculate the forces acting on atoms due to their positions and physical properties during the simulation. The forces are calculated based on the potential energy functions.

To solve the potential function, two kinds of interactions are taken into account and discussed in more detail: bonded and non-bonded interactions. The total potential energy function is the sum of these interactions,

$$V_{total} = V_{bonded} + V_{non-bonded}, \tag{3.2}$$

where V_{bonded} is the potential energy function of bonded interactions and $V_{non-bonded}$ is the potential energy function of non-bonded interactions.

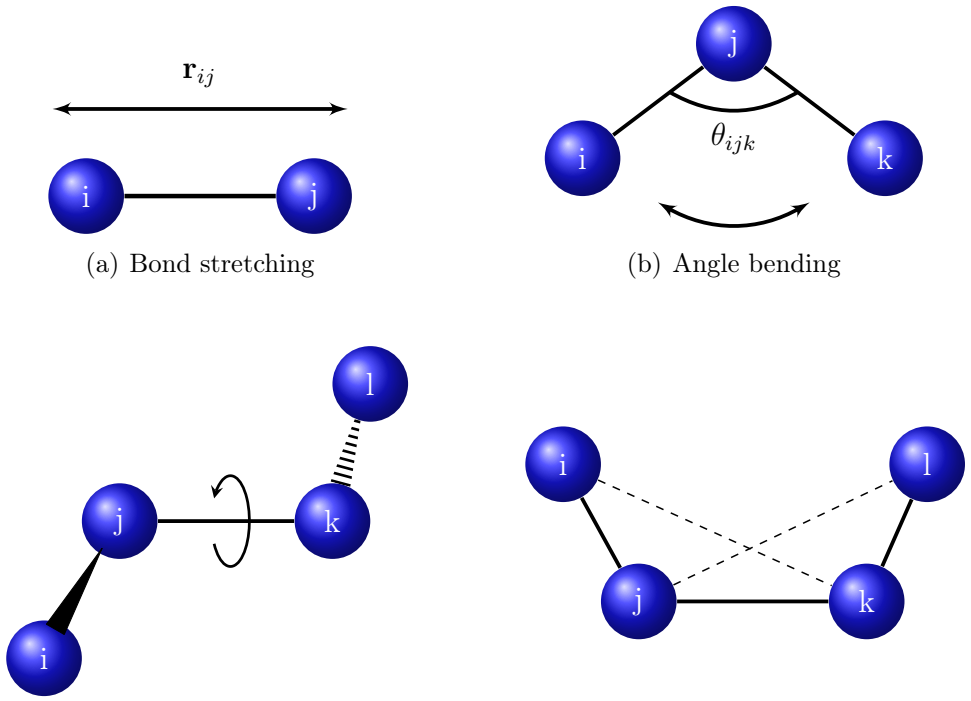
There are many different force fields based on different principles that are specialized for distinct applications [59] and a few examples of them are introduced here. There are classical all-atom force fields, such as OPLS [63] for organic molecules and peptides where the molecular description is atomistic. There are also reactive force fields that can be used to model chemical reactions. Thus, bonds can break and form during a simulation when a reactive force field is employed. One example of a reactive force field is the ReaxFF [64]. Polarizable force fields [65] include a representation of the change in the molecule charge distribution in response to its environment.

There are also coarse-grained (CG) force fields containing a set of parameters that can be used for construction of systems like biological membranes. Coarse-graining is a way to reduce the degrees of freedom of a system. Typically, a few heavy atoms are represented as one bead instead of representing all of the atoms in the system. Then, a coarse-grained molecular description, like MARTINI [66], is used to describe the interactions between these beads. Thus, larger and/or longer simulations can be performed for coarse-grained systems compared with all-atom systems. In this study, the MARTINI force field was applied. This force field was chosen to lower the need for computational resources.

3.2.1 Bonded interactions

Bonded interactions govern the interactions between particles within molecules. These particles are attached to each other via covalent bonds. All covalent bonds are defined in the topology file and cannot change during the simulation if the molecular description is static, such as in the MARTINI force field. Bonded interactions usually include interactions between two, three or four atoms. These interactions describe the bond stretching, angle vibration, dihedral angle torsion, and out of plane bending, see Figure 3.1. The total bonded interaction potential acting on specific atoms is the sum of these different potentials:

$$V_{bonded} = V_{bonds} + V_{angles} + V_{dihedrals}. \tag{3.3}$$



(a) Bond stretching

(b) Angle bending

(c) Proper dihedral angle torsion

(d) Improper dihedral out of plane bending

Figure 3.1 *Bonds, angles, and dihedrals are defined in force fields.*

The stretching of bonds between two covalently bonded atoms i and j is given by harmonic potential

$$V_{bonds} = \frac{1}{2} k_b (r_{ij} - r_0)^2, \tag{3.4}$$

where k_b is the force constant and $(r_{ij} - r_0)$ is the bond length compared to the reference bond length. Harmonic angle potential is defined between a triplet of atoms i, j, and k and can also be represented by a harmonic potential:

$$V_{angles} = \frac{1}{2} k_\theta (\theta_{ijk} - \theta_0)^2, \tag{3.5}$$

where k_θ is the angular force constant and θ_{ijk} is the angle between particles i, j, and k. θ_0 is the corresponding reference angle.

There are proper and improper dihedrals which describe the angles between four successively bonded atoms. The total dihedral potential is the sum of these two potentials:

$$V_{dihedrals} = V_{proper} + V_{improper}. \tag{3.6}$$

Proper dihedrals prevent bond rotation around the jk-axis shown in Figure 3.1 (c). Either the GROMOS periodic function or the Ryckaert-Bellemans potential are typically used to describe potentials for proper dihedrals. The Ryckaert-Bellemans (RB) potential equals

$$V_{RB} = \sum_{n=0}^{5} C_n (cos\phi)^n, \tag{3.7}$$

where C_n is a constant defined in the force field and ϕ is the torsion angle. The form of the periodic potential function is

$$V_{periodic} = k_\phi (1 + cos(n\phi - \phi_0)), \tag{3.8}$$

where n is a constant, k_ϕ is a force constant, ϕ is the torsion angle, and ϕ_0 is the reference torsion angle.

Improper dihedrals are used to ensure that planar structures stay planar and to prevent molecules from changing their chirality meaning transforming into their mirror images. Harmonic potential for the simplest improper dihedral potential is given by

$$V_{improper\ dihedral} = \frac{1}{2}k_\xi(\xi_{ijkl} - \xi_0)^2, \tag{3.9}$$

where k_ξ is the force constant, ξ_{ijkl} is the improper dihedral angle, and ξ_0 is the reference value for the angle defined in the force field.

3.2.2 Non-bonded interactions

Non-bonded interactions are the interactions between atoms that are not covalently linked to each other. The two most common non-bonded interactions are electrostatic (usually described using Coulomb potential, V_C) and van der Waals (usually described by Lennard-Jones potential, V_{LJ}). The total potential energy function for non-bonded interaction is the sum of these two potentials:

$$V_{non-bonded} = V_{LJ} + V_C. \tag{3.10}$$

The V_{LJ} potential between atoms i and j at distance r from each other is

$$V_{LJ} = V_{repulsion} + V_{attraction} = \frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}}, \tag{3.11}$$

where $C_{ij}^{(12)}$ and $C_{ij}^{(6)}$ are constants depending on the types of atoms. The V_{LJ} potential is illustrated in Figure 3.2. The first term of the equation describes the repulsive interaction between the particles that is most significant in short ranges. The second term describes the attractive interaction, which is important in long ranges.

The Coulomb potential describes the interactions between two charged particles i and j. It is given by

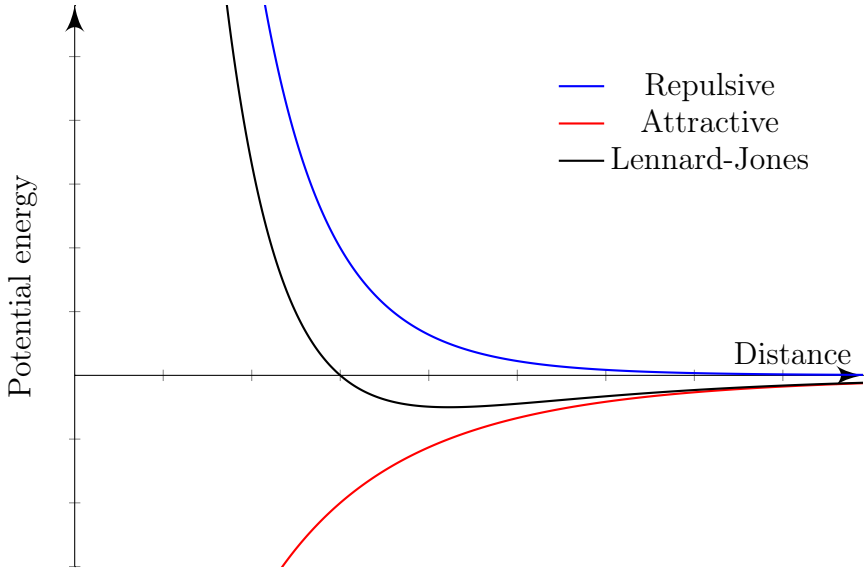


Figure 3.2 *The Lennard-Jones potential energy function is comprised of a repulsive term and an attractive term.*

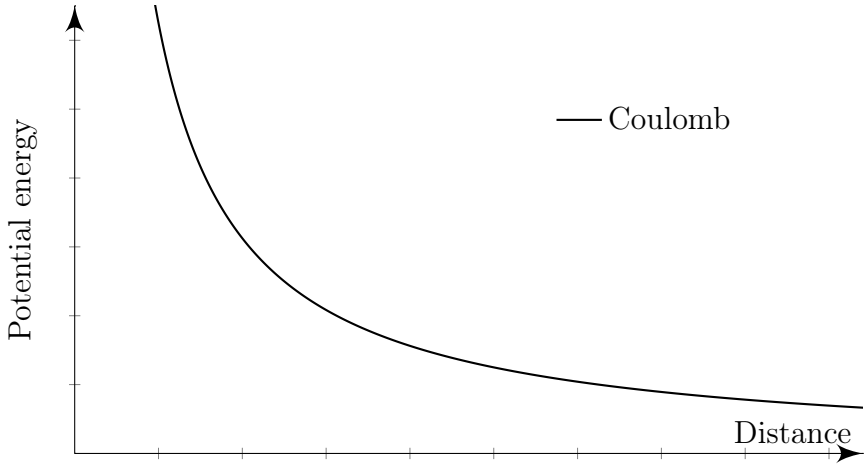


Figure 3.3 *The Coulomb interaction between two particles with equal charges.*

$$V_C = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{r}, \tag{3.12}$$

where q_i and q_j are the charges of the particles and ϵ_0 is the vacuum permittivity. A schematic figure of the Coulomb interaction is shown in Figure 3.3.

There are also other special potential functions, such as restraints, that can be ap-

plied to the molecules. Restraints can be implemented by applying an energy penalty if the molecules deviate from a defined value. For example, an elastic network model restraining the distances between backbone beads can be included to maintain the tertiary structure of a protein during an MD simulation [67].

3.3 Energy minimization

After the molecules are arranged into the simulation box, the configuration of the system may be very far from equilibrium. In these cases the forces acting on particles can be excessively large. For example, if the charged particles are very near to each other the Coulomb interaction acting on these atoms can be very large. Overly large forces can cause an MD simulation to fail, and to eschew this, a robust energy minimization step is performed before performing the actual MD simulation to minimize the initial forces acting on the molecules.

To energy minimize the system, the particles are moved so that the forces are not excessively large and therefore the free energy of the system becomes adequately low. The potential energy functions of molecular systems are very complex landscapes with multiple dimensions and a large number of local minima and maxima. The number of all minima and maxima is so high that it is impossible to survey comprehensively all of them. That is why there is no minimization method that can guarantee to determine the global minimum of the system in any practical amount of time. Fortunately, the nearest local minimum can be found relatively easily. Here, the nearest minimum does not necessarily mean the closest in a geometrical sense but one which can be found by moving down along the steepest local gradient of the energy function.

The steepest descent method is an example of an energy minimization method available in GROMACS. This algorithm uses derivative information of the potential energy function and takes a step in the direction of the force without consideration of previous steps. The algorithm continues to take steps until a user-specified number of steps has been performed or when the values of the forces are smaller than the given value.

There are also more delicate methods to do energy minimization compared with the steepest descent method, but the steepest descent method usually gives sufficiently good results in biological systems. The result of energy minimization does not have

to be perfect since the algorithm gives only the starting structure for the simulation and the structure is usually equilibrated during the simulation. The advantages of steepest descent are that it is a simple and quick method to find robustly the nearest local minimum.

3.4 Equations of motion generate the dynamics

In order to describe the dynamics of the system a classical or an *ab-initio* MD approach can be used. Classical mechanics means that quantum mechanics are not considered: the motion of electrons is ignored and the motion of atoms is described by their nuclei [68, p. 35]. When quantum mechanics need to be taken into account, there are also *ab-initio* methods available [69] to perform quantum mechanical simulations.

In this thesis, classical mechanics is used to describe the motion, i.e. the positions and velocities of atoms during MD simulations. First, the forces acting on a particle need to be calculated. The force, F_i, is defined in the force field as the negative derivative of the potential function $V(r_1, r_1, ..., r_N,)$:

$$\mathbf{F}_i = -\frac{\partial V_i}{\partial \mathbf{x}_i}, \tag{3.13}$$

where $\mathbf{x}_i$ is the position of the particle i. Newton's equation of motion are solved for a system of N interacting atoms to calculate the motion of the particles based on the forces calculated above:

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = \mathbf{F}_i, i = 1...N, \tag{3.14}$$

where m_i is the mass of the particle i and t is time.

Both equations are solved discretely during each time step of the simulation. The time step, Δt, is specified by the user. There are several different algorithms to integrate the equations of motion. The leap-frog algorithm is the default MD integrator in GROMACS. It uses two relations to calculate the positions and velocities:

$$\mathbf{x}_i(t + \Delta t) = \mathbf{x}_i(t) + \mathbf{v}_i(t + \frac{1}{2}\Delta t)\Delta t \tag{3.15}$$

$$\mathbf{v}_i(t + \frac{1}{2}\Delta t) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m_i}\mathbf{F}_i(t), \tag{3.16}$$

where $\mathbf{v}_i$ is the velocity of the particle i.

The resulting coordinates of all particles in the system are written regularly to an output file. The coordinates as a function of time constitute a trajectory of the system. Hence, the trajectory provides atomic or near atomic details of the structures and motions of the system. After some time of the simulation, the system usually reaches an equilibrium state where the average properties of the system do not change over time. Many macroscopic properties can be analyzed from the output file by averaging over an equilibrated trajectory.

As a summary, performing MD simulations includes several steps which are partly similar to real experiments [70, p. 63]: a sample is prepared, measurements are taken, and the gathered data is analyzed statistically to diminish the effects of noise. The workflow of an MD simulation is presented in Figure 3.4. Firstly, the structures and positions of molecules in the system are defined precisely and a force field is chosen to parametrize the particles. Before running the simulation, energy minimization is carried out to minimize the energy of the initial structure. Forces acting on atoms are calculated from the potential energy functions. Then, Newtonian mechanics defines the motion of each particle, which is calculated iteratively for every step of the simulation for a desired amount of time. This generates the output of the time development of positions and velocities of each particle during the simulation. The output file, a trajectory, is finally analyzed.

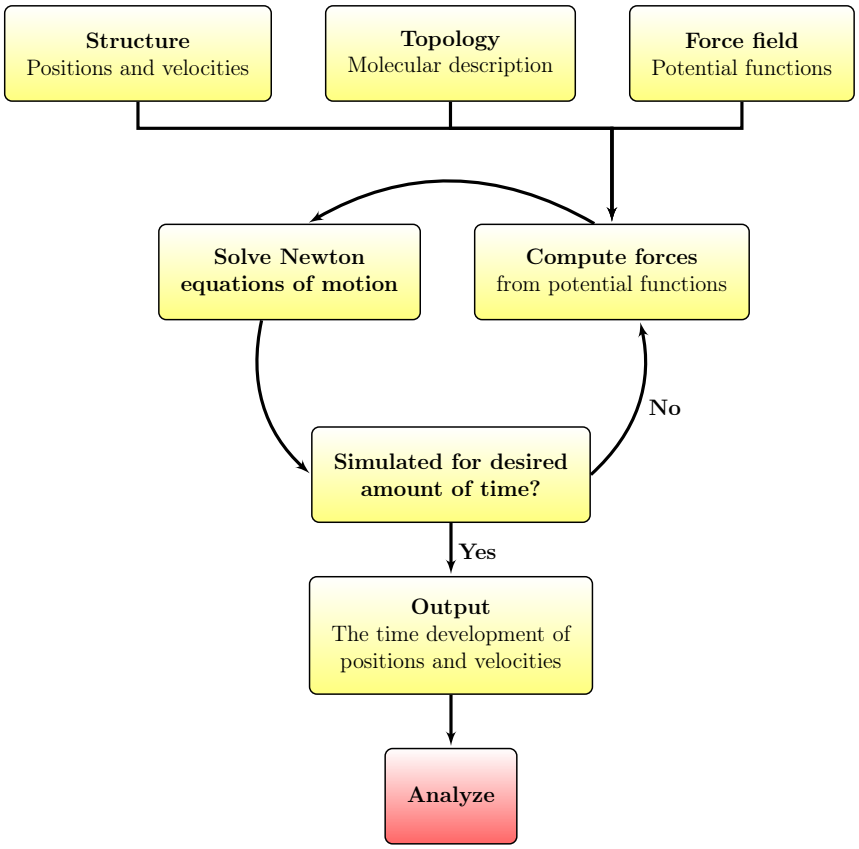


Figure 3.4 *A general algorithm for MD simulations.*

3.5 Thermostats and barostats

Usually biological experiments are conducted in standard conditions in order to allow comparisons of the results between different studies. Therefore, experiments are most often conducted at constant normal temperature and pressure (NpT) conditions at which we usually wish to conduct MD simulations as well. In order to maintain the constant temperature and pressure conditions, different temperature and pressure coupling methods are used. These algorithms are called thermostats for temperature and barostats for pressure coupling.

An often used temperature coupling algorithm is the Berendsen thermostat [71]. It couples the simulation system to an external heat bath of given temperature T_0. If the temperature of the system, T, deviates from the reference temperature, T_0, this algorithm corrects the temperature according to equation

$$\frac{dT}{dt} = \frac{T_0 - T}{\tau_t}, \tag{3.17}$$

where τ_t is a time constant. Therefore, the deviation of temperature decays exponentially and can be modified with the time constant.

There is also a pressure coupling method similar to Berendsen temperature coupling. The Berendsen pressure coupling [71] is an algorithm to keep the system at a constant pressure. In order to do that, it rescales the box size and the coordinates of particles to sustain a given reference pressure, P_0, according to

$$\frac{d\mathbf{P}}{dt} = \frac{\mathbf{P}_0 - \mathbf{P}}{\tau_p}. \tag{3.18}$$

The pressure coupling can be applied isotropically, which means that all dimensions are scaled together or anisotropically scaling the three dimensions separately.

3.6 Boundary conditions

In general, one wishes to study a system that is composed of so many particles that it is not possible to simulate the whole system [59, p. 140], such as a whole cell membrane. Consequently, the whole cell membrane is not modeled but only a part of it. MD systems are always finite and therefore a substantial portion of particles have boundaries that interact with the edges of the simulation box. This is undesirable since the properties of the system are very different from the ones of the box surroundings, which can cause artefacts in the results.

There are multiple ways to handle the boundary effects but the most commonly applied way is the use of *periodic boundary conditions* [59, p. 141] (PBCs). This means that the system is replicated in all dimensions so that it becomes an infinite lattice. This PBC treatment is illustrated in Figure 3.5 (in two-dimensions for clarity). PBC works in a way that when a particle exits the box from one side it simultaneously re-enters the box from the other side. This way there are no artificial boundaries affecting the system.

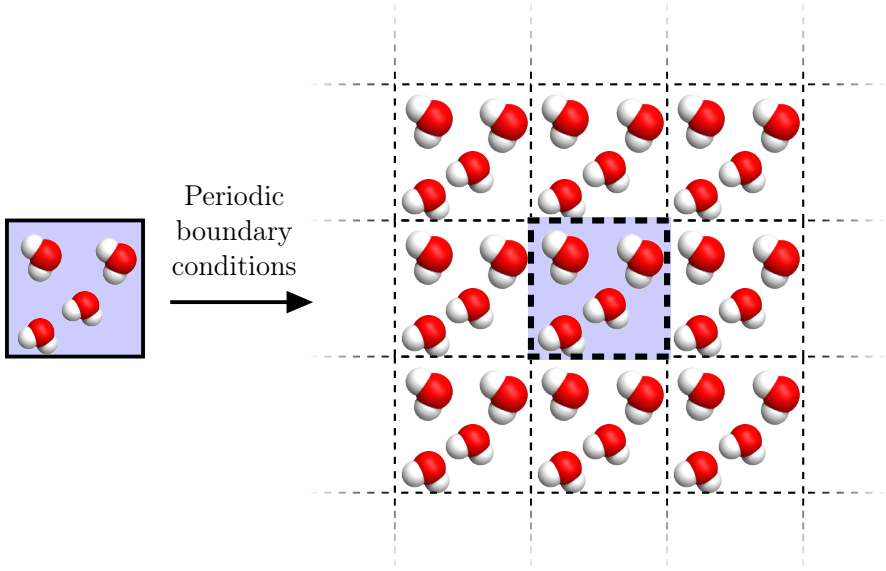


Figure 3.5 *Periodic boundary conditions make the box infinite, containing the unit cell (in the center) and its replicas.*

3.7 Long-range interactions

Long-range interactions comprise Coulomb and Lennard-Jones interactions as described in Section 3.2. These interactions approach zero as the distance between the particles approaches infinity. Hence, the interactions have an infinite range but the significance of these interactions decreases quickly as the distance increases above a certain value.

Calculating the infinite number of interactions between the atoms would take an immense amount of resources for little gain. Moreover, the particles can interact with themselves due to the periodic boundary conditions which causes artefacts. Therefore, cut-off radii are used for van der Waals and electrostatic interactions, after which these interactions are set to zero. The cut-off radius cannot exceed half of the box size so that the particles cannot interact with themselves.

3.8 Limitations

Even though MD simulations are a powerful tool to study biological phenomena, they also have limitations. It is important to be aware of those limitations and take them into account when performing MD simulations.

Firstly, the simulations are usually classical since Newtonian mechanics is used to describe the motions of atoms. At normal temperatures, this approach is accurate enough but there are a few exceptions: Some processes involving protons and/or helium liquid at low temperatures cannot be reliably modeled with classical mechanics. Additionally, bonds that vibrate at high frequency, like practically all bonds vibrating at room temperature, may misbehave when modeled with classical mechanics.

The second limitation is related to the electrons of the atoms that are in the ground state in classical MD simulations. This means that only the motion of atoms is modeled but not that of electrons. This is accurate enough in most applications but naturally electron transfer processes and excited states cannot then be studied.

The force field defines the forces in MD simulations, and there are many force field options to choose from. The simulation results depend on the force field and therefore its careful selection is important. Force fields are only approximate since the potential energy functions are assumed as pair additive, which means that non-bonded forces result from the sum of non-bonded pair interactions. Force fields can be modified by the user if a need arises. However, in most cases they are sufficiently reliable when modeling biological environments.

GROMACS uses a cut-off radius for the long-range interactions: for van der Waals interactions and sometimes also for electrostatic interactions. This can cause accumulation of charges at the cut-off boundary or wrong energies in systems that contain charged particles. Long-range electrostatic algorithms, such as particle-mesh Ewald, should be used in this situation.

The MD system sizes are small compared with natural environments. Therefore, a substantial share of particles have a boundary with the environment which is unwanted. To resolve this, the periodic boundary conditions are usually used. The errors caused by the PBC are small for large systems, however for small systems the PBC may enhance internal correlation. It is important to be aware of whether the system size influences the results.

There are also limitations in the amount of computational resources that can be used to perform MD simulations. This means that the sizes of the systems and the lengths of the simulations are limited. If large system sizes are essential, coarse-grained models have to be used. This substantially lowers the need for resources.

4 MODELS AND ANALYSIS METHODS

In this thesis, the effect of GPCR proteins on the lateral and transmembrane distributions of lipids is studied by using molecular dynamics simulations introduced in the previous Chapter. The structures of the studied systems are first described in this Chapter. Then, the details of the simulations and simulation parameters are discussed. Analysis methods used in this thesis are explicated in Section 4.3. The results obtained with the analysis methods are presented in the following Chapter 5.

4.1 The structures of the studied systems

To study the self-assembly of different lipids around GPCR proteins, we built three systems with a different GPCR protein in each of them. The studied proteins are β_2AR, μ-OR, and Rho presented in Figure 2.3. The structure of β_2AR is modified [72] from the starting configuration 3D4S in the Protein Data Bank (PDB) [24] proposed by Hanson *et al.* [73]. The μ-opioid receptor (4DKL in PDB [24]) and Rho (1U19 in PDB [24]) protein structures are based on the structures proposed by Manglik *et al.* [30] and Okada *et al.* [37], respectively. All of these proteins belong to the GPCR family and have a consistent structure of seven hydrophobic transmembrane segments [11] described in Chapter 2.

Each system is composed of one of the three proteins, 517 lipid molecules of 16 different lipid species, water, and sodium and chloride ions, see Table 4.1. The only differences between the systems, in addition to the different protein in each system, are the amounts of sodium ions and water beads. The numbers of sodium ions and water beads are slightly different in each system due to the different total charges of the proteins that are neutralized by replacing an adequate amount of water beads with sodium ions. The lipid composition mimics the natural lipid composition of biological plasma membranes and is based on the plasma membrane model composition proposed by Ingolfsson *et al.* [74].

Table 4.1 *The numbers of molecules in the studied systems.*

Component	β_2AR	Rho	μ-OR
Protein	1	1	1
POPC	110	110	110
DPSM	21	21	21
DXSM	61	61	61
POPE	35	35	35
POPI	13	13	13
POPS	7	7	7
POPA	4	4	4
PODG	4	4	4
DPCE	4	4	4
DOPC	11	11	11
DOPE	17	17	17
PAPC	32	32	32
PAPE	35	35	35
PAPS	15	15	15
PUPS	7	7	7
CHOL	141	141	141
Water	6 038	6 027	6 044
Na^+	113	124	107
Cl^-	70	70	70

We used a coarse-grained representation of the molecules in order to lower the need for computational resources. We selected the MARTINI force field to parametrize the molecules. The MARTINI lipid topology and coordinate files can be found on the website: Martini Coarse Grain Force Field for Biomolecular Simulations [75]. The corresponding atomistic structures of the studied lipid molecules are shown in Figure 2.2 in Chapter 2.

The objective of this thesis is to study how GPCR proteins affect the lateral and transmembrane distribution of lipids. Therefore, we want to study well equilibrated systems where the initial configuration of the system does not affect the results. Often, MD simulation systems are built in a way that the lipids and proteins are positioned into a bilayer array in the xy-plane [76]. However, lateral lipid diffusion and lipid flip-flops are slow phenomena in standard MD timescales [3]. It would take substantial computational resources to run simulations long enough to obtain well equilibrated systems if the lipids would be arranged into a bilayer from the beginning. To bypass this challenge, the protein is centered into a simulation box and all lipids are placed randomly in the simulation box. The lipids may then

freely form a bilayer around the protein, thus mimicking plasma membranes during the MD simulation. This self-assembly of proteins and lipids is a relatively new way of creating unbiased MD models of biological membranes. To the best of our knowledge, self-assembly is previously utilized in a few studies [77, 78, 79, 80, 81] but not for as complex systems as our model, including both a protein and as many as 16 different lipid species.

We built a total of 50 copies of each system, including one of the three proteins, resulting in a total of 150 simulated systems. The initial positions of the lipids were random in each copy and therefore the starting configuration of each simulation was different. The systems were solvated with 6 221 water beads corresponding to 24 884 water molecules since one water bead represents four water molecules in the MARTINI force field [82]. An adequate number of water beads were changed to Na^+ and Cl^- ions equaling to a physiological salt concentration of $0.15\,mol/dm^3$. Finally, the total negative charge of the systems was neutralized by substituting a sufficient number of water beads with Na^+ ions.

4.2 Simulation parameters

We performed a total of 150 molecular dynamics simulations: 50 repetitions of each system having a random initial configuration with the coarse-grained molecular description and the MARTINI force field. The simulations were conducted using the GROMACS package version 5.0.4 [83].

Before MD simulations, the initial structures were energy minimized with the steepest descent algorithm to find the nearest local minimum of the potential energy function. Then, we began the molecular dynamics simulation: The time step for numerical integration was set to 20 fs. The simulations were carried out at constant 1 bar semi-isotropic pressure. The pressure was controlled with the Berendsen [71] barostat with a time constant of 3 ps. The temperature was also controlled with the Berendsen thermostat and held at constant 310 K temperature with a time constant of 1.5 ps.

A 1.1 nm cutoff was used for both electrostatic and van der Waals interactions. Periodic boundary conditions were applied in all three dimensions. The LINCS [84] algorithm was used to constrain bond lengths. All systems were first simulated at the very least for 400 ns. If a bilayer formed during the first 400 ns of simulation

but the protein was not within the bilayer, the simulation was not extended further or analyzed. All other simulations were extended to 1 μs.

4.3 Analysis methods

All analysis methods used in this study are described in this Section. Both existing GROMACS analysis tools and new analysis tools created for this project were used to analyze different properties of the systems. The new analysis tools were created in Matlab. Also all additional statistical calculations of the results were performed in Matlab.

4.3.1 Area per lipid

We calculated the area per lipid to determine the average area occupied by lipids. The area per lipid (APL) is often used for estimating the thermal equilibration state of a system [85, p. 418]. The area per lipid is given by

$$APL = \frac{A_{membrane}}{N_{lipids}/2}, \tag{4.1}$$

where $A_{membrane}$ is the total area of the bilayer and N_{total} is the total number of lipids in the system. The membrane area is calculated by multiplying the box vectors, obtained with an analysis tool *gmx energy*, in the plane of the bilayer. The area of the protein is not taken into account when calculating APL since the area per lipid is used here to follow the changes in the APL during the simulations.

4.3.2 Number of contacts

Number of contacts between the protein transmembrane surface and each lipid species was studied. A contact is defined so that if one or more beads of a lipid molecule is within a given distance of the protein transmembrane surface, the lipid is considered to be in contact with the protein. Here, the default value of 0.6 nm was used as the upper limit for the contact distance. The number of contacts was calculated for each simulation from the moment when the bilayer was assembled to the end of each simulation.

A GROMACS program *gmx mindist* was used to examine which lipid molecules are in contact with the protein. Next, a script was written in order to calculate the total number of contacts that each lipid species has with the protein. The number of contacts of each lipid species was divided by the number of all lipid contacts with the protein. Finally, the result was averaged over the analysis time and the mean values obtained were averaged over all repetition trajectories of each system. Therefore, the result gives the mean percentage of contacts that each lipid type occupies of all contacts. The error of the result is the standard error of the mean where the sample is the means of the results for each separate simulation repetitions. Thus, the error describes the variation between the results for different simulation repetitions of a specific system.

4.3.3 Interaction times

We calculated the average interaction times of each lipid species with the protein TM part. The average interaction times were calculated from the minimum distances between each lipid and the protein transmembrane surface. The minimum distances were calculated with a GROMACS program *gmx mindist*. The definition of an interaction time is illustrated in Figure 4.1.

An interaction begins when the minimum distance between a certain lipid and protein TM surface is less than 0.6 nm and ends when the distance is more than 0.6 nm during at least two consecutive time steps. If the minimum distance is over 0.6 nm for only one time step the interaction does not end. This allows short conformational shifts to happen without termination of an interaction. If the first interaction ends, the lipid can have a second interaction with the protein if the minimum distance between it and the protein TM surface is again less than 0.6 nm. This way all interaction times are calculated for every lipid molecules during all simulations.

The interaction times are calculated for every simulation from the bilayer assembly to the end of each simulation. The mean of the interaction times is calculated for each lipid species within each simulation repetition. Next, the result is averaged over all repetition trajectories of a system. Therefore, the result is the mean interaction time of each lipid species within each system. The error of the result is the standard error of the mean where the sample is the means of interaction times in each separate simulation repetitions and the error describes the variation between the results for different simulation repetitions of a specific system.

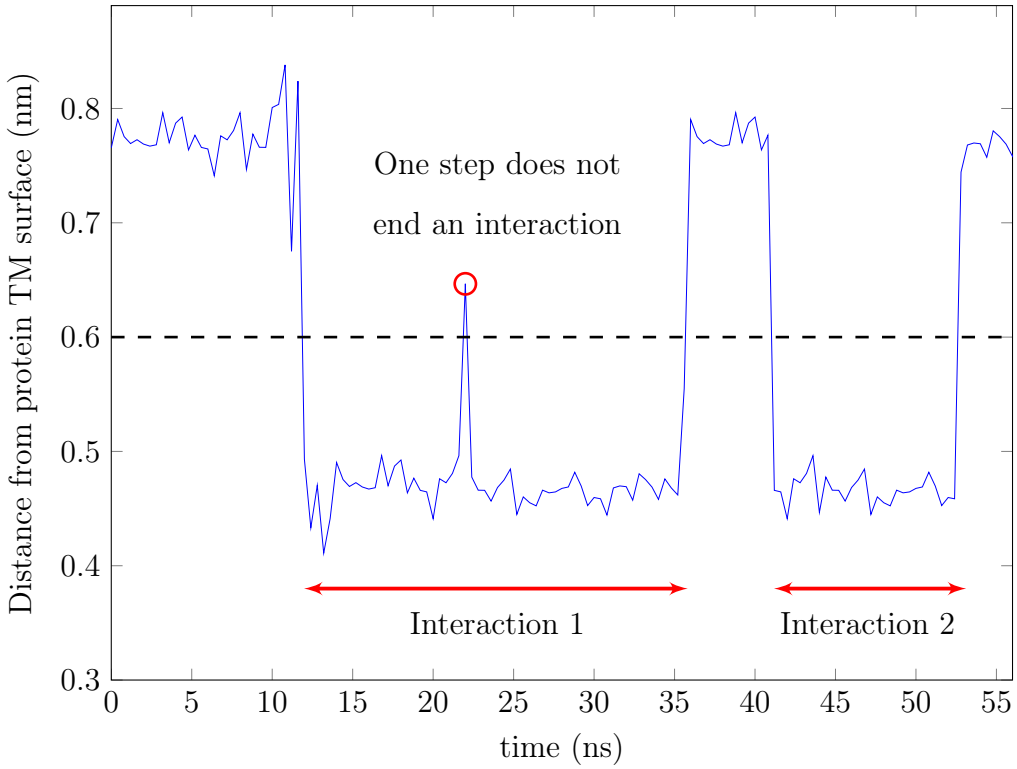


Figure 4.1 *The definition of an interaction time. An interaction begins when the minimum distance between a lipid and the protein TM surface is less than 0.6 nm and ends when the distance is more than 0.6 nm during at least two consecutive time steps.*

4.3.4 Lateral radial distribution functions

Lateral radial distribution functions (RDFs) were calculated to study the lateral distribution of every lipid species in the membrane plane two-dimensionally (2D). RDFs are calculated with a GROMACS program *gmx rdf*.

The radial distribution functions are calculated around the center of mass of the protein transmembrane helices separately for all lipid species. Secondly, RDFs are also calculated for each lipid species around CHOL molecules. The lipid molecules are analyzed bead by bead. The functions were calculated using the last 100 ns (900–1000 ns) of each trajectory and finally averaged over all simulation repetitions of each system.

4.3.5 Density maps

We plotted lateral 2D number density maps of each lipid species to show the average density of the lipids in the bilayer plane around the different proteins. The density maps were calculated separately for the extracellular and intracellular leaflets with a GROMACS program *gmx densmap*.

Before calculating the density maps, the protein was fitted to the same reference structure in each simulation, and in each frame the rotation and translation of the protein was removed. All trajectories were analyzed from the bilayer assembly to the end of the simulations. The result is a set of heat maps showing the average density of each lipid type in the membrane plane averaged over both leaflets, over time, and over all simulation repetitions of each system. The density is divided by the number of analyzed lipids to allow comparison between different lipid species.

4.3.6 The symmetry of bilayer leaflets

We analyzed the symmetry of self-assembled bilayer leaflets. To analyze the symmetry, the bilayer was divided to intracellular and extracellular leaflets. The lipids in the extracellular leaflets have a position above the center of the bilayer and vice versa. To find the center of the bilayer, a mass density profile of all lipids was calculated across the bilayer normal with a GROMACS program *gmx density*. The density profiles were calculated for trajectories where the protein was centered to the simulation box and the results are time averages from the bilayer assembly to the end of the simulations. The local lowest point in the middle of the density profile is the bilayer center, marked in Figure 4.2.

A script was written to divide the lipids to extracellular and intracellular leaflets according to their position in the bilayer. Most lipid types in physiological plasma membranes have a phosphate group which is usually located near the membrane-water interface due to the hydrophilic nature of the lipid head. For lipids containing a phosphate group, it was used to divide the lipids to separate leaflets: if the center of mass of the lipid phosphate group is located above the center of the bilayer, the lipid is in the extracellular leaflet and vice versa.

CHOL, PODG, and DPCE lipids do not have a phosphate group so these lipids require a different analysis method than lipids with a phosphate group. Additionally,

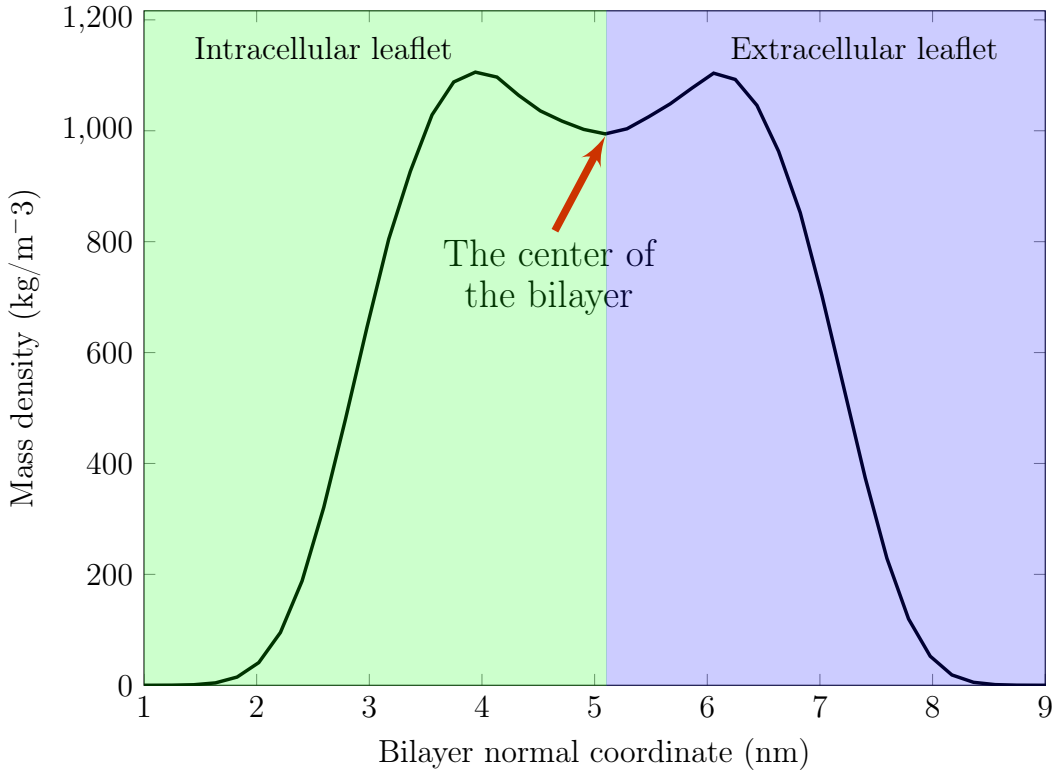


Figure 4.2 *The mass density profile of all lipids in one simulation. The center of the bilayer is marked with an arrow.*

these lipids are less amphiphilic than other studied lipids and therefore these lipids may reside in multiple different orientations in bilayers. Membrane curvature of a portion of studied bilayers adds an additional challenge in assigning these lipids to the right leaflets. For the lipids without a phosphate group, the center of mass of the lipid was used as the coordinate for lipid position. Next, the closest POPC phosphate group to the lipid was searched. Finally, the lipid was assigned to the right leaflet according to the position of the closest POPC phosphate group.

The symmetry of the leaflets was analyzed separately for every simulation in the end of each simulation. The result is the average over all repetition simulations for each system. The error is the standard error of the mean where the sample is the means of the symmetry in separate simulations.

4.3.7 Analysis of protein radius

We calculated the radii of the proteins' intracellular and extracellular portions. First, the proteins were centered to the simulation box. Then, a mass density profile of all lipids was calculated to find the center of the bilayer. The protein beads were divided to intra- and extracellular parts according to their position with respect to the center of the bilayer.

Then, the radius of gyration (R_g) of both parts were calculated with a GROMACS analysis tool *gmx gyrate*. R_g is related to the mass moment of inertia as follows

$$R_g = \sqrt{\frac{I_{axis}}{m}}, \tag{4.2}$$

where the I_{axis} is the moment of inertia about the bilayer normal axis. When the protein is approximated as a cylinder, the mass moment of inertia is given by an equation $I_{axis} = \frac{1}{2}mR^2$ where R is the radius of the protein. Combining the above two equations, the radius of the protein can be obtained with an equation

$$R = \sqrt{2 * R_g^2}. \tag{4.3}$$

We also calculated the error of the result. The given error is the standard deviation.

4.3.8 Analysis of local membrane properties

We analyzed the local membrane properties with a GROMACS tool *g_ lomepro* [86]. The local thickness and area per lipid were analyzed with this tool in the membrane plane.

Thickness and area per lipid were analyzed from the assembly of the bilayer to the end of the simulation separately for each simulation. In each simulation, the protein was fixed to the same reference orientation in the simulation box so that the rotation and translation of the protein were removed.

5 RESULTS AND DISCUSSION

In this thesis, we studied the effect of GPCR proteins on the lateral and transmembrane distributions of lipids. To study the lipid composition around the three different GPCR proteins, we performed a total of 150 MD simulations. Each simulation contained one of the three proteins and 517 lipids. The composition of all systems is described in Table 4.1. The most important analysis results are presented in this Chapter. The results for lipid species POPA, PODG, and DPCE are not shown due to their low concentration that does not result in sufficient statistics to make strong conclusions. The analysis methods are described in Section 4.3. The presented graphs were created in LaTeX and pictures in VMD [87].

5.1 Lipid self-assembly into bilayers

We initially placed the lipids randomly in the simulation box. Amphiphilic lipids can form different lamellar and non-lamellar structures [88], and also during our simulations molecular interactions caused formation of different structures of lipid molecules. In this study, we focus only on lamellar bilayers formed around the protein.

In order to study the equilibrium behavior of lamellar bilayers, the time of bilayer assembly was determined. The APL is an indicator of membrane's phase [85, p. 420] and therefore APL was used to determine the state of self-assembly. We calculated the time development of area per lipid for each simulation. The self-assembly of a bilayer is defined so that it takes place when the area per lipid reaches the value of 0.55 nm^2 and stays between values 0.53–0.57 nm^2 to the end of the simulation.

The time development of area per lipid is shown for one representative β_2AR simulation in Figure 5.1. The lipids self-assemble into a bilayer via three general phases. These phases can be detected from the time development of area per lipid.

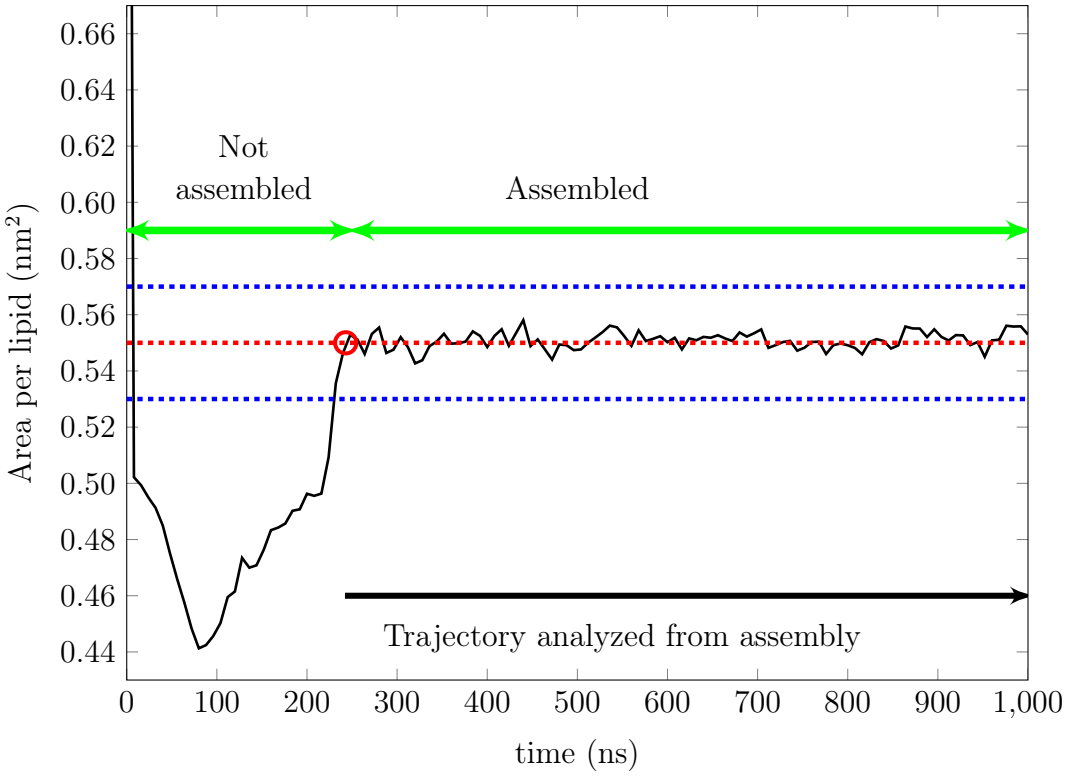


Figure 5.1 *The APL for one representative self-assembly β_2AR simulation. The bilayer is considered to be formed when area per lipid reaches the value of 0.55 nm^2 (red dashed line) and stays between values of 0.53–0.57 nm^2 (blue dashed lines) to the end of the simulation.*

In the beginning of a simulation, there is a large hydrophobic surface of lipids and a protein exposed to the solvent due to their random configurations. This drives the lipids to aggregate with the protein. This can be seen in Figure 5.1 as a rapid drop in APL within the first circa 90 ns of the simulation. As the simulation progresses, the lipids start to assemble towards a lamellar bilayer. At this phase, the APL increases since a lamellar phase occupies a larger area compared with the non-lamellar phase. The APL increases until it reaches approximately the value of 0.55 nm^2. When the area per lipid is more than 0.55 nm^2, all lipid headgroups and water have left the non-polar region of the bilayer and the bilayer is fully assembled. After the assembly, the APL is substantially more stable than during the assembly process. This general pathway of lipid self-assembly into a bilayer is also reported in other MD studies using all-atom and united atom models [77, 78, 79, 80].

Even though the self-assembly is defined numerically, the assembly of the bilayer has additionally been verified visually for each simulation: In Figure 5.2, there is

a β_2AR system where the bilayer has nicely formed around the protein at the time when APL has reached the value of 0.55 nm^2.

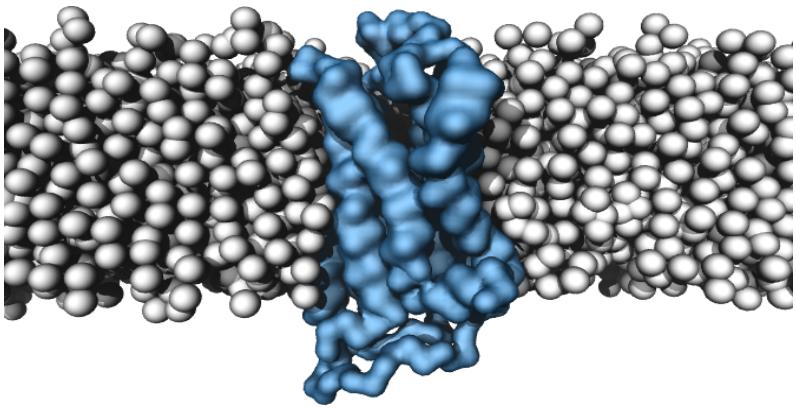


Figure 5.2 *A β_2AR system at the time of assembly where the bilayer has assembled around the protein. Protein (blue) and all lipids (white) are illustrated in the figure.*

The time development of area per lipid is shown for all assembling β_2AR systems in Figure 5.3. For all assembly simulations, the APL shows the same general phases of assembly as described above. The bilayer formation time is shown for all successful simulations in Table 5.1. The time for bilayer formation varies substantially in different simulations even within the repetition simulations including the same protein. Only systems assembling to a bilayer structure in less than 500 ns are analyzed in order to have sufficient statistics for the equilibrated part of the simulations.

The bilayer formation takes from approximately 20 to over 800 ns. A fraction of simulations does not self-assemble into lamellar bilayers at all during 1000 ns. Out of 50 simulations, 64 %, 50 %, and 26 % of the simulations of β_2AR, μ-OR, and Rho systems assemble into bilayers around the protein within 1000 ns, in respective order. Therefore, the β_2AR system has a greater number of successful self-assembly processes than μ-OR or Rho systems. The average bilayer formation times are 200 ns, 310 ns, and 230 ns for β_2AR, μ-OR, and Rho systems, respectively.

As in our study, bilayer formation times have been found to vary greatly also in other self-assembly studies [77, 80]. Skjevik *et al.* [77] showed self-assembly of four different types of phospholipids in separate systems assembling into bilayers in less than 1 microsecond. In their study, the bilayer formation took from 35 to 755 ns in all-atom molecular simulations. The bilayer formation times vary greatly even for the different repetition simulations of the same system, similarly to our results.

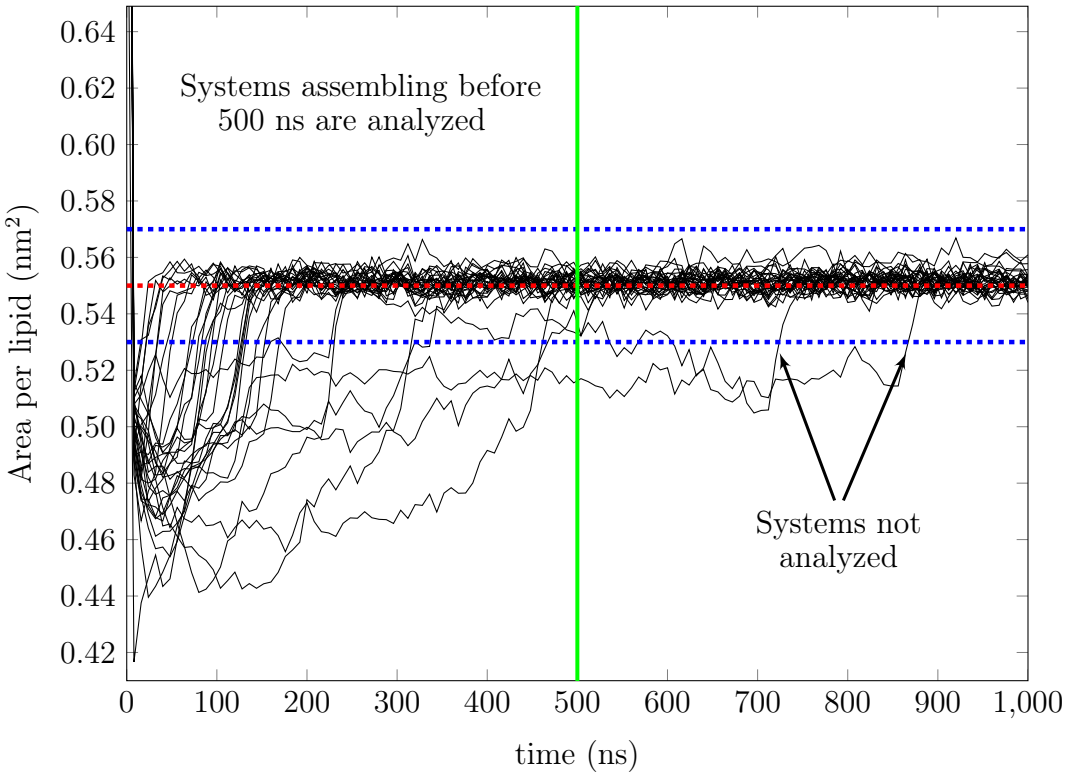


Figure 5.3 *The area per lipid for all self-assembly β_2AR simulations.*

Table 5.1 *The bilayer formation time for all assembling systems simulated in this work. The simulation repetitions of each system are sorted in ascending order based on bilayer formation time.*

(a) β_2AR assembly

Sim. no.[a]	Time (ns)
33	23.6
22	30.0
7	35.6
36	45.6
37	45.6
46	70.0
5	73.2
17	86.4
16	94.4
11	99.6
41	100.0
49	100.0
23	112.0
31	118.4
40	134.8
19	136.0
14	136.8
15	142.0
29	144.0
10	160.8
12	161.2
9	172.8
3	200.0
45	200.0
39	237.6
6	243.2
50	345.6
28	370.4
32	486.8
35	542.0
47	751.6
38	889.2

(b) μ-OR assembly

Sim. no.[a]	Time (ns)
4	64.0
37	80.8
47	91.2
23	99.6
32	110.8
25	120.0
29	134.0
39	145.2
14	146.8
48	149.2
45	158.8
18	164.8
20	176.8
28	202.4
41	220.0
12	264.0
26	281.2
40	312.8
19	385.2
10	468.4
17	694.0
34	699.6
21	700.0
9	876.8
6	964.0

(c) Rho assembly

Sim. no.[a]	Time (ns)
8	40.4
44	64.8
26	86.8
47	90.0
23	115.2
7	145.2
48	147.6
28	208.0
15	276.0
18	284.8
29	306.8
25	508.8
43	722.0

[a] Simulation number.

5.2 Contact preferences between the different lipids and proteins

We calculated the number of contacts of every lipid species with each studied protein. The number of contacts of a specific lipid with the protein divided by the number of total contacts is presented in Table 5.2. Consequently, the results describe the proportion of contacts that one lipid species occupies out of all contacts. The results can be compared with the concentration of each lipid type that is also presented in Table 5.2. The lipids having larger/smaller proportion of all contacts than their concentration are highlighted in red/blue, respectively.

Table 5.2 *The average proportion of total contacts with the TM part of each protein for all lipid species.*

	The proportion of all contacts (%)			Lipid concentration
Lipid	β_2AR	μ-OR	Rho	(%)
POPC	15.5 ± 0.4	14.9 ± 0.4	13.9 ± 0.6	21.3
DPSM	2.0 ± 0.2	2.0 ± 0.2	2.0 ± 0.3	4.1
DXSM	6.2 ± 0.3	7.4 ± 0.3	6.2 ± 0.5	11.8
POPE	5.1 ± 0.2	4.2 ± 0.3	4.5 ± 0.3	6.8
POPI	4.1 ± 0.3	3.7 ± 0.4	3.9 ± 0.4	2.5
POPS	1.2 ± 0.2	1.4 ± 0.2	1.1 ± 0.2	1.4
DOPC	1.7 ± 0.2	1.6 ± 0.2	2.0 ± 0.3	2.1
DOPE	2.9 ± 0.2	2.6 ± 0.3	3.0 ± 0.2	3.3
PAPC	10.4 ± 0.4	10.7 ± 0.4	11.5 ± 0.5	6.2
PAPE	10.5 ± 0.5	12.9 ± 0.5	15.5 ± 0.9	6.8
PAPS	5.9 ± 0.4	7.6 ± 0.5	7.6 ± 0.7	2.9
PUPS	3.6 ± 0.3	2.8 ± 0.3	5.0 ± 0.5	1.4
CHOL	27.8 ± 0.2	25.5 ± 0.3	20.0 ± 0.4	27.3

Lipids POPI, PAPC, PAPE, PAPS, and PUPS occupy a significantly larger proportion of all contacts than their concentration in the systems. The number of contacts of CHOL is slightly larger than its concentration in β_2AR system but smaller than that in the μ-OR and Rho systems. All other lipid species have a low number of contacts compared with the concentration of these lipids in the systems.

According to the number of contacts analysis, all polyunsaturated lipids including four or more double bonds and POPI (a monounsaturated lipid species) are enriched to the protein surface, whereas other monounsaturated lipids do not reside near the protein in large concentrations. The result is significant, since the differences

between the numbers of contacts and lipid concentrations are major and exceed the error bars. For instance, the proportion of PAPS lipid contacts out of all contacts is more than twice its concentration in all systems. Additionally, the results are very consistent between the different systems.

5.3 The interaction times of the lipids with the proteins

The above results show the number of contacts of each lipid species with the protein. Here, we present the lengths of these contacts. The average interaction times of each lipid species with the proteins are shown in Table 5.3. In the table, the lipid species having longer/shorter than 1.5 ns average interaction time are presented with red/blue background, respectively.

Table 5.3 *The average interaction time of lipid species with each protein.*

	The average interaction time (ns)		
Lipid	β_2AR	μ-OR	Rho
POPC	1.02 ± 0.02	0.89 ± 0.02	0.83 ± 0.02
DPSM	0.81 ± 0.04	0.77 ± 0.05	0.77 ± 0.06
DXSM	0.73 ± 0.03	0.67 ± 0.02	0.63 ± 0.04
POPE	1.02 ± 0.03	0.84 ± 0.04	0.84 ± 0.07
POPI	2.5 ± 0.2	2.6 ± 0.4	2.2 ± 0.5
POPS	1.2 ± 0.2	1.4 ± 0.2	1.01 ± 0.08
DOPC	1.11 ± 0.09	1.05 ± 0.07	1.1 ± 0.2
DOPE	1.22 ± 0.06	1.13 ± 0.08	1.04 ± 0.09
PAPC	2.13 ± 0.08	2.2 ± 0.1	2.20 ± 0.09
PAPE	2.1 ± 0.1	2.6 ± 0.2	2.5 ± 0.2
PAPS	3.1 ± 0.2	3.8 ± 0.3	3.0 ± 0.3
PUPS	3.8 ± 0.4	4.0 ± 0.5	4.5 ± 0.5
CHOL	3.11 ± 0.06	2.30 ± 0.06	1.69 ± 0.06

The lipids POPI, PAPC, PAPE, PAPS, PUPS, and CHOL have longer than 1.5 ns average interaction time with all proteins. Lipids POPC, DPSM, DXSM, POPE, POPS, DOPC, and DOPE have shorter than 1.5 ns average interaction time in all systems. PUPS lipids have the longest and DXSM the shortest average interaction time, PUPS having more than a five times longer average interaction time than DXSM.

We also calculated the distribution of the interaction times. The distribution of the interaction times of CHOL with β_2AR is shown in Figure 5.4 as an example.

Even though the average interaction times (Figure 5.3) are only a few nanoseconds, some lipid molecules are in contact with the proteins for hundreds of nanoseconds. Those lipids that stay in contact with the protein for substantially long times are often located in a pocked of a protein. For example, a cholesterol molecule stays in a pocked between helices 2 (H2) and 4 (H4) of β_2AR for hundreds of nanoseconds, see Figure 5.5. This place is also a suggested binding site of cholesterol [29, 42] reported in previous studies. This suggests that our model is sufficiently accurate to reveal possible lipid-binding sites.

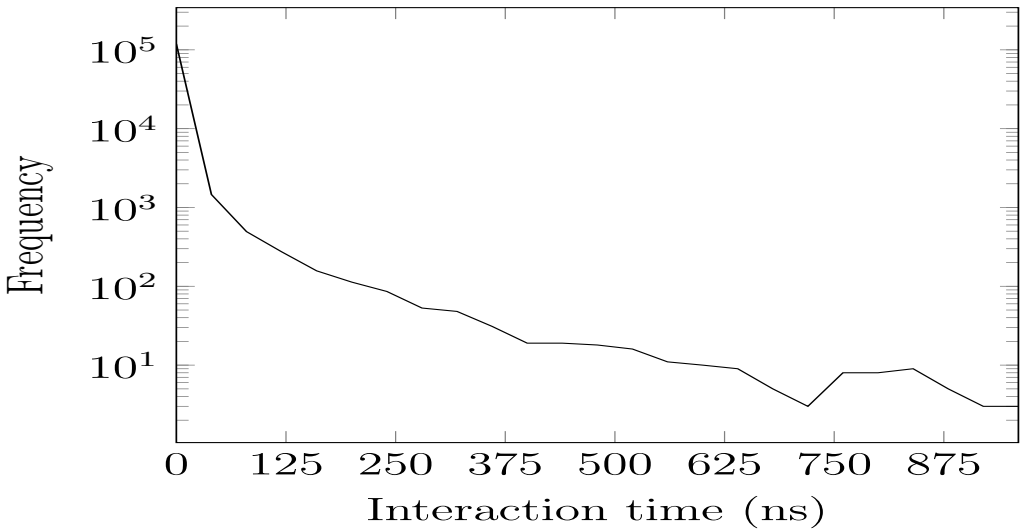


Figure 5.4 *The distribution of CHOL interaction times with β_2AR protein. The result is averaged over all simulation repetitions.*

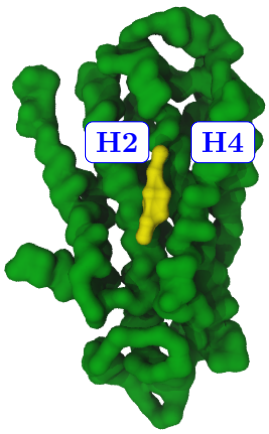


Figure 5.5 *A cholesterol (yellow) molecule in a pocket between helices 2 (H2) and 4 (H4) of β_2AR.*

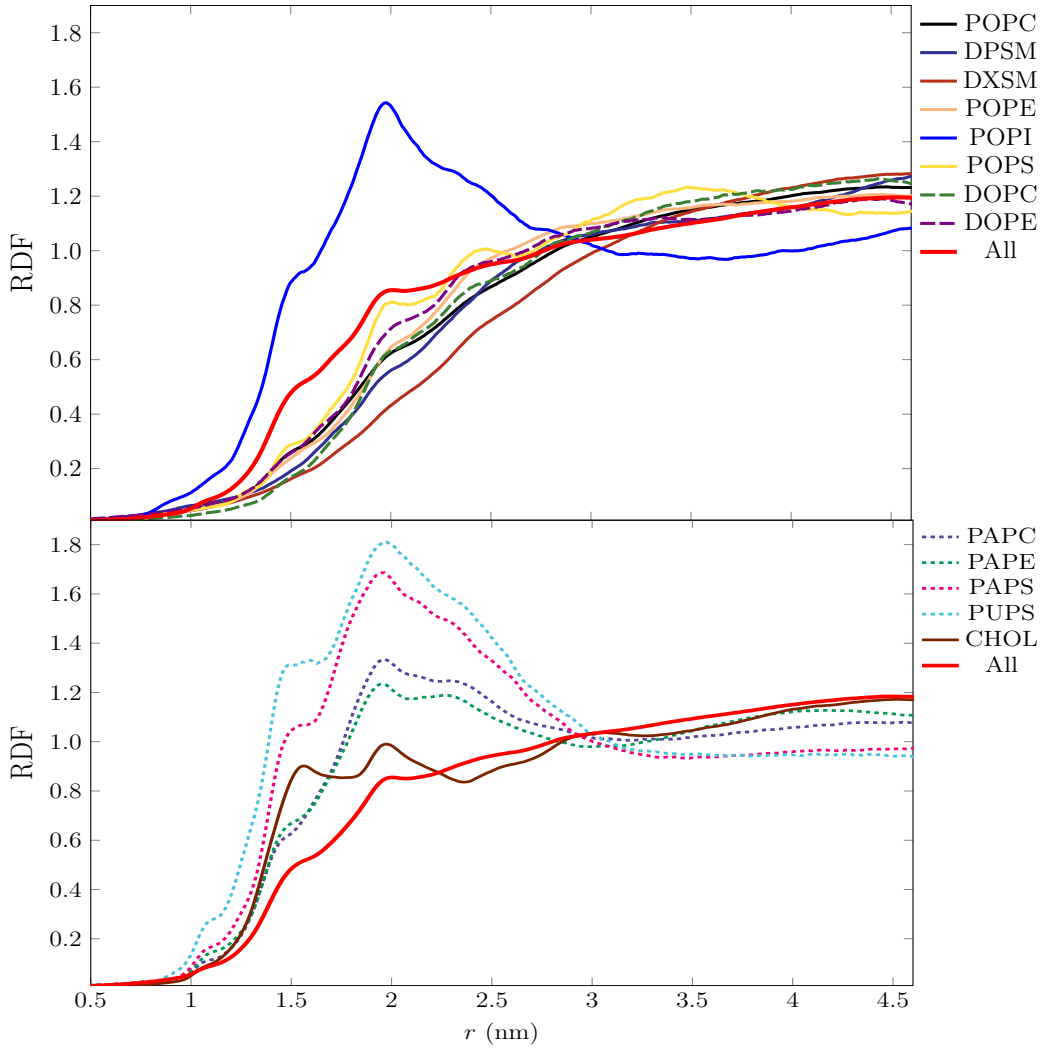


Figure 5.6 *The average radial distribution functions for each lipid species around the center of mass of the $\beta_2 AR$ protein.*

5.4 Radial distribution functions show surface preference for unsaturated tails, phosphatidylinositol, and cholesterol

We computed two dimensional radial distribution functions (RDFs) in the membrane plane to study the packing of different lipid species against the three studied proteins. The average RDFs are presented for β_2AR system in Figure 5.6. The results for μ-OR and Rho systems are given in Appendix A. The results for the lipid species containing one or two double bonds are presented in the upper panel and lipids containing four or more double bonds and CHOL are presented in the lower panel. The results can be compared with the averaged RDF for all lipids shown in red.

We see a considerable difference in the RDFs between the different lipid species. In all systems, the RDF profile for POPC, DPSM, DXSM, POPE, POPS, DOPC, and DOPE peak beyond 3 nm, whereas POPI, PAPC, PAPE, PAPS, PUPS, and CHOL are enriched closer to the protein surface than the other lipids.

To study which parts of the lipids are most in contact with the proteins we also calculated the average lateral RDFs separately for the headgroup and both tails of each lipid species around the proteins' center of mass. The RDFs for POPC, PAPE, and POPI lipids are presented as examples for the β_2AR system in Figure 5.7. All monounsaturated lipid species, except for POPI, do not have any preferred part facing the proteins' surface (see the uppermost panel of Figure 5.7). Meanwhile, the headgroup of POPI is preferred over its tails to the protein surface in all systems (see the middle panel of Figure 5.7). For all lipid species containing four or more double bonds, such as PAPE, the polyunsaturated lipid tail is on average found closer to the protein surface than the saturated tail (see the lowest panel of Figure 5.7).

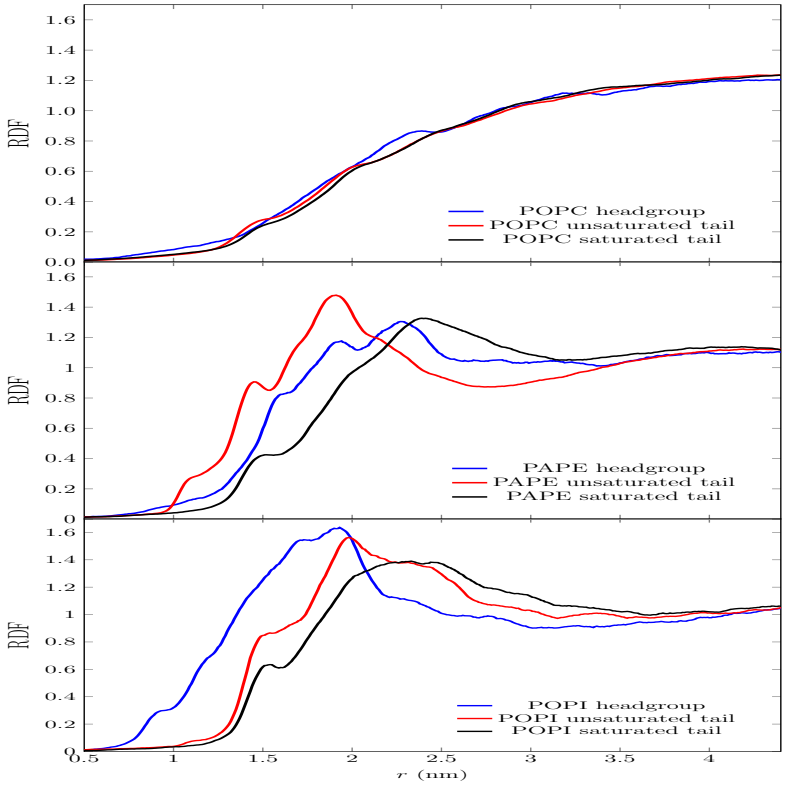


Figure 5.7 *Radial distribution functions for lipids' headgroup and tails around the β_2AR protein.*

We also calculated lateral RDFs for all lipid species around CHOL molecules since cholesterol is suggested to affect lipid raft formation [89]. The results are presented in Figure 5.8 for the β_2AR system. In every system containing a different GPCR protein, cholesterol molecules are found, on average, nearer monounsaturated lipids than polyunsaturated lipids.

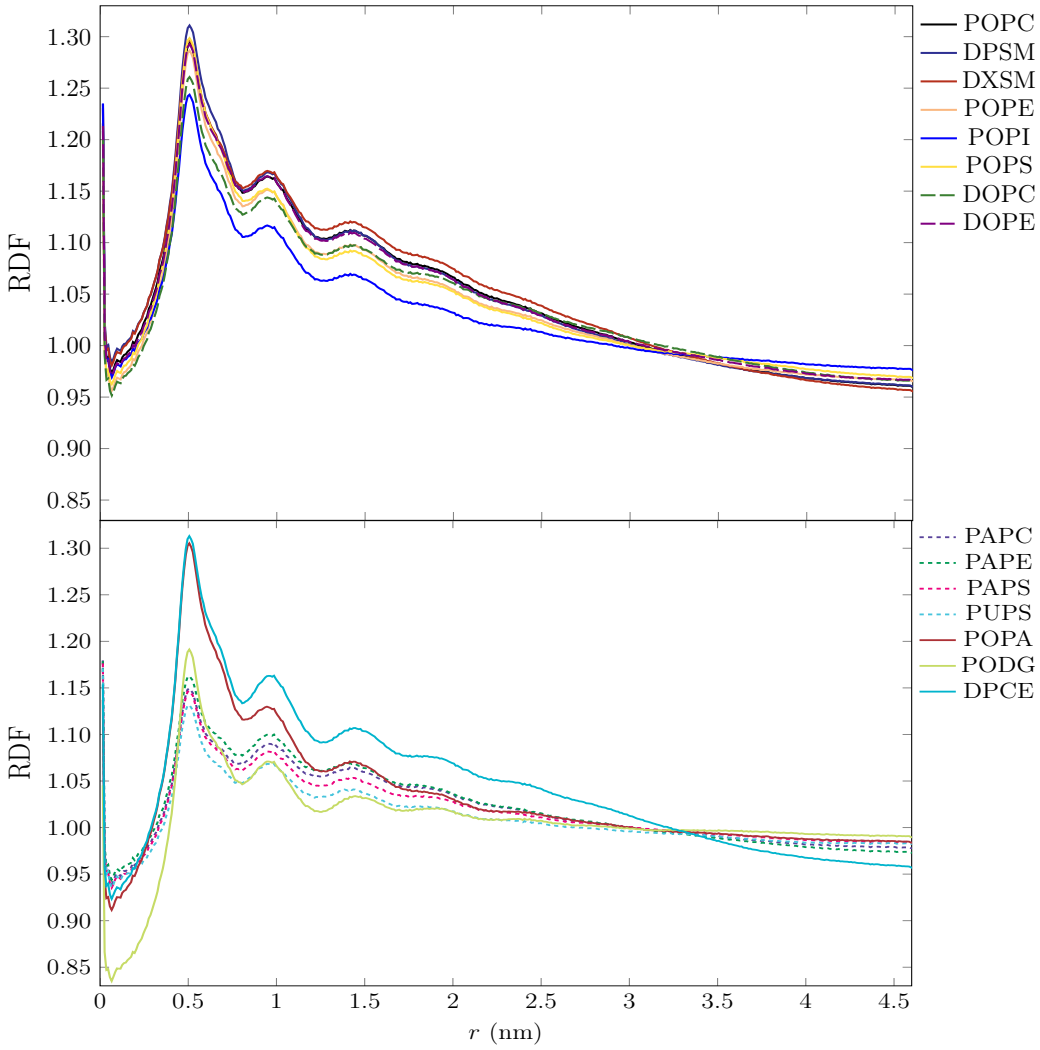


Figure 5.8 *The average radial distribution functions for different lipids around cholesterol in the β_2AR system.*

There is a wide range of studies regarding how cholesterol interacts with different lipid species. Similarly to our results, previous studies have shown that cholesterol interacts stronger with saturated than unsaturated lipids [89, 90]. Therefore, it is likely that cholesterol is enriched near the proteins' surface due to the proteins'

properties and not because of the lipid environment.

5.5 Density maps show polyunsaturated lipid, phosphatidylinositol, and cholesterol preference

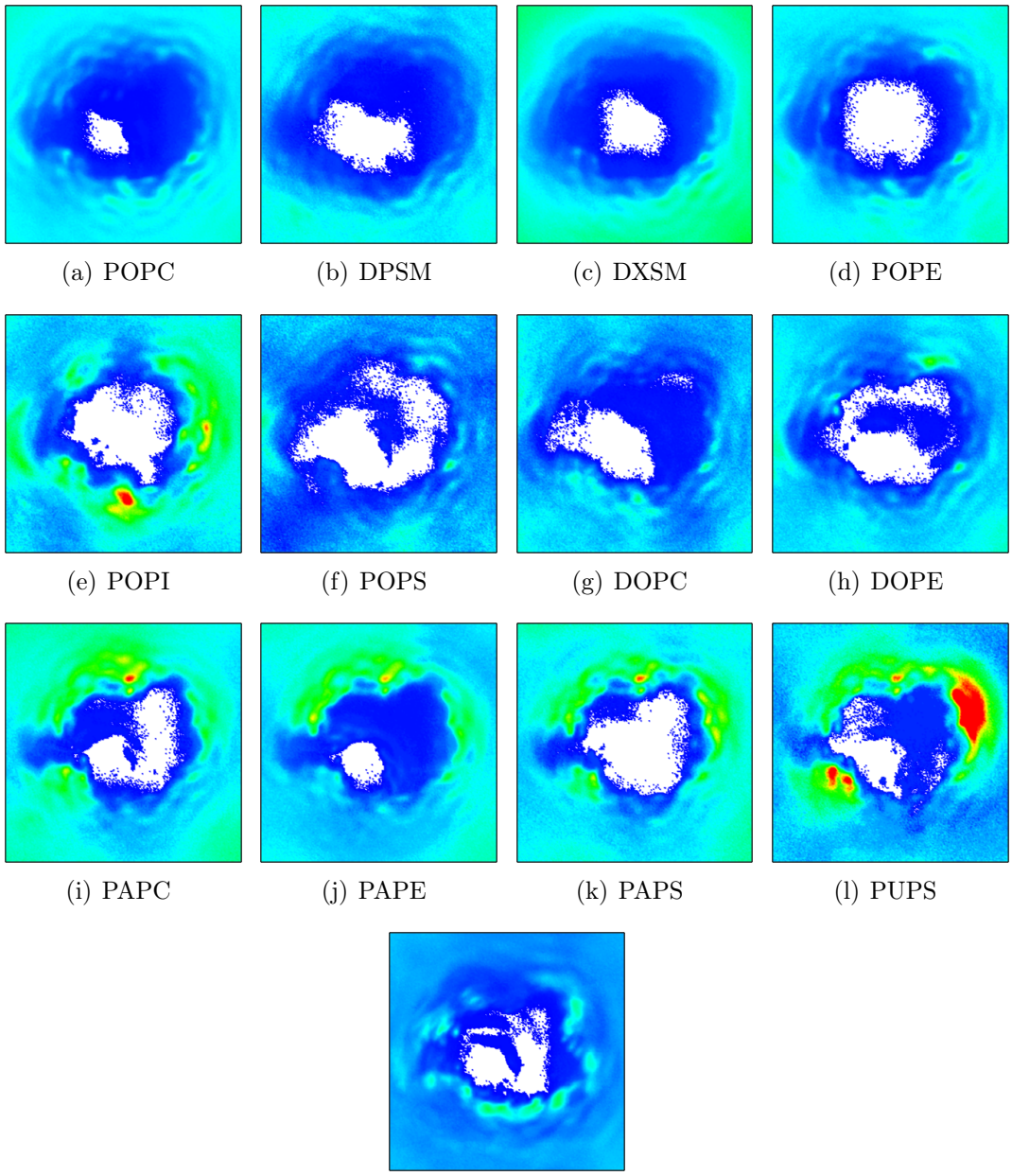
Lateral RDF profiles above show clear differences in the preference of different lipids to reside next to the surface of the proteins. Nevertheless, RDFs contain limited information since they do not distinguish the different portions of the protein surface or the different bilayer leaflets. Therefore, we show lateral density heat maps in the membrane plane for all lipid compounds separately in the extracellular and intracellular leaflets.

Figures 5.9 and 5.10 present the number density maps for different lipid species around β_2AR separately in the extracellular and intracellular leaflets. The white region in the center of the figures is the excluded volume of the protein. The red regions represent high density, and blue regions the low density. The protein is aligned to the same orientation in each figure and the heat map densities are divided by the number of lipids in question so that the figures can be compared.

In the density map figures, POPI, PAPC, PAPE, PAPS, PUPS, and CHOL show peaks near the protein surface both in the extracellular and intracellular leaflets in all systems. However, POPI, PAPE, PAPS, and PUPS are naturally found in the intracellular leaflet and therefore the lateral organization of these lipids in the extracellular leaflet is biologically irrelevant. POPC, DPSM, DXSM, POPE, POPS, DOPC, and DOPE show low densities near the protein surface in all systems compared with the other lipid species.

The density heat maps show that POPI lipids are enriched in two spots near helices 4 and 6 of β_2AR in the intracellular leaflet. The RDFs (Figure 5.6) reveal that the headgroup of POPI is more likely to interact with the protein than either of its tails. These results suggests that POPI may interact specifically with the studied proteins. To the best of our knowledge, this is the first study to show a possible specific interaction of POPI lipids with β_2AR, μ-OR, and Rho proteins.

Polyunsaturated lipids are found on average nearer the protein surfaces than most monounsaturated lipids. Polyunsaturated lipids are enriched around nearly the whole proteins according to the density heat maps. RDF profiles (Figure 5.6)



(a) POPC (b) DPSM (c) DXSM (d) POPE

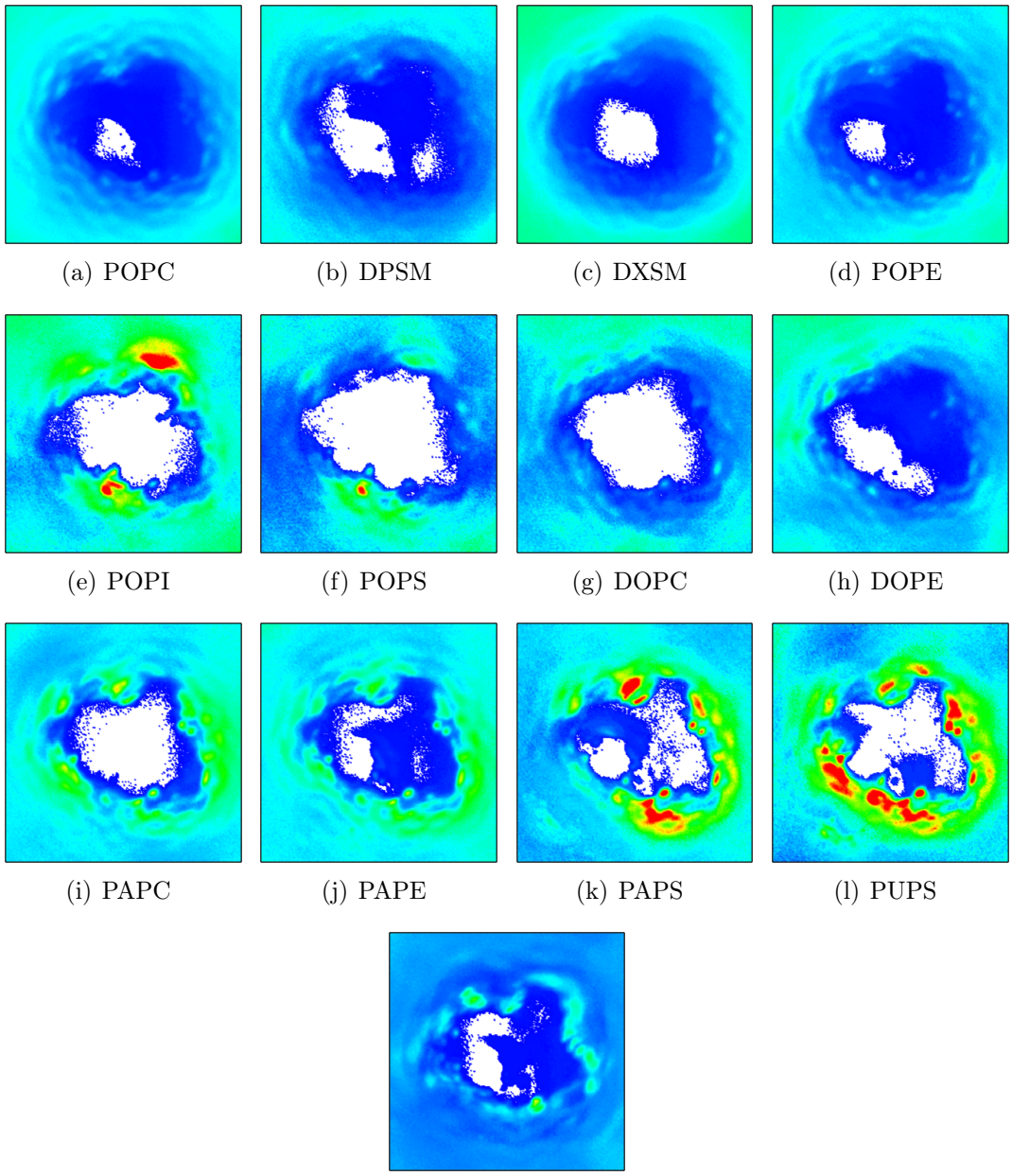
(e) POPI (f) POPS (g) DOPC (h) DOPE

(i) PAPC (j) PAPE (k) PAPS (l) PUPS

(m) CHOL

Figure 5.9 *The number density maps of different lipids around β_2AR in the extracellular leaflet. High density regions are red, and low density regions are blue.*

show that the polyunsaturated tail is preferred to the proteins' surface. Similarly to our results, a few previous studies have shown that polyunsaturated lipid tails are preferred to the surfaces of GPCR proteins [91, 92]. For example, Horn *et al.* [91, p. 75–94] showed by molecular dynamics simulations that polyunsaturated DHA



(a) POPC (b) DPSM (c) DXSM (d) POPE

(e) POPI (f) POPS (g) DOPC (h) DOPE

(i) PAPC (j) PAPE (k) PAPS (l) PUPS

(m) CHOL

Figure 5.10 *The number density maps of different lipids around $\beta_2 AR$ in the intracellular leaflet. High density regions are red, and low density regions are blue.*

chains are found in higher concentrations at the surfaces of rhodopsin and opsin than stearoyl chains. Our study shows more general results with substantially more complex systems. Our results reveal that all lipids in our systems containing four or more double bonds are enriched to the surfaces of all three studied GPCR proteins.

Brewster and Safran [93] have suggested that lipids having one fully saturated tail and another partially unsaturated tail could act as surface active components. These kinds of lipids could lower the line tension between different membrane environments by having the saturated tail preferring the more ordered side and the unsaturated tail orienting towards the less ordered environment. Polyunsaturated lipids can pack towards curved interfaces that saturated lipids cannot due to their relatively straight and rigid tails. Polyunsaturated lipids could lower the energetic cost of incorporating a protein into an ordered lipid bilayer by packing at curved interfaces. Therefore, the preference of polyunsaturated lipids toward the proteins may be entropically driven.

Cholesterol is known to be important for GPCR function [22]. Previous studies have provided possible cholesterol-binding sites at least in the β_2AR [29, 42] and Rho [43] GPCR proteins. Our study shows similar results with multiple possible cholesterol-binding sites in all studied proteins. Our MD simulations show that cholesterol spontaneously and recurrently diffuses to multiple specific sites near the proteins' surfaces. This supports the previous knowledge that cholesterol is important for GPCRs.

The results regarding the lateral organization of lipids obtained with different analysis methods are very consistent. Additionally, the lipids interact with all the three studied GPCR proteins in a very similar way. The studied lipid species are found to interact with the proteins with different affinities. Polyunsaturated lipids including four or more double bonds and POPI are found to interact with the proteins more than the other lipid species (according to the number of contacts analysis). Also, they have longer interaction times with the proteins than the other studied lipid species (according to the interaction time analysis). Further, polyunsaturated lipids reside nearer the proteins' surface than all lipids on average (according to the RDF profiles and density heat maps). As a summary, the GPCR proteins induce domain formation where polyunsaturated lipids with four or more double bonds, cholesterol, and a monounsaturated lipid POPI are enriched near all studied proteins. All other studied lipids are enriched farther than 3 nm from the proteins' center of mass.

5.6 Self-assembled membranes are on average symmetrical

We calculated the transmembrane symmetry of the bilayer leaflets in all systems. The results are presented in Table 5.4 and show that the membrane leaflets are

symmetrical when the results are averaged over all simulation repetitions of each system. However, the error for the symmetry of a lipid type can be as much as 6 percent: when the bilayers are studied separately they can have highly asymmetrical transmembrane lipid compositions. Bilayers studied separately can have even dozens of lipids more in one leaflet than in the other. Also the study of Skjevik *et al.* [77] showed that separate self-assembled membranes can be asymmetrical.

Table 5.4 *The average percentage of each lipid type in the intracellular leaflet in the end of the simulations.*

	The percentage of lipids in the intracellular leaflet (%)		
Lipid	β_2AR	μ-OR	Rho
POPC	51 ± 1	49 ± 1	50 ± 2
DPSM	50 ± 2	50 ± 3	54 ± 4
DXSM	49 ± 2	50 ± 2	51 ± 2
POPE	51 ± 2	49 ± 2	50 ± 3
POPI	53 ± 3	56 ± 4	47 ± 4
POPS	54 ± 4	54 ± 5	44 ± 6
DOPC	50 ± 4	50 ± 4	52 ± 5
DOPE	52 ± 3	49 ± 3	54 ± 4
PAPC	48 ± 2	50 ± 2	51 ± 4
PAPE	50 ± 2	50 ± 2	46 ± 3
PAPS	51 ± 3	58 ± 4	48 ± 5
PUPS	57 ± 3	51 ± 5	46 ± 6
CHOL	51 ± 1	50 ± 1	51 ± 1
All lipids	50 ± 1	50 ± 1	50 ± 1

There are on average more lipids in the intracellular than in the extracellular leaflet in all systems. Approximately, there are 2–4 lipids more in the intracellular than in the extracellular leaflet. To estimate the effect of the proteins' volume, we calculated the radii of the proteins' extracellular and intracellular parts. The results are shown in Table 5.5.

Table 5.5 *The radii of the proteins' intracellular and extracellular parts.*

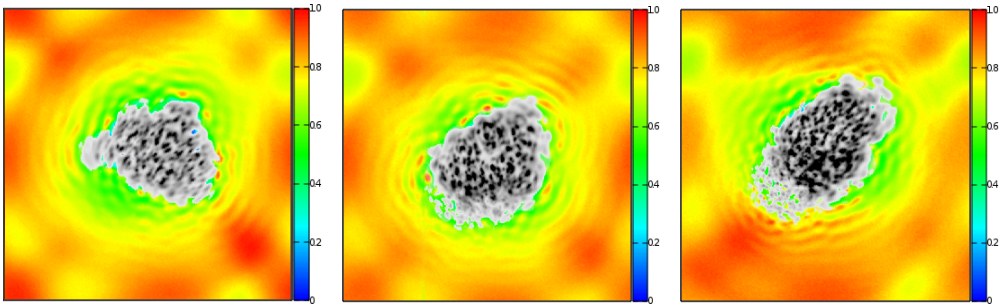
	Intracellular radius (nm)	Extracellular radius (nm)
β_2AR	1.74 ± 0.02	1.85 ± 0.04
μ-OR	1.80 ± 0.02	1.82 ± 0.02
Rho	1.68 ± 0.02	1.85 ± 0.02

The extracellular part of all proteins is larger than their intracellular part. The difference between the radii of intracellular and extracellular parts is the largest in

the Rho protein and smallest in μ-OR. The β_2AR system has the largest difference in the average number of lipids in the extra- and intracellular leaflets. According to the radii, the β_2AR protein occupies 1.24 nm^2 more area in the extra- than in the intracellular leaflet. The area per lipid is circa 0.55 nm^2 so we would expect to find on average 2.3 lipids more in the intra- than in the extracellular leaflet. For β_2AR, there are on average approximately 2–6 lipids more in the intra- than in the extracellular leaflet when the error is taken into account. Therefore, the proteins' larger volume in the extracellular side compared with the intracellular volume can cause the difference in the found numbers of lipids in distinct leaflets in all systems. The results suggest that GPCR proteins do not affect the transmembrane distribution of lipids in self-assembled membranes but the lipids are equally distributed between the leaflets.

5.7 Local membrane properties

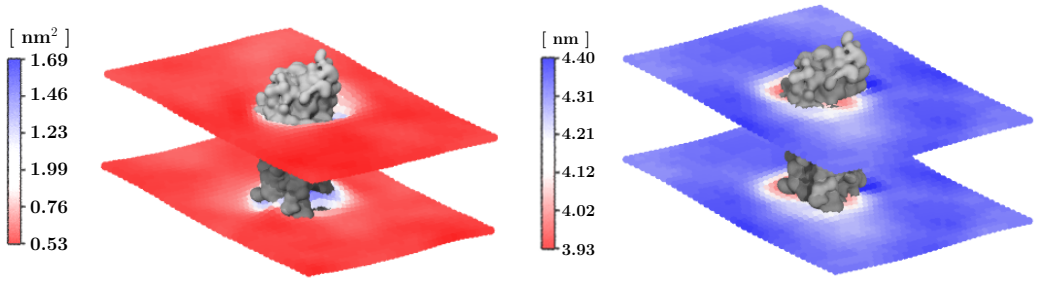
As shown in this thesis, GPCR proteins can affect the lateral lipid distribution but they may also affect other membrane properties. Figure 5.11 shows the lateral number density maps of all lipids in each system averaged over all simulations and both leaflets. The protein is illustrated in gray. The solvation spheres can be seen around the protein. This suggests that the statistics is good in the performed calculations. The figures also show that the number density of lipids is lower near the protein surface than far away from it. Similarly, the RDFs (Figure 5.6) show that the probability of finding any lipid increases as the distance from the protein becomes larger. This suggests that the lipid number concentration is higher far away from the protein than near the protein surface.



(a) β_2AR (b) OPIOID (c) RHOD

Figure 5.11 *The number density maps of all lipids in the systems. Protein is presented in the center of each figure in gray.*

The lipid composition can affect the local APL and thickness of a biological membrane. We calculated the local APL and thickness in all systems separately for all simulation repetitions. The results are shown for one representative β_2AR simulation in Figure 5.12.



(a) APL (b) Thickness

Figure 5.12 *The local membrane properties of the β_2AR system. Protein backbone illustrated in gray in both figures.* ***a.*** *The time averaged APL.* ***b*** *The time averaged local thickness.*

The proteins affect the APL and thickness of the membranes only locally. The results are consistent for all systems including a different GPCR protein. The APL is increased near the proteins' surface. The thickness of the membranes is increased or decreased near the proteins' surface depending on the protein part. The values of the APL and thickness cannot be compared with any previous study due to the complexity of our systems.

6 CONCLUSIONS

There is accumulating evidence that plasma membranes have heterogeneous lateral and transmembrane lipid distributions [47]. The lipid domains with distinct lipid content are suggested to have multiple functions [9, 51, 52]. The lipid environment of integral proteins can influence the protein activity by changing the general bilayer properties or by specific lipid-binding interactions. Despite the proposed importance of lipid rafts, their exact structures, roles, and forces driving domain formation are not well understood [47, 53]. To reveal new insights into lipid domains, the effects of β_2AR, μ-OR, and Rho GPCR proteins on the lateral and transmembrane lipid distributions in plasma membrane models were studied in this thesis.

Lipid domains are difficult to study due to their small size [5, 9], which is not to be resolved by conventional experimental methods. Therefore, this study was conducted using molecular dynamics simulations with a coarse-grained molecular description to obtain near atomistic resolution of the studied systems. The previous challenges in MD simulations studying lipid rafts have been the slow diffusion of lipids in MD time scales. This has caused a challenge, since due to the slow lipid diffusion the starting configuration of a conventionally built MD model bilayer can affect the results if the system is not equilibrated for a sufficiently long time. This challenge was here overcome by using self-assembly of proteins and lipids to obtain unbiased membranes.

This study showed that self-assembly of proteins and lipids can be used to generate unbiased membrane models with MD simulations. The proteins and lipids can be placed randomly to a simulation box and the molecular interactions during an MD simulation cause the self-assembly of the molecules into different structures including lamellar bilayers. When self-assembly is used to generate the membranes, the starting configuration does not affect the results when the resulting data is analyzed statistically for a sufficient number of simulation repetitions.

This study revealed that the three studied GPCR proteins (β_2AR, μ-OR, and Rho)

can induce lateral domain formation in spontaneously formed plasma membrane models. Since the structures of all GPCR proteins are highly conserved [10], the results suggest that all GPCR proteins may induce domain formation in plasma membranes. The lipid distribution around the proteins was analyzed with multiple methods and the results were consistent and significant. The number of contacts, the interaction times, the radial distribution functions, and the density map analysis methods were utilized to study the lateral distribution of lipids. The analysis methods showed that the lipids containing four or more double bonds, a monounsaturated lipid type POPI, and CHOL were enriched to the proteins' surface. All the other studied lipids were enriched more than 3 nm away from the proteins' center of mass. We also showed that the GPCR proteins studied in this thesis do not affect the transmembrane distribution of the lipids since the self-assembled membranes were on average symmetrical even though the separate membranes were slightly asymmetrical.

The RDFs revealed that the POPI headgroup was found on average nearer the proteins' surface than either of its tails. According to the density map analysis POPI was found to interact specifically with the protein at two separate sites. To the best of our knowledge, this is the first study suggesting that there may be two POPI-binding sites at the surfaces of the GPCR proteins. The RDFs also revealed that the unsaturated tails of polyunsaturated lipids were preferred at the protein surface over the saturated tails. Polyunsaturated lipids were found to interact with nearly all parts of the protein TM surfaces according to the density map analysis. Therefore, the polyunsaturated tails may be entropically drawn to the surfaces of the proteins to lower the line tension between the more ordered lipid bilayer and the less ordered protein surface. Additionally, cholesterol was found to diffuse repeatedly to multiple separate sites at the surfaces of the GPCR proteins both in the extra- and intracellular leaflets.

As a conclusion, the GPCR proteins studied in this thesis did not affect the transmembrane distribution of lipids but had a significant effect on the lateral lipid organization. The results of this thesis are significant for understanding lipid domains formed around GPCR proteins. It is known that the lipid environment is very important for the functioning of GPCR proteins. This study revealed that GPCR proteins can themselves induce domain formation in plasma membrane models and therefore affect their lipid environment. We found that certain lipid types, especially POPI and CHOL, interact specifically and reproducibly with particular parts

of the GPCR proteins. These lipids may affect the protein function through direct binding. The lipid content was also significantly different near proteins and far from them. This changes the physical and chemical properties of the membrane which also affects the function of the proteins.

Further examination of topics explored in this thesis is needed to reveal the interactions between the protein and the lipids more specifically. An analysis regarding the interactions between lipids and all amino acids of the proteins would show the more detailed interaction sites of lipids with the proteins and possibly reveal the more detailed binding sites of lipids. Other interesting questions are how GPCRs affect other membrane properties, such as lipid order parameter, and what is the maximum size of a lipid domain induced by a GPCR protein, which would give an estimate of the range of protein's influence on the lipid organization. For the reliability of the results in this thesis, the study needs to be repeated with all-atom MD simulations and experimental methods as far as they are feasible. Nonetheless, MD simulations have been shown to be advantageous in studying lipids rafts and domains to reveal their more detailed biological functions.

BIBLIOGRAPHY

[1] B. Alberts, D. Bray, K. Hopkin, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter, *Essential Cell Biology*. Garland Science, 3 ed., 2010.

[2] P. Niemelä, M. Hyvönen, and I. Vattulainen, "Atom-scale molecular interactions in lipid raft mixtures," *Biochimica et Biophysica Acta*, vol. 1788, no. 1, pp. 122–135, 2009.

[3] G. van Meer, D. Voelker, and G. Feigenson, "Membrane lipids: where they are and how they behave," *Nature Reviews Molecular Cell Biology*, vol. 9, no. 2, pp. 112–124, 2008.

[4] A. Lee, "Lipid-protein interactions in biological membranes: a structural perspective," *Biochimica et Biophysica Acta*, vol. 1612, no. 1, pp. 1–40, 2003.

[5] K. Simons and J. Sampaio, "Membrane organization and lipid rafts," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 10, pp. 1–17, 2011.

[6] Z. Korade and A. Kenworthy, "Lipid rafts, cholesterol, and the brain," *Neuropharmacology*, vol. 55, no. 8, pp. 1265–1273, 2008.

[7] L. Pike, "The challenge of lipid rafts," *Journal of Lipid Research*, vol. 50, no. Suppl, pp. S323–S328, 2009.

[8] D. Marquardt, B. Geier, and G. Pabst, "Asymmetric lipid membranes: towards more realistic model systems," *Membranes*, vol. 5, no. 2, pp. 180–196, 2015.

[9] J. Hancock, "Lipid rafts: contentious only from simplistic standpoints," *Nature Reviews Molecular Cell Biology*, vol. 7, no. 6, pp. 456–462, 2009.

[10] D. Rosenbaum, S. Rasmussen, and B. Kobilka, "The structure and function of G-protein-coupled receptors," *Nature*, vol. 459, no. 7245, pp. 356–363, 2009.

[11] B. Kobilka, "G protein coupled receptor structure and activation," *Biochimica et Biophysica Acta*, vol. 1768, no. 4, pp. 794–807, 2007.

[12] A. Manglik and B. Kobilka, "The role of protein dynamics in GPCR function: insights from the β_2AR and rhodopsin," *Current Opinion in Cell Biology*, vol. 27, pp. 136–143, 2014.

[13] P. Insel, C.-M. Tang, I. Hahntow, and M. Michel, "Impact of GPCRs in clinical medicine: genetic variants and drug targets," *Biochimica et Biophysica Acta*, vol. 1768, no. 4, pp. 994–1005, 2008.

[14] D. Nelson and M. Cox, *Lehninger principles of biochemistry*. Sara Tenney, 5 ed., 2008.

[15] G. Cooper, *The Cell*. Sinauer Associates, 2000.

[16] A. Dupuy and D. Engelman, "Protein area occupancy at the center of the red blood cell membrane," *PLoS Computational Biology*, vol. 105, no. 8, pp. 2848–2852, 2008.

[17] J. Goose and M. Sansom, "Reduced lateral mobility of lipids and proteins in crowded membranes," *PLoS Computational Biology*, vol. 9, no. 4, pp. 1–10, 2013.

[18] T. Heimburg, *Thermal Biophysics of Membranes*. Wiley-VCH, 2007.

[19] M. Sud, D. Cotter, A. Brown, E. Dennis, C. Glass, A. Merrill, R. Murphy, C. Raetz, D. Russel, and S. Subramaniam, "LMSD: LIPID MAPS structure database," *Nucleic Acids Research*, vol. 35, no. Database issue, pp. D527–D532, 2007.

[20] "The LIPID MAPS lipidomics gateway." Available (accessed on 18.9.2015): http://www.lipidmaps.org/.

[21] S. George, B. O'Dowd, and S. Lee, "G-protein-coupled receptor oligomerization and its potential for drug discovery," *Nature Reviews Drug Discovery*, vol. 1, no. 10, pp. 808–820, 2002.

[22] T. Pucadyil and A. Chattopadhyay, "Role of cholesterol in the function and organization of G-protein coupled receptors," *Progress in Lipid Research*, vol. 45, no. 4, pp. 295–333, 2006.

[23] V. Cherezov, D. Rosenbaum, M. Hanson, S. Rasmussen, F. S. Thian, T. S. Kobilka, H.-J. Choi, P. Kuhn, W. Weis, B. Kobilka, and R. Stevens, "High-resolution crystal structure of an engineered human β_2-adrenergic G protein-coupled receptor," *Science*, vol. 318, no. 5854, pp. 1258–1265, 2007.

[24] H. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. Shindyalov, and P. Bourne, "The protein data bank," *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000.

[25] T. Higgins, "Molecular characterization of β_2-adrenergic receptor haplotypes," Master's thesis, University of North Carolina at Chapel Hill, 2008.

[26] M. Johnson, "Molecular mechanisms of β_2-adrenergic receptor function, response, and regulation," *Journal of Allergy and Clinical Immunology*, vol. 117, no. 1, pp. 18–24, 2006.

[27] J. Baker, "The selectivity of β-adrenoceptor antagonist at the human β_1, β_2 and β_3 adrenoceptors," *British Journal of Pharmacology*, vol. 144, no. 3, pp. 317–322, 2005.

[28] S. Dorph and C. Binder, "Evaluation of the hypotensive effect of beta-adrenergic blockade in hypertension," *Journal of Internal Medicine*, vol. 185, no. 1-6, pp. 443–448, 1969.

[29] X. Cang, Y. Du, Y. Mao, Y. Wang, H. Yang, and H. Jiang, "Mapping the functional binding sites of cholesterol in b2-adrenergic receptor by long-time molecular dynamics simulations," *Journal of Physical Chemistry B*, vol. 117, no. 4, pp. 1085–1094, 2013.

[30] A. Manglik, A. Kruse, T. Kobilka, F. Thian, J. Mathiesen, R. Sunahara, L. Pardo, W. Weis, B. Kobilka, and S. Granier, "Crystal structure of the μ-opioid receptor bound to a morphinan antagonist," *Nature*, vol. 485, no. 7398, pp. 321–326, 2012.

[31] R. Santhappan, A. Crowder, S. Gouty, B. Cox, and T. Côté, "Mu opioid receptor activation enhances regulator of G protein signaling 4 association with the mu opioid receptor/G protein complex in a GTP-dependent manner," *Journal of Neurochemistry*, vol. 135, no. 1, pp. 76–87, 2015.

[32] C. Contet, B. Kieffer, and K. Befort, "Mu opioid receptor: a gateway to drug addiction," *Current Opinion in Neurobiology*, vol. 14, no. 3, pp. 370–378, 2004.

[33] J. Li, P. Edwards, M. Burghammer, C. Villa, and G. Schertler, "Structure of bovine rhodopsin in a trigonal crystal form," *Journal of Molecular Biology*, vol. 343, no. 5, pp. 1409–1438, 2004.

[34] K. Palczewski, "G protein-coupled receptor rhodopsin," *Annual Review of Biochemistry*, vol. 75, pp. 743–767, 2006.

[35] R. Kazmin, A. Rose, M. Szczepek, M. Elgeti, E. Ritter, R. Piechnick, K. Hofmann, P. Scheerer, P. Hildebrand, and F. Bart, "The activation pathway of human rhodopsin in comparison to bovine rhodopsin," *Journal of Biological Chemistry*, vol. 290, no. 33, pp. 20117–20127, 2015.

[36] R. Stenkamp, D. Teller, and K. Palczewski, "Crystal structure of rhodopsin: a G-protein-coupled receptor," *ChemBioChem*, vol. 3, no. 10, pp. 963–967, 2002.

[37] T. Okada, M. Sugihara, A. Bondar, M. Elstner, P. Entel, and V. Buss, "The retinal conformation and its environment in rhodopsin in light of a new 2.2 a crystal structure," *Journal of Molecular Biology*, vol. 342, no. 2, pp. 571–583, 2004.

[38] C. Hunte, "Specific protein-lipid interactions in membrane proteins," *Biochemical Society Transactions*, vol. 33, no. 5, pp. 938–942, 2005.

[39] J. Israelachvili, J. Mitchell, and B. Ninham, "Theory of self-assembly of lipid bilayers and vesicles," *Biochimica et Biophysica Acta*, vol. 470, pp. 185–201, 1977.

[40] P. Quinn, "Lipid-lipid interactions in bilayer membranes: married couples and casual liaisons," *Progress in Lipid Research*, vol. 51, no. 3, pp. 179–198, 2012.

[41] F. Contreras, A. Ernst, F. Wieland, and B. Brügger, "Specificity of intramembrane protein-lipid interactions," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 6, pp. 1–19, 2011.

[42] D. Gater, O. Saurel, I. Iordanov, W. Liu, V. Cherezov, and A. Milon, "Two classes of cholesterol binding sites for the β2ar revealed by thermostability and NMR," *Biophysical Journal*, vol. 107, no. 10, pp. 2305–2312, 2014.

[43] G. Khelashvili, A. Grossfield, S. Feller, M. Pitman, and H. Weinstein, "Structural and dynamic effects of cholesterol at preferred sites of interaction with rhodopsin identified from microsecond length molecular dynamics simulations," *Proteins*, vol. 76, no. 2, pp. 403–417, 2009.

[44] G. van Meer, "Dynamic transbilayer lipid asymmetry," *Cold Spring Harbor Perspectives in Biology*, vol. 3, no. 5, pp. 1–11, 2011.

[45] P. Devaux and R. Morris, "Transmembrane asymmetry and lateral domains in biological membranes," *Traffic*, vol. 5, no. 4, pp. 241–246, 2004.

[46] J. Singer and G. Nicolson, "The fluid mosaic model of the structure of cell membranes," *Science*, vol. 175, no. 4023, p. 720, 1972.

[47] P. Niemelä, S. Ollila, M. Hyvönen, M. Karttunen, and I. Vattulainen, "Assessing the nature of lipid raft membranes," *PLoS Computational Biology*, vol. 3, no. 2, pp. 304–312, 2007.

[48] K. Simons and R. Ehehalt, "Cholesterol, lipid rafts, and disease," *Journal of Clinical Investigation*, vol. 110, no. 5, pp. 597–603, 2002.

[49] J. Berg, J. Tymoczko, and L. Stryer, *Biochemistry*. W. H. Freeman, 5 ed., 2002.

[50] M. Holz, S. Heil, and A. Sacco, "Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate ^{1}H NMR PFG measurements," *Physical Chemistry Chemical Physics*, vol. 2, no. 20, pp. 4740–4742, 2000.

[51] K. Simons and E. Ikonen, "Functional rafts in cell membranes," *Nature*, vol. 387, no. 6633, pp. 569–572, 1997.

[52] L. Pike, "Lipid rafts: heterogeneity on the high seas," *Biochemical Journal*, vol. 378, no. Pt 2, pp. 281–292, 2004.

[53] E. London, "How principles of domain formation in model membranes may explain ambiguities concerning lipid raft formation in cells," *Biochimica et Biophysica Acta*, vol. 1746, no. 3, pp. 203–220, 2005.

[54] Ü. Coskun, M. Grzybek, D. Drechsel, and K. Simons, "Regulation of human EGF receptor by lipids," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, no. 22, pp. 9044–9048, 2011.

[55] R. Brown, "Sphingolipid organization in biomembranes: what physical studies of model membranes reveal," *Journal of Cell Science*, vol. 111, no. 1, pp. 1–9, 1998.

[56] C. Hofsäß, E. Lindahl, and O. Edholm, "Molecular dynamics simulations of phospholipid bilayers with cholesterol," *Biophysical Journal*, vol. 84, no. 4, pp. 2192–2206, 2003.

[57] P. Niemelä, M. Hyvönen, and I. Vattulainen, "Structure and dynamics of sphingomyelin bilayer: insight gained through systematic comparison to phosphatidylcholine," *Biophysical Journal*, vol. 87, no. 5, pp. 2976–2989, 2004.

[58] T. Schlick, ed., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer Science+Business Media, 2006.

[59] H. Berendsen, *Simulating the Physical World*. Cambridge University Press, 2007.

[60] J. Perilla, B. Goh, K. Cassidy, B. Liu, R. Bernardi, T. Rudack, H. Yu, Z. Wu, and K. Schulten, "Molecular dynamics simulations of large macromolecular complexes," *Current Opinion in Structural Biology*, vol. 31, pp. 64–74, 2015.

[61] M. Allen, *Introduction to Molecular Dynamics Simulation*, vol. 23. NIC Series, 2004.

[62] M. Abraham, D. van der Spoel, E. Lindahl, B. Hess, and the GROMACS development team, *GROMACS user manual version 5.0.4*. www.gromacs.org, 2014.

[63] W. Jorgensen, D. Maxwell, and J. Tirado-Rives, "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids," *Journal of the American Chemical Society*, vol. 118, no. 45, pp. 11225–11236, 1996.

[64] A. van Duin, S. Dasgupta, F. Lorant, and W. Goddard, "ReaxFF: a reactive force field for hydrocarbon," *Journal of Physical Chemistry*, vol. 105, no. 41, pp. 9396–9409, 2001.

[65] C. Baker, "Polarizable force fields for molecular dynamics simulations of biomolecules," *WIREs Computational Molecular Science*, vol. 5, no. 2, pp. 241–254, 2015.

[66] S. Marrink, J. Risselada, S. Yefimov, P. Tieleman, and A. de Vries, "The MARTINI force fied: coarse grained model for biomolecular simulations," *The Journal of Physical Chemistry*, vol. 111, no. 27, pp. 7812–7824, 2007.

[67] X. Periole, M. Cavalli, S.-J. Marrink, and M. Ceruso, "Combining an elastic network with a coarse-grained molecular force field: structure, dynamics, and intermolecular recognition," *Journal of Chemical Theory and Computation*, vol. 5, no. 9, pp. 2531–2543, 2009.

[68] A. Leach, *Molecular Modelling Principles and Applications*. Pearson Education, 2 ed., 2001.

[69] T. Kühne, "Ab-initio molecular dynamics," *WIREs Computational Molecular Science*, vol. 4, no. 391, pp. 1–13, 2014.

[70] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, vol. 1. Academic Press, 2 ed., 2002.

[71] H. Berendsen, J. Postma, W. van Gunsteren, A. DiNola, and J. Haak, "Molecular dynamics with coupling to an external bath," *Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.

[72] M. Manna, W. Kulig, M. Javanainen, J. Tynkkynen, U. Hensen, D. Müller, T. Rog, and I. Vattulainen, "How to minimize artifacts in atomistic simulations of membrane proteins, whose crystal structure is heavily engineered: β_2-adrenergic receptor in the spotlight," *Journal of Chemical Theory and Computation*, vol. 11, no. 7, pp. 3432–3445, 2015.

[73] M. Hanson, V. Cherezov, M. Griffith, C. Roth, V. Jaakola, E. Chien, J. Velasquez, P. Kuhn, and R. Stevens, "A specific cholesterol binding site is established by the 2.8 a structure of the human beta2-adrenergic receptor," *Structure*, vol. 16, no. 6, pp. 897–905, 2008.

[74] H. Ingólfsson, M. Melo, F. van Eerden, C. Arnarez, C. Lopez, T. Wassenaar, X. Periole, A. de Vries, P. Tieleman, and S. Marrink, "Lipid organization of the plasma membrane," *Journal of the American Chemical Society*, vol. 136, pp. 14554–14559, 2014.

[75] "Martini coarse grain force field for biomolecular simulations." Available (accessed on 22.10.2015): http://www.cgmartini.nl.

[76] C. Kandt, W. Ash, and P. Tieleman, "Setting up and running molecular dynamics simulations of membrane proteins," *Methods*, vol. 41, no. 4, pp. 475–488, 2007.

[77] Å. Skjevik, B. Madej, C. Dickson, K. Teigen, R. Walker, and I. Gould, "All-atom lipid bilayer self-assembly with the AMBER and CHARMM lipid force fields," *Chemical Communications*, vol. 51, no. 21, pp. 4402–4405, 2015.

[78] A. de Vries, A. Mark, and S. Marrink, "The binary mixing behavior of phospholipids in a bilayer: a molecular dynamics study," *Journal of Physical Chemistry B*, vol. 108, no. 7, pp. 2454–2463, 2004.

[79] S. Marrink, E. Lindahl, O. Edholm, and A. Mark, "Simulation of the spontaneous aggregation of phospholipids into bilayers," *Journal of the American Chemical Society*, vol. 123, no. 35, pp. 8638–8639, 2001.

[80] S. Esteban-Martín and J. Salgado, "Self-assembling of peptide/membrane complexes by atomistic molecular dynamics simulations," *Biophysical Journal*, vol. 92, no. 3, pp. 903–912, 2007.

[81] K. Scott, P. Bond, A. Ivetac, A. Chetwynd, S. Khalid, and M. Sansom, "Coarse-grained MD simulations of membrane protein-bilayer self-assembly," *Structure*, vol. 16, no. 4, pp. 621–630, 2008.

[82] S. Yesylevskyy, L. Schäfer, D. Sengupta, and S. Marrink, "Polarizable water model for the coarse-grained MARTINI force field," *PLoS Computational Biology*, vol. 6, no. 6, pp. 1–17, 2010.

[83] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. Mark, and H. Berendsen, "GROMACS: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[84] B. Hess, H. Bekker, H. Berendsen, and J. Fraaije, "LINCS: a linear constraint solver for molecular simulations," *Journal of Computational Chemistry*, vol. 18, no. 12, pp. 1463–1472, 1997.

[85] H. Martinez-Siera and T. Róg, *Biomolecular Simulations, Methods and Protocols*, vol. 924. Springer Science+Business Media, 2013.

[86] V. Gapsys, B. Groot, and R. Briones, "Computational analysis of local membrane properties," *Journal of Computer-Aided Molecular Design*, vol. 27, no. 10, pp. 845–858, 2013.

[87] W. Humphrey, A. Dalke, and K. Schulten, "VMD: visual molecular dynamics," *Journal of Molecular Graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[88] R. Koynova and B. Tenchov, "Transitions between lamellar and non-lamellar phases in membrane lipids and their physiological roles," *OA Biochemistry*, vol. 1, no. 1, pp. 1–9, 2013.

[89] J. Silvius, "Role of cholesterol in lipid raft formation: lessons from lipid model systems," *Biochimica et Biophysica Acta*, vol. 1610, no. 2, pp. 174–183, 2003.

[90] S. Wassall and W. Stillwell, "Polyunsaturated fatty acid-cholesterol interactions: domain formation in membranes," *Biochimica et Biophysica Acta*, vol. 1788, no. 1, pp. 24–32, 2009.

[91] M. Filizola, ed., *G Protein-Coupled Receptors - Modeling and Simulation*. Springer Science+Business Media, 2014.

[92] S. Feller, K. Gawrisch, and T. Woolf, "Rhodopsin exhibits a preference for solvation by polyunsaturated docosohexaenoic acid," *Journal of the American Chemical Society*, vol. 125, no. 15, pp. 4434–3345, 2003.

[93] R. Brewster and S. Safran, "Line active hybrid lipids determine domain size in phase separation of saturated and unsaturated lipids," *Biophysical Journal*, vol. 98, no. 6, pp. L21–L23, 2010.

APPENDIX A. RADIAL DISTRIBUTION FUNCTIONS FOR μ-OPIOID RECEPTOR AND RHODOPSIN SYSTEMS

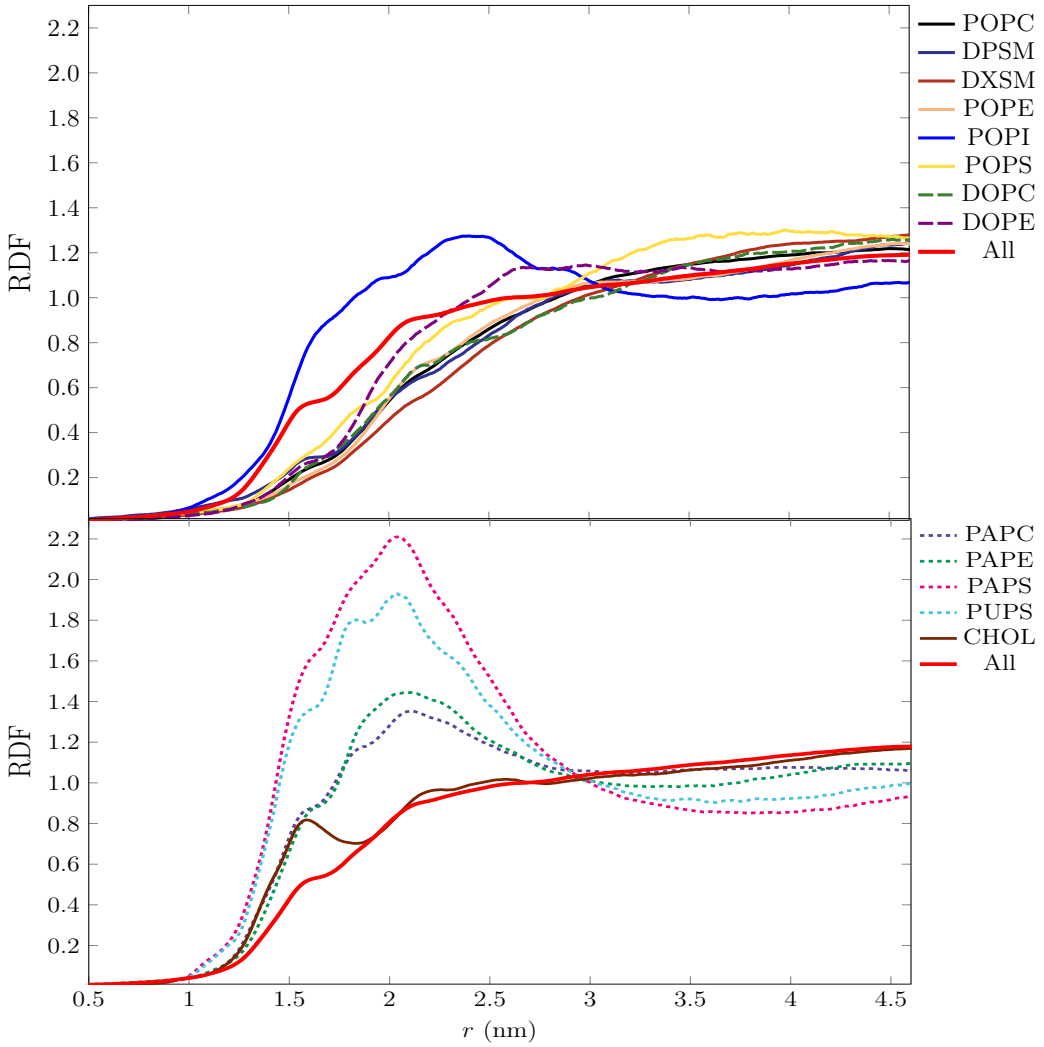


Figure A1 *Average radial distribution functions for different lipids around the center of mass of μ-opioid receptor.*

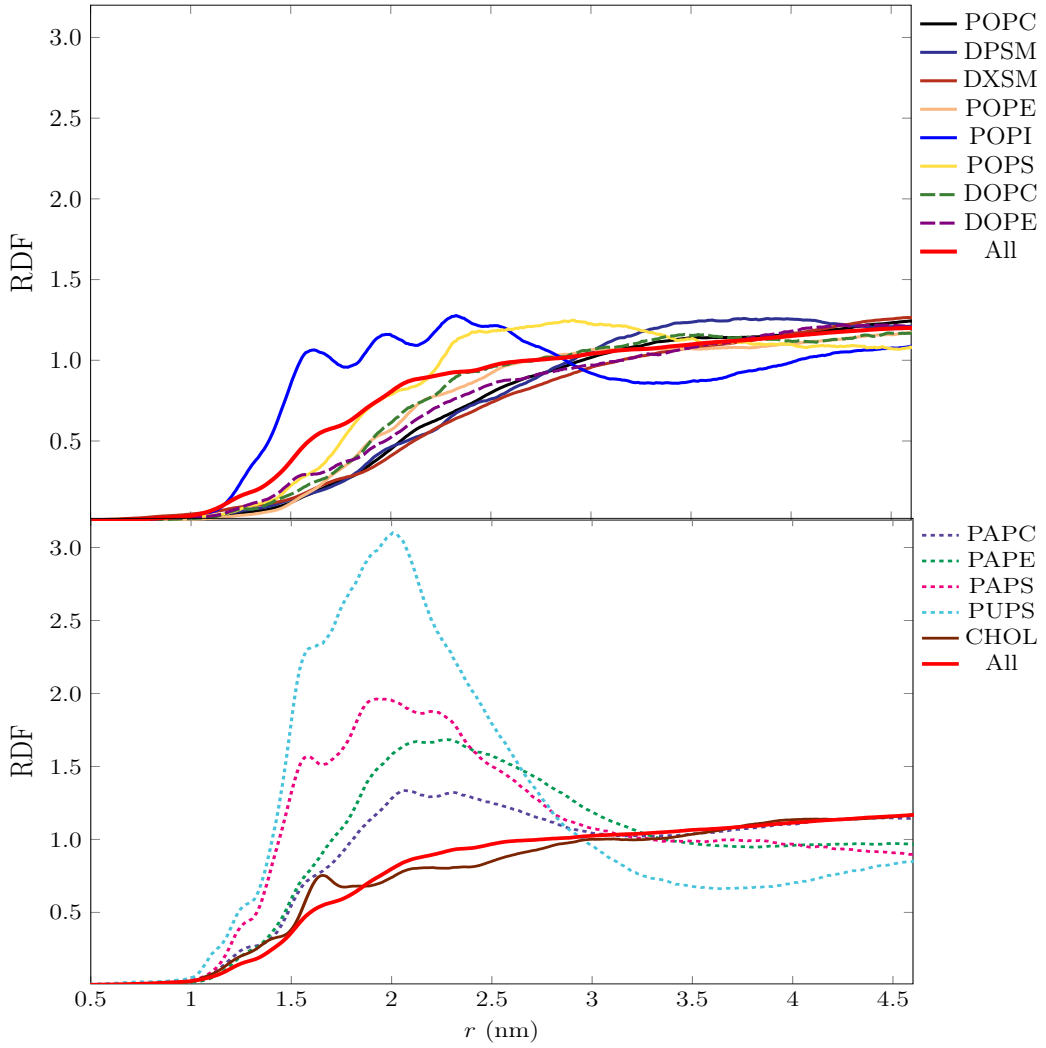


Figure A2 *Average radial distribution functions for different lipids around the center of mass of rhodopsin.*