TAMPERE UNIVERSITY OF TECHNOLOGY

TUOMO NIEMINEN

A STUDY OF RHODOPSIN AS A POTENTIAL PHOSPHOLIPID SCRAMBLASE

Master of Science Thesis

Examiner: Ilpo Vattulainen
Examiner and topic approved in the
Faculty of Science and Environmental
Engineering council meeting on
5 September 2012

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Teknis-luonnontieteellinen koulutusohjelma
NIEMINEN, TUOMO: Tutkimus rodopsiinista potentiaalisena scramblaasiproteiinina
Diplomityö, 73 sivua
Joulukuu 2012
Pääaine: Teknillinen fysiikka
Tarkastaja: Professori Ilpo Vattulainen
Ohjaaja: DI Matti Javanainen
Avainsanat: Apoptoosi, biologinen fysiikka, flip–flop, flippaasi, molekyylidynamiikkasimulaatiot, rodopsiini, scramblaasi, umbrella sampling

Solukalvon lipidien siirtyminen kalvon yhdeltä puolelta toiselle (flip–flop) on soluelämän kannalta elintärkeä prosessi. Tämän hitaan prosessin kulkua nopeuttavat tietyt integraaliset kalvoproteiinit, joita kutsutaan flippaaseiksi ja scramblaaseiksi. Nämä kalvoproteiinit nähtävästi alentavat solukalvon keskellä olevaa vapaaenergiakynnystä sellaiselle tasolle, että fosfolipidien flip–flop-ilmiö tapahtuu huomattavasti nopeammin joko ulkoisten energialähteiden kuten ATP:n avulla, tai vaihtoehtoisesti pelkästään kalvon lämpövärähtelyn vaikutuksesta. Flippaasien olemassaolosta on tiedetty jo pitkään, mutta niiden nopeuttaman flip–flop-ilmiön tapahtumaketjusta ei ole vielä selkeyttä eikä kyseisiä proteiineja ole vielä kyetty yksilöimään. Scramblaasien olemassaolosta on käyty kiihkeätä keskustelua, sillä ainoastaan niihin liittyviä geenisekvenssejä on pystytty tunnistamaan.

Scramblaasiproteiinien toimintamekanismien laajempi ymmärtäminen voisi olla hyödyllistä monien solutasolla alkavien sairauksien ymmärtämisessä. Esimerkiksi ohjattu solukuolema ilmeisesti lähtee käyntiin tietyn scramblaasiproteiinin aktivoituessa, jolloin fosfatidyyliseriini-lipidit siirtyvät nopeasti solun pinnalle. Tämä toimii signaalina solun ulkopuolisille fagosyyteille merkitsemällä solun valmiiksi apoptoosille. Apoptoottiset solut yhdistetään usein syöpäkasvaimien etenemiseen. On ehdotettu, että mutaatiot scramblaasigeeneissä voivat edistää syövän leviämistä. Oikeanlainen scramblaasien aktivoituminen näissä tilanteissa auttaisi häiriintyneiden syöpäsolujen tuhoamisessa.

Viimeaikaiset kokeelliset tutkimukset ovat osoittaneet, että opsiini, yksi silmän verkkokalvon valoreseptorisoluissa sijaitseva G-proteiinikytkentäinen reseptori, olisi eräs flippaasiproteiini. Samat tutkimukset ovat vahvasti näyttäneet, että tunnetuin opsiiniproteiini, rodopsiini, olisi mahdollisesti eräs scramblaasiproteiini. Tämän tutkimuksen tarkoituksena on perehtyä rodopsiinin scramblaasiominaisuuksiin laskennallisten menetelmien kautta käyttäen molekyylidynamiikkasimulaatioita. Saatuja tutkimustuloksia voidaan käyttää täydentämään kokeellisia tuloksia sekä antamaan suuntaa jatkotutkimuksille. Fosfolipidien mahdollisia translokaatioreittejä tutkitaan ohjattujen molekyylidynamiikkasimulaatioiden avulla ja näiden reittien vapaaenergiaerot lasketaan ns. umbrella sampling –menetelmän avulla. Saadut tulokset osoittavat, että rodopsiini alentaa kalvon vapaaenergiakynnystä merkittävästi toimien täten flippaasina. Lipidien havaittu translokaatioreitti sijaitsee rodopsiinin pinnan läheisyydessä.

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Science and Engineering
NIEMINEN, TUOMO: A study of rhodopsin as a potential phospholipid scramblase
Master of Science Thesis, 73 pages
December 2012
Major: Technical Physics
Examiner: Professor Ilpo Vattulainen
Supervisor: M.Sc. Matti Javanainen
Keywords: Apoptosis, biological physics, flip–flop, flippase, molecular dynamics simulations, rhodopsin, scramblase, umbrella sampling

Transbilayer phospholipid transfer (flip–flop) is a vital process for all cellular life. This intrinsically slow process is facilitated by integral membrane proteins called flippases and scramblases, which lower the free energy barrier to such a level that phospholipid flip–flop can rapidly occur either with help of external energy sources such as ATP or by thermal fluctuations alone. The existence of flippases has been known for some time, but the identities of the phospholipid translocators are still unknown. The existence of scramblases is less clear, since only related gene sequences have been identified so far. Also, the mechanisms by which phospholipids carry out flip–flops are yet to be identified.

Further knowledge on scramblase activities could be helpful in understanding and preventing many diseases. For instance, programmed cell death begins when phosphatidylserine flips in an uncontrollable manner from the cytosolic leaflet to the extracellular side and this process is assumed to take place due to certain membrane proteins. This event in turn creates a signal for extracellular phagocytes by marking the cell for apoptosis. Apoptotic cells are often linked to tumor progression in which it is suggested that mutations in the scramblase genes can facilitate cancer growth. Proper scramblase activation during cancer progression could be used to remove harmful cells.

Recent experimental studies have suggested that opsin, one of the G protein-coupled receptors found in photoreceptor cells of the retina, is a phospholipid flippase and its counterpart rhodopsin could be a possible scramblase. The purpose of this thesis is to use atomistic molecular dynamics simulations to examine the phospholipid scrambling properties of rhodopsin. The results can be used as a basis for further research regarding the subject. The possible flipping routes are investigated using steered MD simulations, and the free energy differences along the translocation paths are calculated by the umbrella sampling method. The results show that rhodopsin significantly lowers the free energy barrier of the membrane, thus functioning as a flippase. The observed flip–flop route taken by the phospholipids is located near the surface of rhodopsin.

PREFACE

This Master of Science Thesis was carried out in the Biological Physics group of the Department of Physics of Tampere University of Technology between June 2012 and November 2012. The molecular dynamics simulations were conducted using the computing services of CSC — IT Center for Science.

I would like to thank my examiner Prof. Ilpo Vattulainen for the chance to work with the Biological Physics group and for this interesting thesis subject. This thesis could not have been done without his knowledge on the subject and the resources of the group. This project has really piqued my interest in cellular level biology and encouraged me to continue researching this field as a post-graduate student.

Special thanks go to my supervisor Matti Javanainen for his advice during the various stages of the project and for new ideas regarding the subject. I would also like to thank my colleagues Hanna, Heikki, Joni and Sami in the Bat Cave for all the help and discussions during the past summer and autumn. Thanks should also be directed to Tomasz Róg and Reinis Danne for their assistance with the force fields. I would also like to express my gratitude to my friends and everyone at Hiukkanen for the great student years I had.

Last, but not least, I am grateful of the support of my family and Liisa during my studies and the writing process of this thesis.

Tampere, November 2012

Tuomo Nieminen

CONTENTS

SYMBOLS AND ABBREVIATIONS

A	Prefactor in Arrhenius equation
$A\left(\xi\right)$	Potential of mean force
α	Weight parameter in Ewald summation
$\mathbf{b}$	Box vector matrix
β	Inverse temperature
β_{ij}	Isothermal compressibility
C	Concentration
C_n	Ryckaert–Belleman torsion parameters
E	Potential energy
ΔE	Electric potential difference
$\mathrm{erfc}\left(x\right)$	The complimentary error function
ϵ_{ij}	Depth of the potential well in Lennard-Jones interaction
ϵ_r	Relative permittivity of the medium
ϵ_0	Permittivity of the vacuum
$\mathbf{F}_i$	Force acting on particle i
F_j	Free energy constants in umbrella sampling
$\mathcal{F}\left(Q\right)$	Fourier transform of the charge grid
ΔF	Helmholtz free energy difference between two configurations
f	Binding energy in secondary structure analysis
ΔG	Change in Gibbs free energy
g_i	Statistical inefficiency
θ_{ijk}	Angle between two atoms with the atom j in the center
θ^0_{ijk}	Reference angle in angle bending
i	The imaginary unit
K	Kinetic energy of the system
K_0	Target kinetic energy
K_i	Force constant for system restraining in umbrella sampling
k	Rate of flip–flop
k_B	Boltzmann constant
k^r_{ij}	Force constant in bond stretching
k^θ_{ijk}	Force constant in angle bending
k_ϕ	Force constant in proper dihedrals
k_ξ	Force constant in improper dihedrals
L	Largest box matrix element
$\mathbf{m}$	Reciprocal space vector

m_i	Mass of particle i
N	Number of degrees of freedom
$\mathbf{n}$	The box vector
n	Periodicity
$\mathbf{P}$	Pressure tensor
$\mathbf{P}_{\text{ref}}$	Reference pressure
$P(\xi)\,\mathrm{d}\xi$	Probability distribution of the system
$P_i^b(\xi)$	Biased distribution of the system
$P_u^b(\xi)$	Unbiased distribution of the system
$P^u(\xi)$	Global unbiased distribution
$p_i(\xi)$	Weights of unbiased distributions
ϕ_a	Reference angle in proper dihedrals
ϕ_{ijkl}	Angle between planes ijk and jkl
q_i	Charge of particle i
R	Molar gas constant
$\mathbf{r}_i$	Position of particle i
$\mathbf{r}_i^{\text{ref}}$	Reference position for particle i
r_{ij}	Distance between particles i and j
$\mathbf{r}_{ij,\mathbf{n}}$	The real distance between two charges in electrostatics
r_{ij}^0	Reference distance in bond stretching
σ_{ij}	Van der Waals distance between two particles
σ_{PMF}	Standard deviation of bootstrapped potentials of mean force
T	Absolute temperature
T_0	Temperature of the heat bath
t	Time
Δt	Time-step
τ_i	Integrated autocorrelation time of umbrella window i
τ_p	Time constant in pressure coupling
τ_T	Time constant of exponential decay for temperature coupling
V	Volume of the simulation box
$V(\mathbf{r}_1, ..., \mathbf{r}_N)$	Potential energy function
$V_a(\theta_{ijk})$	Potential function of angle bending
$V_b(r_{ij})$	Potential function of bond stretching
V_{bonded}	Bonded potential function in OPLS
$V_{\text{c}}(r_{ij})$	Coulomb potential
$V_d(\phi_{ijkl})$	Potential function of proper dihedrals
V_{dir}	Direct space sum in Ewald summation
$V_{id}(\xi_{ijkl})$	Potential function of improper dihedrals
$V_{\text{LJ}}(r_{ij})$	Lennard-Jones potential

$V_{\text{non-bonded}}$	Non-bonded potential function in OPLS
V_{rec}	Reciprocal sum in Ewald summation
V_{total}	Total potential function in OPLS
V_0	Constant term in Ewald summation
$\mathbf{v}_i$	Velocity of particle i
$\mathbf{W}$	Matrix parameter of pressure coupling
W	Work
$W_b(\xi)$	Bootstrapped potential of mean force
$\mathrm{d}W$	Wiener noise term
$w_i(\xi)$	Bias potentials in umbrella sampling
ξ	Reaction coordinate in umbrella sampling
ξ_i^C	Predetermined position of an umbrella window
ξ_{ijkl}	Angle between the planes ijk and jkl in improper dihedrals
ξ_0	Reference angle in improper dihedrals
Z	Faraday constant

AA	All-atom representation
ABC	ATP-binding cassette
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
DPPC	Dipalmitoylphosphatidylcholine
ER	Endoplasmic reticulum
GPCR	G protein-coupled receptor
LINCS	Linear Constraint Solver algorithm
LJ	Lennard-Jones interaction
MD	Molecular dynamics
OPLS	Optimized Potentials for Liquid Simulations
PC	Phosphatidylcholine
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PLSCR	Phospholipid scramblase
PM	Plasma membrane
PMF	Potential of mean force
PS	Phosphatidylserine
POPC	1-palmitoyl-2-oleoylphosphatidylcholine
RB	Ryckaert–Belleman's function
R.E.D.	RESP ESP charge Derive
RMS	Root mean square
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
RNA	Ribonucleic acid
SM	Sphingomyelin
SSM	Stearoylsphingomyelin
UA	United-atom representation
VMD	Visual Molecular Dynamics
WHAM	Weighted histogram analysis method

1. INTRODUCTION

Phospholipid translocation from one membrane leaflet to another (flip–flop) is a vital process in sustaining cellular life [1]. Even though the phospholipid flip–flop takes place very slowly in protein-free lipid membranes with the time scale usually being seconds to minutes, it yet has to occur at a relatively high rate to maintain such cellular activities as membrane growth, signalling, endocytosis, exocytosis and programmed cell death [1, 2]. This suggests that passive phospholipid flip–flop in protein-free environments is complemented by other processes that speed up the flip–flop process.

The polar headgroups of phospholipids prefer to stay away from the hydrophobic interior of the membrane, thus having to overcome a free energy barrier in order to begin the flip–flop process [1]. This barrier, however, is high enough not to be overcome by thermal fluctuations alone, resulting in a low intrinsic flip–flop rate [2,3]. In order to lower the free energy barrier for entering the membrane interior, specialized membrane proteins are embedded in a bilayer [4]. These integral membrane proteins facilitate phospholipid flip–flop by reducing the height of the free energy barrier to a level where it can be easily overcome with the help of external energy sources such as ATP, or even by thermal fluctuations alone [1,3]. A common factor between these proteins is a transmembrane domain, which interacts with the phospholipid undergoing flip–flop [1]. The term 'flippase' is mainly used to refer to unidirectional transbilayer lipid translocator proteins that can be either ATP-dependent or ATP-independent [1, 5]. It is also used to refer to a bidirectional transbilayer lipid translocator in the endoplasmic reticulum [5]. The term 'scramblase' is reserved for bidirectional ATP-independent transbilayer lipid translocator proteins activated by an overload of cytosolic calcium [1, 5].

Even though the existence of flippases has been known for a while, not much is known of the flipping process or even the proteins in question. Even the existence of scramblase proteins has not clearly been proven yet. The flippase proteins have so far only been identified as members of certain well-known protein families, and of scramblases, only a few genes are known so far [6]. Of the mechanics of the flipping process, there have been various predictions [2, 7, 8], but no consensus has so far been reached on the preferred flipping route.

Recently, the first molecular identification of an ATP-independent phospholipid

flippase was carried out as an integral membrane protein from the opsin protein group was identified as one [3]. The results have given reason to believe that another protein from that group, rhodopsin, could be a phospholipid scramblase [3, 9].

In this thesis, molecular dynamics simulations were conducted to study the possible scramblase activities of rhodopsin. The use of atomistic molecular dynamics allows us to create an atomic-scale model of a biological membrane with the desired phospholipids, proteins and other molecules embedded in it. This model can then be studied in a desired ensemble at much shorter time scales and better resolution than would be possible experimentally. With molecular dynamics such statistical averages of the flip–flop process as the change in free energy along the flipping path can be calculated with relative ease. Various flipping paths can also be compared by using steered molecular dynamics to force the phospholipid to follow the desired route across the membrane. The data obtained from the trajectories allows us to determine if the routes are even remotely realistic.

However, one should not forget the importance of experimental studies for comparison. Both experimental and computational methods have their weaknesses, but they can be used to complement each other quite effectively. The results obtained in this thesis are compared with previous experimental data, and many similarities are found.

This thesis is divided into seven chapters with the main purpose of examining phospholipid flip–flop facilitated by monomeric rhodopsin, and studying the possible flipping mechanisms and the free energy differences along the flipping paths. The flipping mechanisms are studied with the help of steered molecular dynamics simulations, in which a phospholipid is pulled along a certain path while examining the pulling force applied to its headgroup. Once a favored flipping route is found, the free energy difference is calculated by umbrella sampling, which uses biased molecular dynamics to evaluate the potential of mean force along a reaction path.

The second chapter of this thesis focuses on the biological background behind the functions of cell membranes and proteins. The phospholipids of our model membranes are introduced together with a discussion their properties. The properties of cell membranes are discussed along with the cellular components that are relevant to this thesis. The basics of protein structures are also discussed, and a short review into membrane proteins is given.

The third chapter concentrates on the lipid flip–flop event by discussing the various methods through which lipid flip–flops can occur, and by reviewing the current knowledge on flippase and scramblase proteins. Rhodopsin, the protein under study, is also introduced and its molecular structure is discussed thoroughly. The chapter closes with discussion of apoptosis, one of the most important cellular activities arising from lipid flip–flop.

The fourth chapter concentrates on the basics of molecular dynamics simulations. The main focus of the chapter is on the simulation engine GROMACS and the OPLS force fields that are utilised in this work. The chapter begins with a short overview of the simulation process and computational methods, moving then into the theory behind the implemented force fields, temperature and pressure coupling methods and bond constraints. The basic theory behind umbrella sampling and its analysis methods are also discussed.

In the fifth chapter, the stages undertaken in our molecular dynamics simulations are discussed. The model membranes and related simulation parameters are discussed together with the ideas behind the steered molecular dynamics simulations.

The sixth chapter focuses on the results obtained from the simulations. The possible rhodopsin-mediated flipping routes are discussed with the obtained free energy curves supported by calculations and experimental data. A free energy curve is also obtained for a protein-free bilayer with a similar lipid composition, to provide a reference for rhodopsin-mediated flip–flop. The stability of the system is analyzed by examining fluctuations in the protein structure.

Finally, the seventh chapter summarizes the thesis with a short review of the results and discussion regarding future directions that could be taken with regards to combining molecular dynamics simulations with experimental research of transbilayer lipid motion.

2. BIOLOGICAL BACKGROUND

Cells are the basic structural units of all living organisms. Eukaryotic cells found in plants and animals are much more complex than their prokaryotic counterparts found in bacteria. They are complex structures surrounded by a plasma membrane (PM), which separates them from the surrounding environment providing protection from outside dangers. Plasma membranes are formed of two layers of lipids — usually comprised of fatty acid chains linked by a headgroup — and proteins embedded in the membrane. These proteins allow certain substances to pass through the otherwise inpregnable bilayer. Many types of organelles can be found inside the cell, each with its specific function and most surrounded by a phospholipid bilayer. Most notable of these bilayers is the endoplasmic reticulum (ER) that encloses a cisternal space called the lumen. Other cellular organelles include the nucleus, which contains the genetic information of the cell; the mitochondria, the main sources of ATP; the Golgi device, which has important functions in the synthesis of new proteins and lipids; and ribosomes, that use the RNA from the nucleus to synthesize new proteins.

In this chapter the biological background of cell membranes and the molecules within are discussed. Basic knowledge on the types and functions of phospholipids and cell membranes is necessary in order to understand the aims of this thesis. In addition, knowledge of protein structure and the functions of different types of membrane proteins is helpful in understanding further discussion on flippase proteins, which are the main subject of this thesis.

First, a short overview of the structures and the types of phospholipids of our model membranes are reviewed. Next, the physical structure and biological functions of the cell membrane are discussed together with a short review of certain important cellular components. Basics of protein structures are discussed to help understand how different proteins are formed. Finally, a short review of membrane proteins and their functions is given as a prelude for the next chapter.

2.1 Phospholipids in Cell Membranes

Lipids are important molecules in all living systems due to their ability to form borders around prokaryotic or eukaryotic cells. Their chemical structures may vary considerably even though they have common features such as insolubility in water and solubility in nonpolar organic solvents [10]. They consist of fatty acid chains

(derivatives of hydrocarbon chains), which can either be unsaturated or saturated depending on whether they contain double bonds or not. These fatty acid chains can attach themselves to other molecules through esterification or ether bonds in which a covalent bond between two atoms is formed and water is removed [11]. Glycerol with its three hydroxyl groups is the most common molecule these hydrocarbon chains attach themselves to [12].

Phospholipids are lipids that consist of a polar hydrophilic headgroup, a phosphate group, a glycerol backbone and two hydrocarbon tails. The type of the phospholipid is defined by the types of the headgroup and the hydrocarbon tails. Phospholipids are amphipathic molecules meaning that their polar headgroups can form bonds with water molecules while the hydrophobic tails try to avoid water. Therefore phospholipids have a tendency to naturally form, for instance, micelles or lipid bilayers in aqueous solutions, in which the hydrophilic heads prefer to stay in contact with water while the hydrophobic tails stay away from it [13].

Phospholipids are mainly synthesized in the endoplasmic reticulum (ER) of the cell. The newly synthesized phospholipids must be transported from the biogenic membrane to various organellar membranes such as the Golgi device or the plasma membrane [14]. Their intermembrane transport through the aqueous part of the cell is made possible by vesicles that enclose the hydrophobic tails [7]. The parts of the ER membrane that bud off as vesicles fuse with the target membrane once they have reached their destination [7]. The area of the target membrane and the volume of the enclosed organelle is subsequently increased [7]. The cytosolic side of the transported part of the ER membrane remains on the cytosolic side of the target membrane [14].

Three types of membrane lipids were used in our model membranes: POPC, SSM and cholesterol. POPC, or 1-palmitoyl-2-oleoylphosphatidylcholine, consists of one palmitic acid chain and one oleic acid chain joined together by a glycerol backbone, which in turn is connected to a positively charged phosphatidylcholine (PC) headgroup [1]. The palmitic tail is saturated and the oleic tail unsaturated due to a *cis*-double bond in it, creating a kink in the structure of the tail. Phosphatidylcholines are major components of biological membranes found in every cell of the human body [1]. The chemical structure of POPC is presented in Fig. 2.1.

Sphingolipids are derivatives of sphingosine, an amino alcohol with a long hydrocarbon tail, containing an additional fatty acid chain attached to the sphingosine amino group [11]. Sphingomyelin (SM) is a phospholipid which consists of two hydrocarbon tails bound to a polar phosphocholine headgroup containing serine [14]. If the second fatty acid chain is stearic acid, then the newly formed phospholipid is called stearoylsphingomyelin (SSM). Sphingomyelins are constructed in the Golgi device from a ceramide backbone that is synthesized in the ER [14]. A high con-

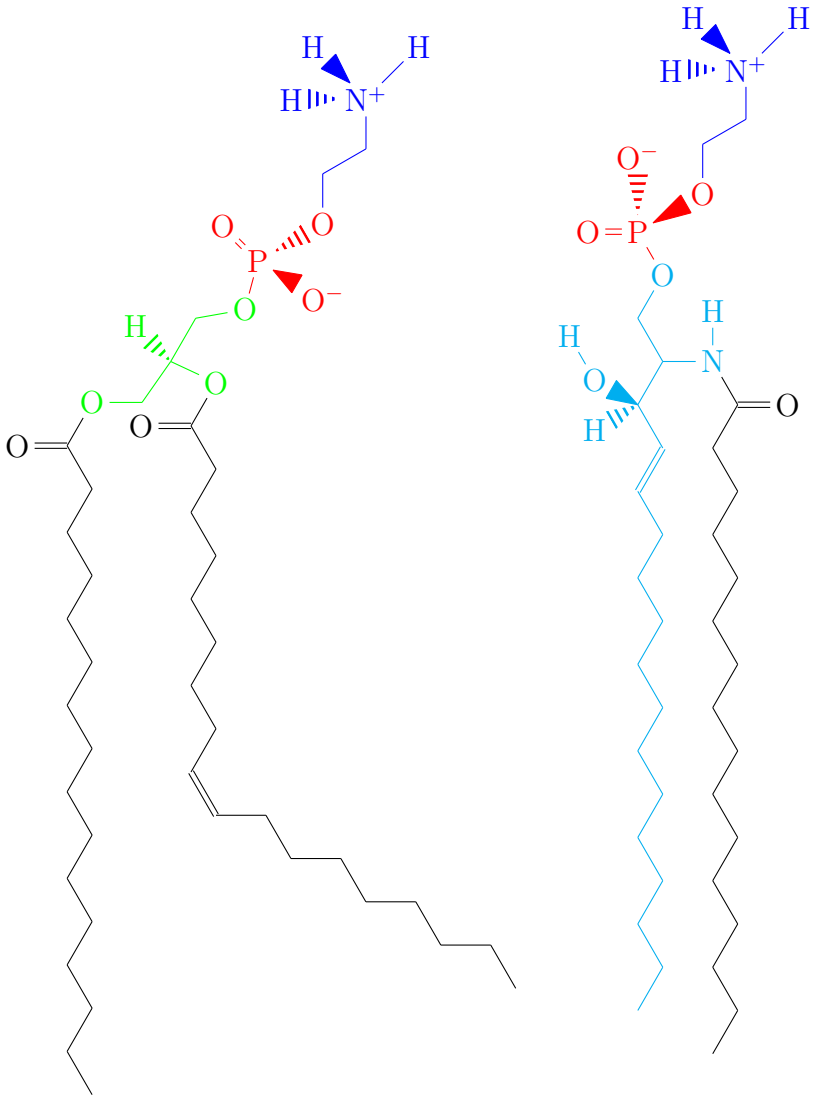

Figure 2.1: The chemical structures of POPC and SSM. Blue color represents the choline headgroup, red the phosphate group, green the glycerol backbone of POPC, cyan the sphingosine part of SSM and black the hydrocarbon tails.

centration of sphingomyelin in a membrane is thought to have a stabilizing effect on the membrane structure due to the extensive hydrogen bonding caused by the free hydroxyl group in sphingosine [11, 15]. When located on the surface of a plasma membrane, sphingolipids act as recognition sites for nearby cells, which is useful for various purposes including programmed cell death [12, 16]. The chemical structure of SSM is presented in Fig. 2.1.

Cholesterol is an amphipathic molecule with a characteric structure consisting of a polar hydroxyl group and a rigid nonpolar sterole group of four carbon rings [10]. It is an important membrane lipid being abundant in the plasma membranes of eukaryotic cells [1]. A stable bilayer structure can not be formed with only cholesterol, though, since the sterol group makes it too hydrophobic and it can thus

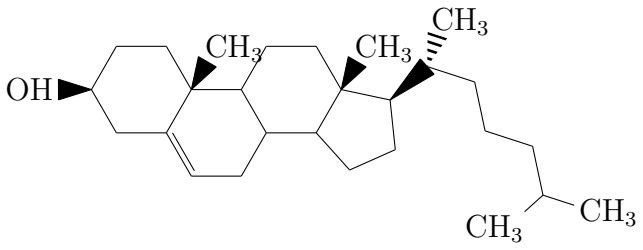

Figure 2.2: Cholesterol consists of a polar headgroup connected to a rigid steroid ring structure on the opposite side of which is a short nonpolar hydrocarbon tail.

only be found in bilayers mixed with phospholipids [11]. Cholesterol has a stiffening effect on the bilayer structure as it increases the ordering of both the saturated and monounsaturated chains of phospholipid tails [17]. The stiffening increases steeply until the bilayer contains about 30 mol-% cholesterol in its composition after which the bilayer can be considered to be in a gel-like state [17]. Cholesterol and sphingomyelin are usually found together in plasma membranes and their effects on each other have been studied for instance in Ref. [18]. The chemical structure of cholesterol is presented in Fig. 2.2.

2.2 Properties of Cell Membranes

Cell membranes such as plasma membranes or endoplasmic reticula are formed out of two leaflets of phospholipids with proteins embedded in them. They are surrounded by water on both sides and therefore the phospholipids are naturally arranged in such a way that their hydrophobic tails face each other in the interior of the membrane while their hydrophilic heads are located on the outer sides of the leaflets, exposed to water [19]. This sort of behavior results in a bilayer structure. Typical lengths of the hydrocarbon tails of phospholipids are usually between 10 and 20 carbon atoms meaning that most lipid bilayers are 3 to 6 nanometres thick [19]. The lipid bilayer can be either symmetric or asymmetric in lipid composition depending on the lipid distribution, which is affected by certain membrane proteins.

The hydrocarbon tails of phospholipids spontaneously aggregate to an energetically favored structure forming a hydrophobic core at the interior of the bilayer through which no hydrophilic solutes can easily diffuse. This impermeable barrier therefore prevents most spontaneous movement of molecules through the bilayer. The proteins embedded in the membrane, however, lower this barrier allowing necessary transportation of molecules between the lumen and the cytosol, or the cytosol and the extracellular region [12]. As both intracellular and intercellular movement of molecules and ions is vital to cell life, these transporter proteins are of utmost importance.

The membranes of animal cells contain five major types of phospholipids, which

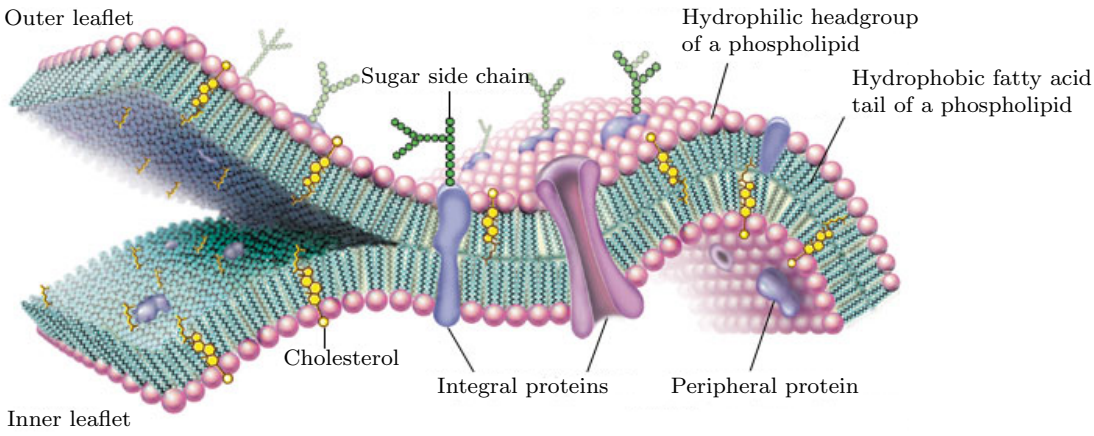

Figure 2.3: A schematic drawing of a cell membrane. On both sides of the membrane is a large amount of water, which has been omitted from the picture. Figure modified from Ref. [22].

are phosphatidylcholine (PC), phosphatidyethanolamine (PE), phosphatidylserine (PS), sphingomyelin (SM) and cholesterol [14]. In plasma membranes PC and SM can mainly be found in the outer leaflet while PE and PS are usually located in the cytoplasmic leaflet resulting in asymmetric distribution of phospholipids in the plasma membrane [14, 20, 21]. Cholesterol can be found in both leaflets of plasma membranes [14]. The plasma membranes of human erythrocytes have a molar composition of about 21 % PC, 29 % PE and PS, 21 % SM and 26 % cholesterol [11].

The endoplasmic reticulum is a eukaryotic cell organelle enclosing a cisternal space called the lumen. It forms a network of interconnected internal membranes throughout the cytosol [1]. The endoplasmic reticula are either smooth or rough, the latter of which contains ribosomes essential for the synthesis of specific plasma membrane proteins such as opsin [1, 13]. These proteins are then translocated by vesicles through the Golgi device into the plasma membrane [13].

The lumen of the ER occupies over 10 % of total cell volume and has a central role in the biosynthesis of both lipids and proteins [1]. It also has a special function in storing and releasing calcium ions that are used in cell signalling [13]. The release of calcium is controlled by certain membrane proteins that act as Ca^{2+}-pumps between the cytosol and the lumen [1].

The main structural differences between the PM and the ER is in their distribution of lipids. Whereas lipid asymmetry is vital for the PM, the ER prefers a symmetric distribution of phospholipids [1]. The two layers of the plasma membrane are asymmetric due to the fact that new lipids are synthesized in the cytosolic leaflet of the endoplasmic reticulum from which they are transported to the inner leaflet of the PM and not all types of phospholipids end up on the outer leaflet [1]. On the other hand, the two layers of the endoplasmic reticulum are symmetric,

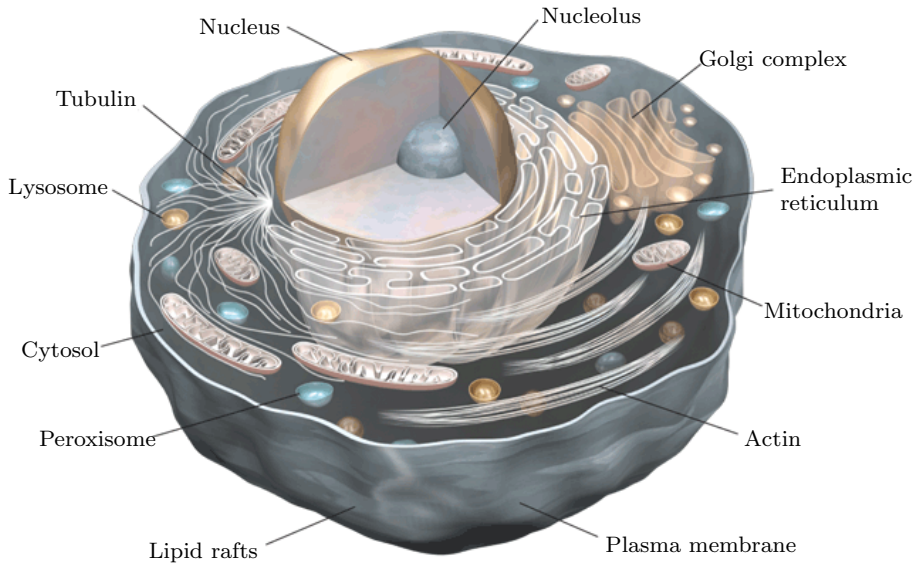

Figure 2.4: A schematic drawing of the interior of the cell. Figure modified from Ref. [23].

presumably due to a scramblase protein, which is a phospholipid translocator that induces rapid transbilayer movement of all phospholipids in the ER and therefore half of the newly synthesized phospholipids rapidly flip to the lumenal leaflet [1]. The phospholipid equilibration results in a symmetric growth of both halves of the bilayer, because the translocation mechanism is not lipid specific [1].

The asymmetric distribution of phospholipids in the PM results for instance in membrane curvature, mechanical stability of the membrane, regulation of biogenesis and function of membrane proteins making it vital to cellular activities [4,6]. The loss of lipid asymmetry in plasma membranes can result in aggregation of thrombocytes and it is also associated with apoptotic cells [6, 24].

Cell membranes are not stiff structures. Lipids can rotate freely about their axis and move laterally via diffusion so that their acyl chains remain in the membrane interior. The amount of free area and its distribution in a membrane has a strong effect on the rate of lateral diffusion and the permeability of the membrane [25]. A typical phospholipid can diffuse laterally at as high a velocity as several micrometers per second at normal body temperature [11]. The lateral diffusion of lipids gives the bilayer a fluid-like composition, which also causes lateral movement of membrane proteins. Lateral diffusion of phospholipids is also vital to all reactions occurring in the cell membrane [11].

The bilayer phases can be characterized into an ordered crystalline gel-like phase, a liquid-ordered phase or a disordered liquid phase and the types of phospholipids in a plasma membrane strongly affect its properties [26]. Kinks created by *cis*-bonds in the fatty acid chains cause the lipid composition to be much looser and disordered whereas long saturated chains can pack together much more tightly, regularly

forming a much stiffer structure [10]. The fluidity of the bilayer is thus greater in a disordered structure. Cholesterol interacts with the hydrophobic tails of the phospholipids, immobilizing them and hardening the bilayer structure by restricting the random lateral diffusion of phospholipid molecules and increasing their order [11,25]. The permeability properties of the bilayer are also enhanced by cholesterol due to its immobilizing effect [1]. Studies have also demonstrated that bilayers with sphingomyelin in them are thicker and more gel-like due to their ordered structure [11]. The phase behavior can be related to many membrane events such as protein sorting, signalling and membrane fusion [26]. Change from one phase to another happens at a characteristic main transition temperature, which is lower in membranes containing lipids with short saturated tails [1].

In liquid state, lipid bilayers can have separate tightly packed and highly ordered domains [1]. Bilayers composed of PC, SM and cholesterol might have strong enough van der Waals forces between the saturated hydrocarbon tails to hold two adjacent molecules together for a while, possibly transiently assembling them into so called lipid rafts with the help of membrane proteins [1]. The group of lipids arranged into a raft can then diffuse freely in the given monolayer. These lipid rafts are thought to have membrane organizing properties that aid in the bioactivities of the membrane, but there is also much controversy surrounding their existence and biological functions [27].

2.3 Protein Structure

Proteins are naturally occurring polymers of amino acids, where each amino acid residue is joined to the next one by a covalent peptide bond. With 20 different types of amino acid residues, and hundreds of residues in a protein, the possible amino acid combinations in a protein are nearly limitless. The types of the amino acid residues and their sequence along with the conformation of the protein contribute to all protein functions including signalling, molecule transport, catalyzing chemical reactions and affecting the structure of the cytoplasm [11].

The three dimensional structure of a protein depends on the number and sequence of the amino acid residues. The long polypeptide chain folds into a conformation which minimizes its free energy. The native conformation is strongly affected by the different hydrophobic properties of the amino acid residues, which are believed to dominate the forces that drive protein folding [19]. The polar side chains of the amino acids tend to gather to the outside of the protein to be able to interact with water, while the nonpolar side chains form a hydrophobic core inside the protein [1].

The folded proteins are thermodynamically stable and their unfolding by, for instance, high differences in temperature will cause many of their properties to disappear [19]. Other factors, such as interactions with surrounding environment, can

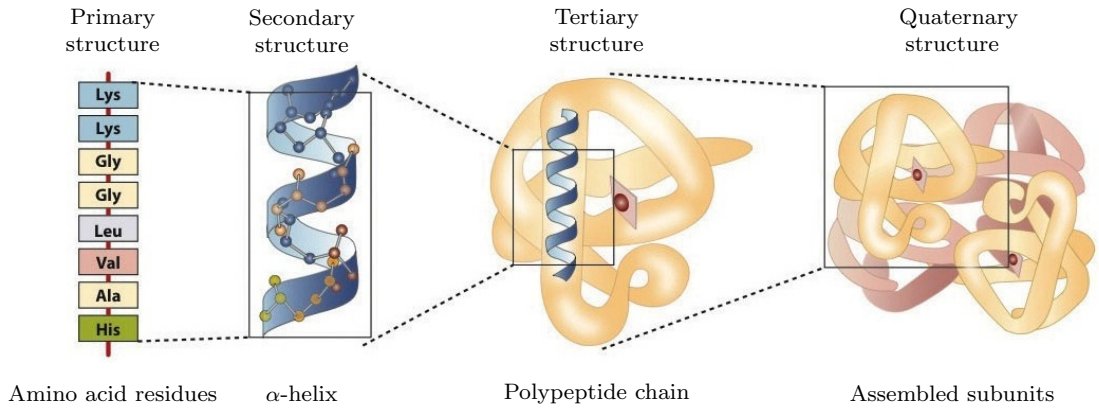

Figure 2.5: Illustration of the four levels of protein structures. Figure modified from Ref. [28].

also affect the structure of the protein by either making it more stable or causing the native conformation to unfold. For instance, different kinds of lipids in membranes have different effects on the stability of membrane protein alpha-helices [19].

The protein structure can be described by four levels: primary, secondary, tertiary and quarternary structures. The sequence of amino acids in the protein builds up the primary structure. Two proteins are said to be homologous to each other if they are significantly similar to each other by their primary structures [29]. The secondary structure consists of the local patterns in the structure such as alpha-helices, beta-sheets, turns, loops and coils. The secondary structures form different few amino acid residue long motifs depending on the amino acid sequence, some of which have specific biological functions. One such motif is the so called EF-hand motif consisting of two alpha-helices E and F connected together by a loop, which has a calcium binding property [29]. The tertiary structure of the protein is defined by the way different motifs are arranged into domains, which are basic units of tertiary structures. A critical determinant of the tertiary structure is for instance the localization of hydrophilic amino acids on the protein surface and of hydrophobic amino acids in the protein interior [14]. Homology in the primary structure results in similar tertiary structures in the homologous domains of the protein [29]. The quarternary structure of the protein consists of several polypeptide chains that associate three-dimensionally in a specific way [29]. It includes the interactions between the different polypeptide chains in proteins that are composed of more than one polypeptide [14].

The alpha-helix is a right-handed polypeptide chain that coils around itself in a spiral manner to form a helical structure [10]. Each turn of the alpha-helix is 3.6 residues or about 5.4 Å long, i.e. each carbonyl (C = O) bond in residue i is connected to a NH-group in a following residue $i + 4$ by a hydrogen bond [10, 29].

The alignment of these hydrogen bonds as well as the dipole moments caused by the CO- and NH-groups is along the helical axis [29]. The sidechain groups are directed outward from the helix. In certain cases the alpha-helices can loosen into π-helices or tighten into 3_{10}-helices [29, 30]. The looser π-helices are right-handed and have 4.1 residues per each turn and similar hydrogen bonds between its residues i and $i + 5$ [30]. The tighter 3_{10}-helices are also right-handed and have only 3 residues in each turn and hydrogen bonds between residues i and $i + 3$ [30]. All membrane crossing helices are in a hydrophobic environment and therefore most of the side chains in them are also hydrophobic [29].

The beta-sheets consist of two peptide chains lying side by side in a pleated-sheet arrangement forced by the planarity of peptide bond forces, in which the sidechain groups alternate above and below the plane of the sheet [10]. They are the second most common structure in proteins after alpha-helices. The peptide chains are held together by interchain hydrogen bonds and form a repeating pattern every 7 Å [10]. Beta-sheets are either parallel or antiparallel depending on whether the beta strands run in the same biochemical direction or in opposite directions [29].

Turns are a three or four residue long sites where the overall directions of the polypeptide chains are reversed, easily distinguishable at the ends of helices [11, 31]. They allow large proteins to form very compact structures and can also be classified by the amount of residues it takes for a turn to occur [11]. If a turn is more than a few residues long, it is called a loop [11].

2.4 Membrane Proteins and Molecule Trafficking

Cell membranes would be impregnable for most molecules without membrane proteins [13]. They can be divided into two groups: integral and peripheral membrane proteins. Integral membrane proteins are amphipathic and tightly attached into the membrane as a result of hydrophobic interactions between their transmembrane alpha-helices and bilayer lipids, and possibly also by a lipid anchor [13]. In order to detach them from the membrane, one has to use either detergents or nonpolar organic solvents which break the hydrophobic interactions between the protein and the membrane [13, 32]. Peripheral membrane proteins attach themselves only temporarily either to the lipid bilayer by lipid anchors or via electrostatic interactions and hydrogen bonds, or to integral membrane proteins. In the former case they are also known as lipid-anchored proteins [32]. The locations of membrane proteins on the plasma membrane are not static due to the fluidity of the bilayer, which forces them to move laterally along with the lipids [32].

Membrane proteins have an important role in the functionality of the cell. They affect the permittivity of the membrane, regulate the cell functions and most importantly they act as channels for ions and solutes to travel through the lipid bilay-

ers [32]. The part of the integral membrane protein which is embedded inside the membrane is called the transmembrane domain [14]. It typically consists of 20–25 nonpolar or hydrophobic amino acid residues coiled into an alpha-helix [14]. Usually these alpha-helices pass through the membrane several times opening up a channel for polar molecules to pass through the lipid bilayers [32].

Not all membrane proteins have transmembrane alpha-helices though, and the ones that do are called transporters [32]. The transporter proteins are either active or passive depending on whether they require an outside energy source or not, respectively [13, 32]. The most common external energy source in cells is adenosine triphosphate (ATP) [13]. Upon releasing energy, ATP loses a phosphate atom and turns into adenosine diphosphate (ADP) [13]. The amount of energy released by the transition from ATP to ADP for a single ATP molecule is about 20 $k_{\mathrm{B}}T$, which in the normal body temperature of 310 K is about 51.5 kJ/mol [13].

Passive trans-membrane transportation is caused by diffusion, which is a spontaneous process occurring from a region of a higher concentration towards a region of a lower concentration. In the case of lipid bilayers there is a free energy barrier between the two monolayers, meaning that for the transmembrane diffusion to begin, a certain activation energy is required [13]. Small nonpolar molecules can cross cell membranes with relative ease, but larger molecules and polar molecules encounter a high free energy barrier [14]. This barrier exists, because the polar molecule has bonded with surrounding water molecules, which are difficult to transport through the hydrophobic bilayer [13]. The passive transporters lower this activation energy by creating a transmembrane channel through the bilayer. Weak interactions between the transporter protein and the transported molecule replace the bonds between the molecule and the water molecules making the repulsive forces caused by the nonpolar core of the bilayer much smaller [13]. The transporter protein quickens the intercellular movement of molecules by several orders of magnitude [13].

The direction of passive transportation can be deduced from the concentration gradient. In active transportation, moving molecules against their concentration gradient requires an external energy source such as ATP [13]. The amount of energy required for active transportation can be crudely estimated from the equation

$$\Delta G_t = RT \ln \frac{C_2}{C_1} + ZF\Delta E, \tag{2.1}$$

where in the first term $R = 8.3145 \frac{\mathrm{J}}{\mathrm{mol} \times \mathrm{K}}$ is the molar gas constant, T the absolute temperature and C_2/C_1 the ratio of concentration between the two solutions, when $C_2 > C_1$. The second term accounts for charges if the solute is an electrolyte. In the second term Z is the charge of the solute, $F = 96.480 \frac{\mathrm{kJ}}{\mathrm{V} \times \mathrm{mol}}$ the Faraday constant and ΔE the electric potential difference between the two compartments, which for

a eukaryotic cell membrane is typically about 70 mV [13, 32].

If the energy for the active transportation is obtained from proteins such as ATPases, the transporter proteins are primary active transporters [13]. If the energy comes only from an electrochemical gradient, the transporter proteins are secondary active transporters [13].

The majority of primary active transporters are P-type ATPases, integral proteins that transport cations using energy gained from the hydrolysis of ATP [13]. P-type ATPases phosphorylate themselves during the cation pumping process by absorbing a phosphate ion from ATP, which turns into ADP in the process [1, 13]. The best understood P-type transporter ATPase is the P-type Ca^{2+} ATPase found in the sarcoplasmic reticulum membrane of skeletal muscle cells, which actively pumps Ca^{2+}-ions out of the cell in order to maintain a steep Ca^{2+} gradient across the plasma membrane [1]. Other P-type cation pumps function in the same way, maintaining an electrochemical ion gradient across the plasma membrane, the directions of which depend on the ion [1].

Other primary active transporters include the ATP-binding-cassette (ABC) transporters, constituting the largest family of membrane transporter proteins. Each member of the ABC transporter protein family contains two highly conserved ATP-binding domains, which undergo a conformational change as a result of ATP binding. This conformational change opens a channel inside the protein for ions to pass through [12]. Dissociation of ADP as a result of ATP hydrolysis returns the conformation to its original state [12]. Most ABC transporters are unidirectional and in eukaryotic cells their main function is to export small molecules out of the cytosol [1].

The secondary active transporters obtain their energy mainly from ion gradients of Na^+ or H^+ generated by primary transporters, making them dependent on the energy obtained from ATP-dependent mechanisms [12]. They typically function by coupling the spontaneous flow of ions down the concentration gradient to the simultaneous pumping of molecules by a secondary active transporter against it, thus transforming electrochemical energy into vectorial movement [12, 13, 33].

The forms of secondary active transport can be allocated into three categories. Uniporters bind to one molecule at a time, moving them down the concentration gradient along with the ions sustaining it [13]. Symporters are involved in moving two types of molecules in the same direction, typically against the concentration gradient, whereas antiporters move two types of molecules in opposite directions exchanging one molecule to another [13].

The transporting capabilities of membrane proteins are not limited only to ions or small molecules, but they extend to larger molecules such as phospholipids as well. The functions of these active and passive lipid transporters are discussed further in the next chapter.

3. TRANSBILAYER LIPID MOTION

Cellular life is sustained by the rapid transbilayer motion of polar lipids [2]. Biosynthesis of new phospholipids occurs only on the cytosolic leaflet of the endoplasmic reticulum in eukaryotes and thus newly synthesized lipids must be flipped to the opposing leaflet at a high enough rate for the bilayer to be balanced [6]. The energetic cost for the translocation of the polar headgroup of the lipid is very high, therefore resulting in a slow intrinsic flipping process [4]. Hence, cell growth can not be sustained by spontaneous lipid movement alone. For this reason, cells have specialized membrane proteins embedded in the bilayer that facilitate lipid transport across the bilayer with the help of external energy sources such as ATP [6]. Such ATP-dependent lipid translocators are called flippases or floppases depending on whether they move lipids from the inner monolayer to the outer monolayer or vice versa, and the transbilayer lipid motion mediated by these membrane proteins is called flip–flop [2, 4, 21]. Meanwhile, some Ca^{2+}-activated membrane proteins are able to facilitate lipid flip–flop across the bilayer without the use of external energy sources. These ATP-independent phospholipid flippases are called scramblases [4, 21].

Transbilayer lipid motion occurs both in the plasma membrane and the endoplasmic reticulum. In the situations described in this thesis, the inner monolayer is always located on the cytosolic side of the membrane and the outer monolayer on the extracellular side of the membrane in the PM, or on the lumenal side of the membrane in the ER.

This chapter begins with a discussion of different methods for transbilayer lipid motion followed by discussion of flippase and scramblase proteins and their functions. The putative scramblase protein researched in this thesis is rhodopsin and thus its molecular structure and key functions are discussed in further details. This chapter closes with a short review of programmed cell death, an example of a process triggered by controlled flip–flop of phosphatidylserine.

3.1 Methods of Transmembrane Lipid Flip–flop

The energy required for a lipid with a PC headgroup to flip across the middle of a DPPC bilayer has been calculated using umbrella sampling by Tieleman and Marrink [34] to be about 80 kJ/mol. A later study by Bennett et al. [35] examined the effect of cholesterol components of a bilayer on phospholipid flip–flop and found the free

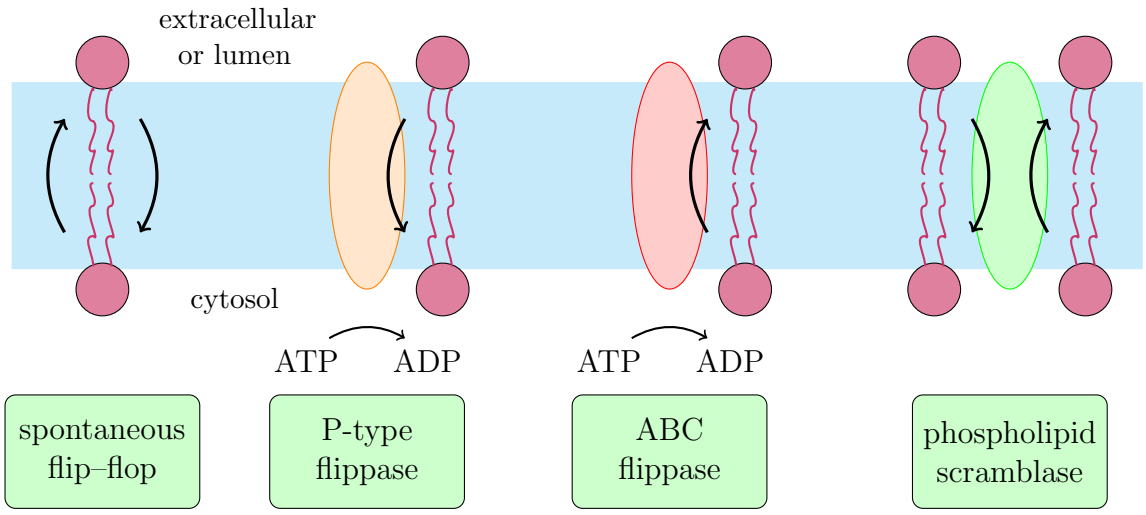

Figure 3.1: Four different mechanisms for transbilayer lipid motion. Spontaneous flip–flop is bidirectional and requires no external energy. P-type flippases facilitate inward movement of phospholipids with the external energy obtained from ATP. Likewise, ABC flippases facilitate outward movement of phospholipids by the use of ATP. Scramblase proteins are bidirectional and Ca^{2+}-dependent, but require no external energy.

energy barrier to increase to about 110 kJ/mol at around 40 mol-% cholesterol concentration. The typical time constants for flip–flop in both studies varied from hours to weeks. Due to the high barrier imposed by the bilayer and the subsequent low intrinsic flip–flop rate, the phospholipid must find a way for its polar head group to have an easier access through the membrane interior.

Due to the amphipathic nature of phospholipids, it is clear that at least part of the flip–flop process must be similar for all phospholipids. The hydrophobic acyl chains are likely to stay in the interior of the membrane while the polar headgroup follows a path with a lower free energy barrier. Two possible solutions have been suggested for mediating lipid transportation across the hydrophobic interior of the membrane [8]. These methods include a pore-mediated process, and a flipping process facilitated by specific membrane proteins. These processes have been demonstrated in Fig. 3.2.

Passive lipid translocation across the membrane has been studied extensively. It can occur intrinsically, or by a pore-mediated process in which a water pore forms in the membrane, through which the hydrophilic headgroup of a phospholipid can easily flip across by a diffusive process [36, 37]. The average durations for pore-mediated lipid flipping have been found to be between 10 and 130 ns for several different types of bilayers, and the flip–flop rate has been found to increase with increasing temperature [36, 37]. It has also been suggested that the translocation of one lipid would promote the translocation of a second lipid from the receiving side, resulting in a repeating process of substitution and mixing of lipids [37]. The limiting factor for lipid flip–flop in this case is the rate at which water pores are

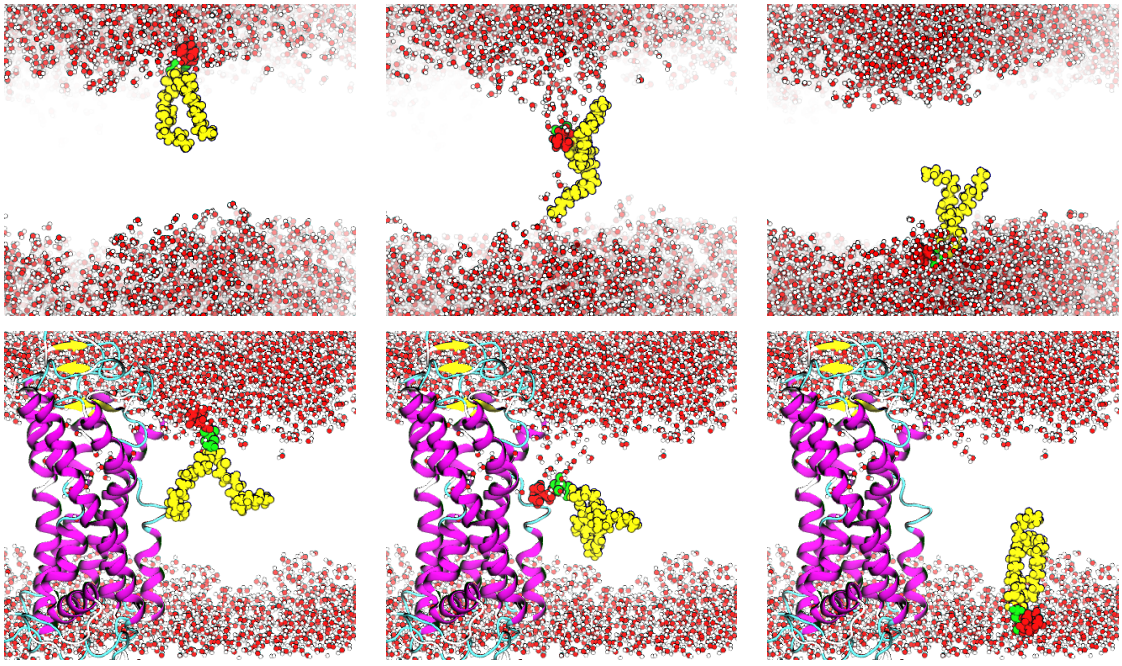

Figure 3.2: Lipid movement across the membrane can either be mediated by water pores or by a specific transporter protein. The head group of the phospholipid moves either along a water pore or the surface of the membrane protein while the acyl chains desorb from the membrane leaflet and travel to the opposing leaflet. The two series of figures show parts of the translocation process for one lipid. The lipid bilayer is omitted from the pictures for the sake of clarity.

spontaneously formed in the interior of the membrane, which is an unlikely event due to its high cost of free energy [36].

The rate of passive transportation is not high enough to sustain cell growth by itself and therefore a second, faster method must exist for lipid flip–flop. By studying bilayers with and without membrane proteins embedded in them, the rate of lipid flip–flop has been found to be orders of magnitude faster in membranes that include proteins [20]. Therefore it has been suggested that phospholipid flip–flop is also protein-mediated. Since then, the related membrane proteins have been narrowed down to a few categories and several suggestions for their relevance in lipid translocation process have been made [2, 6]. A possible method of translocation is shown alongside the pore-mediated process in Fig. 3.2, in which the phospholipid regards the membrane protein in a similar way as the water pore in passive translocation. A possibility exists that water pores form rapidly inside the membrane protein, providing the polar headgroup with easy access through the bilayer. A so called “slip-pop”-model was suggested by Kol et al. [8] in which the transient disturbances in the bilayer caused by motion of membrane proteins would make the phospholipid slip into the interior of the bilayer and pop out on the other side. In the vicinity of a transmembrane peptide, the lipid–lipid interactions, are replaced by

lipid–peptide interactions and increased thermal motion in the bilayer might cause nearby phospholipids to enter the transmembrane channel and flip across the membrane [7]. It has even been suggested that during its translocation the phospholipid could be completely encapsulated by the transmembrane protein, although intuition would consider that to be unlikely [2].

The transmembrane alpha-helices of primary and secondary active transporters form a channel for ions to pass through the membrane bilayer by lowering the free energy barrier that the translocating molecule has to overcome. Thus it is intuitive to say that the same is true for the phospholipid headgroup. The transmembrane channel could thus be a favorable route for the polar headgroup to take.

As the large number of proposed lipid translocation methods suggest, the exact method with which proteins flip lipids is still unknown. It is, however, clear that transmembrane lipid movement is catalyzed by membrane proteins promoting flip–flop. It has been hypothesized that transmembrane peptides or integral membrane proteins with transmembrane alpha-helices induce transbilayer lipid motion on biogenic membranes merely by their presence, and this has been proven e.g. by Kol et al. [38] *in vitro* in *E. coli*. However, phospholipid flipping has been shown to be specific only to certain proteins by a more recent study [3].

Another extensively studied subject is the relation between the bilayer composition and the flip–flop rate. A high content of PE or cholesterol has a reducing effect on the rate of phospholipid flip–flop, which is likely due to the increased order of acyl chains [8]. A higher charge on the polar headgroup seems to lower the rate of flip–flops according to the fact that the free energy barrier of the membrane is dependent on the charge [8]. This headgroup dependence, however, is not observed in biogenic membranes, because the polar side chains of transmembrane alpha-helices interact with the lipid headgroup [8]. It is possible that flippases induce the formation of lipid rafts from sphingolipids and cholesterol, or other molecular packing defects close to the protein–lipid interface, which in turn enhances the flip–flop rate [39]. The bilayer may also be destabilized by dolichols, which function as glycosyl residue carriers across the ER membrane, thus increasing the rate of their flipping [40].

3.2 Flippases – ATP-dependent Lipid Translocators

The existence of transmembrane lipid transporters was first suggested 40 years ago by Bretscher et al. to explain the asymmetry of biological membranes [20]. Since then the mechanisms of transbilayer lipid motion have been researched extensively and flippases have been recognized as members of well-known protein families [6]. The exact methods by which flippases transport phospholipids across the membrane are yet to be clarified. The inward movement of phospholipids is carried out by P_4-type ATPases and the outward movement is mediated by ABC transporter proteins,

both of which are driven by energy obtained from ATP hydrolysis [6]. The energy is used to exceed the energy requirements of lipid transport coming from the energy barrier of the membrane and the concentration gradient of ions.

The inwardly directed lipid flippases are proteins of a large P-type ATPase superfamily consisting of 10 transmembrane helices [6]. They are exclusively found in eukaryotic cells, and a total of 14 different P_4-ATPases are encoded in the human genome [6, 41]. Still their physiological and cellular functions remain widely unknown.

Purified ATP8a1, a member of the P_4-ATPase group, has been proven to be a flippase activated only by a certain stereoisomer of PS, showing that there is lipid selectivity in flippase activity [42]. ATP8a2, found in the photoreceptor outer segments of the retina, has also been identified as a flippase [6]. It was shown to flip PS specifically from the inner leaflet of liposomes to the outer leaflet [43]. Five P_4-ATPases (Drs2p, Dnf1p, Dnf2p, Dnf3p and Neo1p) that can be found in yeast seem to act in regular membrane trafficking processes and are believed to function similarily also in higher eukaryotes [6, 44]. The plant equivalent for an aminophospholipid translocase is ALA1 in the Arabidopsis protein, which is homologous with yeast Drs2p and bovine ATP8a1, transporting both PS and PE [45].

The outwardly directed lipid flippases comprise a group of proteins from the ABC superfamily [6]. About 50 ABC proteins, organized into seven subfamilies (ABCA to ABCG), have a physiologically important role in humans either facilitating outward translocation of lipids or regulating cell environments [46]. Many hereditary diseases have been linked to mutations in the ABCA genes, and dysfunctioning ABC proteins can lead to numerous other diseases [47]. The mechanisms with which ABC proteins transport substrates have been studied, but the precise methods are still unclear.

The first ABC protein to be recognized as a lipid flippase was the ABCB4 protein, also known as MDR3 in humans, which is a PC exclusive flippase in the plasma membranes of hepatocytes [6]. Studies have shown that the lack of ABCB4 or its dysfunction results in formation of cholesterol gallstones and severe liver diseases due to defects in PC translocation [48]. ABCA4, linked with the visual cycle, is another ABC protein found to show flippase activity with a conjugate of retinal-PE [47]. Mutations in ABCA4 interferes with the flipping process, causing the retinal-PE conjugate to accumulate inside the photoreceptor disks, resulting in a loss of vision [47]. Other ABC proteins showing flippase activity include several ABC drug transporters, such as Cdr1p, Cdr2p, Yor1p and Pdr5p, which do not seem to be lipid specific [6]. Purified ABCB1 has been shown to be a flippase for NBD-labelled phospholipids and sphingolipids [49]. ABCC1, also known as multidrug resistance-associated protein MRP1, has been shown to be a flippase for PC and sphingolipids [50]. Other ABC proteins mediate membrane lipid transfer from the

outer leaflet to a soluble acceptor and are therefore not considered to be true lipid translocases [6].

Lipid flippase activity has also been observed in bacteria, mainly in *E. coli*. Flip–flop in bacterial cytoplasmic membranes is rather independent of the phospholipid headgroup and does not require metabolic energy [8]. Dedicated flippase proteins that function in bacteria have not yet been identified even though flip–flop in bacteria has been shown to be sensitive to proteases [6]. The *E. coli* protein translocon was thought to facilitate flippase activity in the bacterial inner membrane due to its structure, but only a minimal contribution was observed [51]. However, several ABC transporters such as MsbA, an essential protein in *E. coli*, are prominent candidates, and purified MsbA has been shown to exhibit flippase activity for a large variety of phospholipids in an ATP-dependent manner [52].

3.2.1 The Biological Functions of Flippases

The main biological functions of flippases are related to biogenesis. When new membranes are being generated, the synthesis of new membrane lipids occurs mainly on the cytosolic leaflet of the endoplasmic reticulum, but also on the cytosolic leaflet of the Golgi device and the mitochondrion [53]. Some of these newly synthesized lipids remain on the cytosolic leaflet of the ER and half of them have to be flipped to the lumenal side in order to grow the bilayer symmetrically [1].

During cell formation, newly synthesized lipids are imported from the ER and the Golgi device to the cytosolic side of the plasma membrane via transport vesicles [53]. From there half of the lipids have to be flipped to the extracellular side of the PM to ensure stable cell growth [1, 54]. Since the intrinsic rate of lipid diffusion across the membrane is very slow (about $10^{-15}\mathrm{s}^{-1}$), flippases play an important role in the translocation of phospholipids as the rate of phospholipid synthesis and flipping has to be equal to the rate of cell growth [2, 6]. Flippases greatly increase the rate of transverse lipid diffusion making cell growth possible [6].

Flippases also play an important role in retaining both the lipid asymmetry of the plasma membrane and the lipid symmetry of the endoplasmic reticulum. The P_4-ATPases of the PM transport both PS and PE to the extracellular leaflet, keeping PC and SM confined to the cytosolic leaflet and resulting in an asymmetric distribution of phospholipids [53]. The translocation mechanism in the ER is ATP-independent and not specific to any lipid, resulting in rapid transbilayer movement of lipids across the ER membrane that preserves a equilibrated lipid structure that is considered to be symmetric [1].

In some cases, the number of lipids in the two monolayers are kept different on purpose. Inwardly directed flippases can create a mass imbalance between the leaflets, thus bending the membrane, which is a prerequisite for endocytosis and

vesicle formation and also useful for absorbing substances from the exoplasm [5, 12]. Conversely, outwardly directed flippases can cause a defect in endocytosis in the same way. It has been shown that even a 1 % difference in the ratio between lipids in the two leaflets will result in a substantial change on membrane structure [55]. Due to the limitations on protein-free flip–flop, membrane bending and vesicle budding would be difficult to accomplish without the aid of flippase proteins [5].

The constitution of the exoplasmic leaflet defines many intercellular properties of the cell. A high level of PC and cholesterol has to be sustained on the exoplasmic leaflet of the bilayer to keep the membrane surface inert and stable in addition to keeping the energy barrier functioning [5]. Flip–flop of glycosphingolipids to the outer leaflet is essential for signal transduction and intercellular recognition [5]. Exposure of platelet membrane PS, for instance, is essential in regulating blood coagulation, since PS acts as a secondary messenger linking platelet activation to the activation of prothrombin, a protein that eventually stems blood loss [56].

Important flippase functions are not limited only to outward flipping as it is equally important to deplete certain phospholipids from the exoplasmic leaflet. For instance, a high exposure of PS on the cell surface acts as a signal for macrophages that the cell in question marks for apoptosis and should be devoured [5]. Therefore, in healthy cells flippases prevent PS from accumulating to the exoplasmic leaflet of the membrane.

Absence of flippases or their dysfunction for any reason will cause phospholipids to accumulate to one leaflet of the bilayer resulting in dramatic deformations of the membrane [2]. Translocation of wrong phospholipids across the membrane can also have dire consequences for membrane structure [57].

3.3 Scramblases – ATP-independent Lipid Translocators

The third group of phospholipid flippases constitute a group of ATP-independent lipid translocator proteins found both in plasma membranes and endoplasmic reticula of eukaryotic cells [1]. The flippases found in self-synthesizing membranes such as the ER or bacterial cytoplasm are commonly referred to as biogenic membrane flippases, whereas the flippases localized to the PM are called scramblases [1].

Although many of the functioning flippase proteins have been identified, the identity and biological functions of scramblase proteins still remain unknown, despite the fact that their existence has been known for quite some time [58]. Progress has however been made in the study of the genes related to the scrambling process.

Phospholipid scramblases (PLSCRs) are a group of bidirectional phospholipid translocators that function without the use of external energy source such as ATP [6]. The scramblases found in the ER are always active resulting in constant exchange of lipids between the two leaflets of the membrane [6]. This activity is not specific to

any lipid, leading to complete randomization of the newly synthesized lipids between leaflets [6]. Meanwhile, the scramblases of the PM are only activated by increased concentration of cytosolic Ca^{2+} resulting in disruptions in the membrane asymmetry, which in turn has serious consequences for the cell [1, 6].

Four related phospholipid scramblase genes have so far been identified in humans, named as hPLSCR1 - hPLSCR4 [59]. Respectively to hPLSCR1, the predicted open reading frames encode proteins with 59 % hPLSCR2, 47 % hPLSCR3 and 46 % hPLSCR4 homology [59].

The most studied member of the family is hPLSCR1, being involved in rapid transbilayer phospholipid redistribution in the plasma membrane. The PLSCR1 protein is a 37 kDa type II endofacial membrane protein comprising a total of 318 amino acid residues with a single transmembrane segment [21]. Its essential amino acid domains include the N-terminal region in amino acids 1–85, which may bind other proteins; the DNA binding domain in amino acids 86–118, which enables the protein to interact with DNA; the cysteine-palmitoylation motif in amino acids 184–189, which provides a membrane anchorage for the protein essential for the scramblase activity; the calcium binding motif in amino acids 273–284, which supposedly binds the calcium ions necessary for the phospholipid scrambling activity; and the transmembrane helix near the C-terminus in amino acids 291–309 [21, 60, 61].

The calcium binding motif of hPLSCR1 was thought to be an EF-hand-like structure consisting of two short alpha-helices connected by a loop [62]. Binding of Ca^{2+} by that loop is supposed to induce a conformational change in the protein enabling the scramblase activity [62]. Mutations leading to replacement of any of the amino acids in the EF-hand resulted in notable reduction in the scrambling activity [62]. A more recent structural model of hPLSCR1, however, showed that calcium does not bind to the EF-hand-like structure, but to a new class of calcium binding motifs found in all scramblases [63]. In addition to Ca^{2+} ions, this motif is also known to bind Mg^{2+} ions [63]. Ionic interactions of Ca^{2+} or Mg^{2+} with this motif lead to a change of conformation around the region of the motif, but its physiological implications are still unclear [63]. It is believed, however, that this change in conformation eventually leads to the events related to programmed cell death, cell activation, or blood coagulation [63, 64].

Little is known of the other three PLSCR genes. The expression of hPLSCR2 mRNA is the most restricted, found mainly in human testis [59]. hPLSCR3 is much more common, found as expressed in a number of tissues including heart, kidney, pancreas and spleen along with hPLSCR1 and hPLSCR4 [59]. Notably hPLSCR4 was not detected in peripheral blood lymphocytes [59]. hPLSCR1 and hPLSCR3 were not detected in the brain [59]. Confined to mitochondria, hPLSCR3 has been found to show scrambling activities in mitochondria by regulating biosynthesis of

cardiolipins and subsequently maintaining the mitochondrial structure and controlling apoptosis [65]. The biological functions of hPLSCR2 and hPLSCR4 are still to be determined [60].

3.4 Rhodopsin

Although four PLSCR genes have been identified, a single protein that acts as a scramblase is yet to be named. Recent studies have indicated that opsin is a phospholipid flippase [3], and rhodopsin, which is converted from opsin by addition of a 11-*cis*-retinal, could be a possible scramblase [9]. The method by which rhodopsin flips lipids is yet unknown, but believed to be related to water molecules that find their way inside the helical bundle of rhodopsin [3]. The water molecules would cause rhodopsin to have a hydrophilic core, which in turn would allow the polar headgroups of phospholipids easy access through the bilayer, while the acyl chains remain in the interior of the membrane. This agrees with the simulation data gathered from pore-mediated lipid translocation [37], where the polar headgroup moves along the water pore and the acyl chains turn around in the membrane. The flipping routes and the possibilities that they function are further discussed in the results section.

3.4.1 The Molecular Structure

As a key protein in human vision, rhodopsin is a thoroughly researched molecule. Its molecular structure has been determined at as low as 2.2 Å resolution by crystallization by Okada et al. [66] and can be found in the Protein Data Bank (PDB) database by its identifier 1U19. Rhodopsin is a member of the largest subfamily of G protein-coupled receptors (GPCRs), heterotrimeric guanine nucleotide -binding proteins, constituting about 90 % of the whole family [66, 67].

All GPCRs have a similar structure consisting of seven transmembrane alpha-helices connected to each other by six protein loops of varying lengths [67]. Like all proteins they consist of a long chain of amino acid residues. The residues are numbered by the order in which they appear in the amino acid chain of the protein and together they form the backbone of the protein [10]. Conserved water molecules can be found in the transmembrane channels of all GPCRs including bovine rhodopsin, and they are believed to have some importance in their biological functions such as changes in conformation [68].

The protein opsin comprises a total of 348 amino acid residues. Rhodopsin differs from opsin only by a 11-*cis*-retinal covalently linked to Lys296 and it has a total weight of about 40 kDa [67]. The retinal chromophore is attached to the lysine residue via protonated Schiff-base-linkage, an N substituted imine, which is

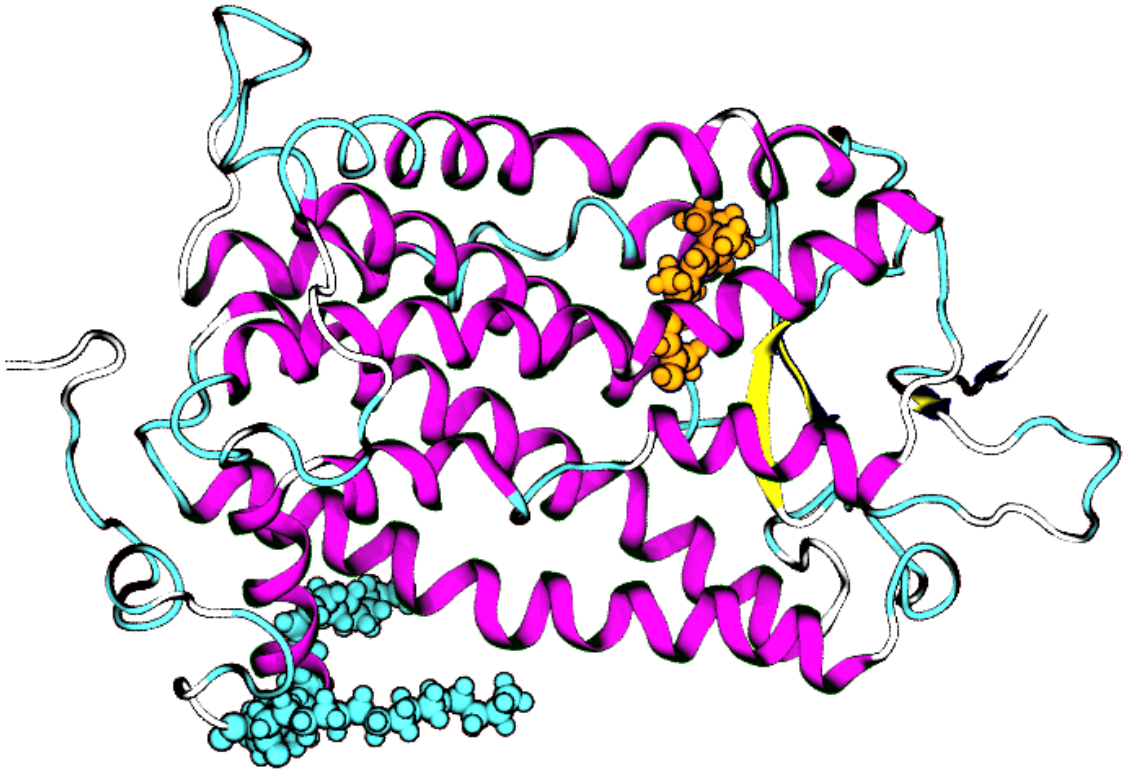

Figure 3.3: Rhodopsin as visualized in VMD (Visual Molecular Dynamics) [70] and rendered using Tachyon ray tracing library [71]. The cytosolic side of the membrane lies on the left side of the image and the extracellular side on the right side of the image. Two palmitoyl chains have been added to Cys322 and Cys323 to act as lipid anchors between the protein and the membrane and are drawn in the figure in cyan color using the VDW-scheme. The retinal molecule located inside the transmembrane channel is attached to Lys296 and is drawn in the figure in orange color using the VDW-scheme. Different colors correspond to different secondary structures. The purple ribbons represent transmembrane alpha-helices, the yellow arrows beta-sheets, and the cyan and white curves turns and coils, respectively. Parts of the alpha-helices have loosened into turns and coils over time in the molecular dynamics simulations.

a functional group containing a carbon-nitrogen double bond [67]. The protonated Schiff-base is stabilized by electrostatic interactions with negatively charged Glu113, which constrains the retinal to its inactive state [69]. The 11-*cis*-double bond of the retinal changes its conformation upon absorbing a photon turning the retinal into a *all-trans*-conformation [66].

The seven transmembrane alpha-helices (H-I to H-VII) are made up of a total of 194 amino acid residues located in the following helices: Trp35 to Gln64 in H-I, Pro71 to His100 in H-II, Pro107 to Val139 in H-III, Asn151 to Val173 in H-IV, Asn200 to Gln225 in H-V, Glu247 to Thr277 in H-VI, and Ile286 to Tyr306 in H-VII [67].

The residues in the extracellular region include 74 amino acids forming a compact structure comprised of interhelical loops and beta sheets (E-I to E-III) and an NH2-terminal tail. The amino acids are located in the following chains: Met1 to Pro34

in the NH2-terminal tail, Gly101 to Gly106 in E-I, Gly174 to Asn199 in E-II, and His278 to Pro285 in E-III [67].

The 70 amino acid residues in the cytoplasmic region are located in three loops (C-I to C-III) and a COOH-terminal comprising His65 to Thr70 in C-I, Cys140 to Glu150 in C-II, Leu226 to Ala235 and Ser240 to Ala246 in C-III, and Ile307 to Pro327 and Ser334 to Ala348 in the COOH-terminal region [67].

It is also notable that the side chains of Cys322 and Cys323, which can be found on the outside of rhodopsin, are suitable for attachment of palmitic acids, which would act as lipid anchors between rhodopsin and the bilayer. These palmitoyl chains are believed to be essential for the scramblase activity of hPLSCR1 either by promoting the correct conformation for the calcium-binding site or by accelerating the transbilayer lipid movement [21, 72]. Thus, they have been added to our rhodopsin. Detailed studies on the behavior of these palmitoylations have suggested that the palmitoyl chain in Cys323 potentially works as a lipid anchor, while the palmitoyl chain in Cys322 interacts more with the protein itself and might thus be related with the stability of the dark state of rhodopsin [73]. The palmitoylation might also act as a targeting signal for other membrane proteins and thereby control its localization [74]. For instance, palmitoylated hPLSCR1 has been shown to localize to the PM, whereas the absence of palmitoylation in hPLSCR1 directs it to the nucleus [74].

3.4.2 The Function of Rhodopsin

The light and colour sensitive area in the eye is called the retina, which comprises millions of light-sensitive rod cells and colour-sensitive cone cells. Rhodopsin can be found as an integral protein on the outer segments of each disk of these cells, and it is functional in visual phototransduction [12]. It activates upon the absorption of a photon by the covalently attached 11-*cis*-retinal, which changes its conformation to *all-trans* in the absorption process. The change in the conformation of the retinal forces additional conformational changes in the rhodopsin, which begins the visual transduction [12].

Like all integral membrane proteins, opsin is synthesized on the rough endoplasmic reticulum [1]. Its signal sequence directs it to the plasma membrane of the cell through the Golgi device, which modifies its original structure [1]. The parts of the plasma membrane containing opsin bud off internally forming stacks of disk shaped membranes enveloped by a plasma membrane into the outer segments of photosensitive rods located in human and animal retina [75].

Rhodopsin can not be synthesized in human cells by itself, i.e., the synthesis requires an external source. Such a source can be found in Vitamin A, which is produced from β-carotene obtained from vegetables such as carrots [12, 76]. Vitamin

A is crucial in forming new 11-*cis*-retinals which along with opsin proteins convert to rhodopsins, and it has been shown that vitamin A deficiency results in a lack of vision in dark environments [76, 77].

3.5 Apoptosis

There are two primary ways for a cell to die. The first is necrosis, caused by external factors that damage the cell membranes so that the cells break down and the cellular molecules are released to the space between cells [13]. The second is apoptosis, a naturally occurring programmed cell death in which nearby phagocytes (white blood cells) devour the dying cells [13]. Unlike necrosis, apoptosis is a controlled process which does not cause infections [13].

Apoptosis is a way for cells to control the time of their own death, which is useful for organisms that have unneeded or potentially dangerous cells [12]. It is characterized by cell shrinkage, blebbing of the plasma membrane and nuclear condensation and fragmentation [64]. The cells break down into small cell portions surrounded by membranes in order to prevent the cellular components from ending up in the intercellular space [13]. These vesicles are then devoured by phagocytes, and the amino acids and nucleotides in the proteins of the cell are subsequently released in the degradation process, which allows them to be reused in neighboring cells [12]. It plays a complementary role to the cell generation process, balancing the number of cells in the organism [78]. Apoptosis is triggered by an outside signal that activates the so called death receptors in the cell, eventually causing the elevation of Ca^{2+} concentration in the cytosol, which in turn activates hPLSCR1 [79].

The apoptotic processes dependent on Ca^{2+} have been linked to caspases, cysteine-dependent aspartate-specific proteases, that function as mainstream executioners of apoptosis [64]. Caspases are homologues of the cell-death protease CED-3, and their overexpression is known to result in apoptotic cell death [64]. Once caspases are activated, they begin to interfere with Ca^{2+} homeostasis by cleaving the enzymes that control the Ca^{2+} flux in and out of the cell, resulting in an overload of Ca^{2+} in the cell [64, 80]. Once enough Ca^{2+} has been accumulated, the scramblases begin to function leading to externalization of phosphatidylserine. Although phosphatidylserine exposure in apoptosis has been shown to being Ca^{2+}-dependent, it is not Ca^{2+}-regulated [64].

In the plasma membrane, phosphatidylserine distribution between the two leaflets is regulated by the bidirectional Ca^{2+}-dependent scramblase and unidirectional flippases. In normal conditions the scramblase is inactive, whereas the flippases function to prevent PS from accumulating to the extracellular leaflet. A surplus of Ca^{2+} in the cytosol activates the scrambling activity of phospholipids, while having an inhibiting effect on the aminophospholipid translocase activity [64].

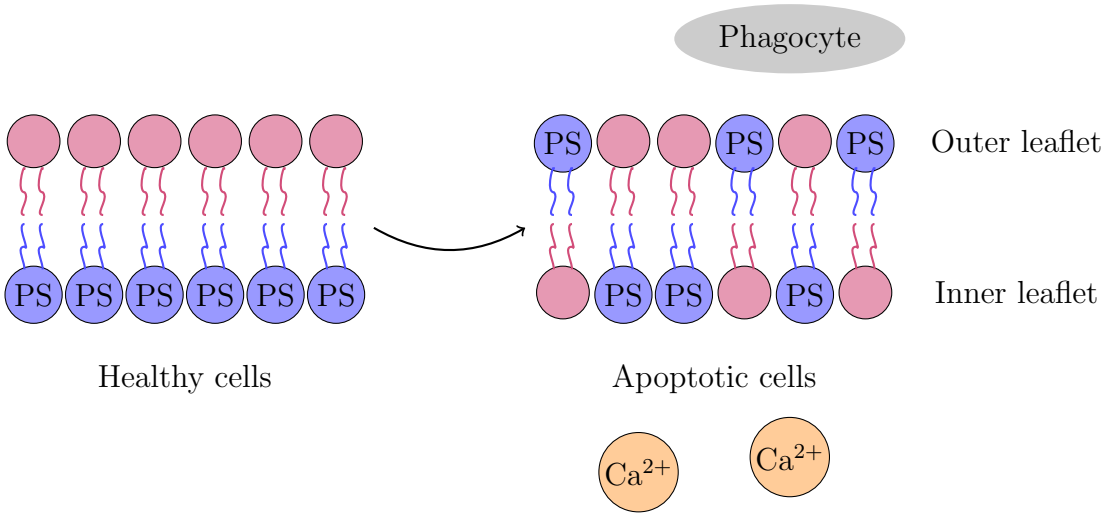

Figure 3.4: An illustration of apoptosis. On the left is a healthy cell bilayer with an inactive scramblase. An overload of calcium ions in the cytosol activates the scramblase resulting in externalization of phosphatidylserine. PS on the outer leaflet then acts as a marker for phagocytosis.

The activation of the phospholipid scramblase and the dysfunction of phospholipid flippases causes PS to rapidly flip–flop to the extracellular leaflet of the bilayer, thus interfering with the lipid asymmetry in the bilayer [21, 79]. The appearance of PS on the usually PS-free cell surface along with changes in the functions of surface molecules that prevent phagocyte recognition of viable cells from repulsive to adhesive then act as a marker for extracellular phagocytes aiding in the recognition and removal of apoptotic cells [64, 79]. The phagocytes then recognize the apoptotic cells and digest them prior to the rupturing of the plasma membrane, thus preventing cellular components from releasing and ending up as intercellular waste [12, 80]. The efficiency with which apoptotic cells are terminated by phagocytes compared with necrotic cells is a distinguishing characteristic of apoptosis [64].

Although hPLSCR1 is involved in the externalization of PS, it has been shown that PS exposure during apoptotic events is specific to cell type and does not correlate with the expression level of hPLSCR1, which suggests that other cellular components might also be involved in apoptosis [81].

Observations of apoptosis have suggested that it contributes to a high rate of cell death in malignant tumors and could even promote the progression of tumors [82]. Identification of apoptotic cells would thus be helpful in preventing cancer from spreading. Unfortunately certain mutations during cancer progression interfere with the genes that control apoptosis, and failure of apoptosis can result in uncontrolled cancer growth [11, 13]. In a model where interrupted apoptosis results in cancer growth, the tumor cells contain the enzymes that are necessary for carrying out apoptosis, but lack the protein responsible for launching the process [80]. It was

suggested that the protein in question is the tumor suppressor p53, which may be inactive in cancerous cells [80, 83]. Since then, new protein families that might restore the function of p53 have been under study [80].

In addition to preventing tumor progression, further knowledge on apoptosis might be used in curing Alzheimer's disease. The death of nerve cells in Alzheimer's disease is thought to be the result of apoptotic processes, and thus anti-apoptotic drugs have been used in helping Alzheimer's patients [84]. However, there is much discussion on whether the neurons are really dying due to apoptosis as there is a lack of characteristic signs [84].

4. THE BACKGROUND OF MOLECULAR DYNAMICS SIMULATIONS

Even though the imaging of nanoscale systems such as cells is possible with modern techniques, the optical limit prevents the observation of atomic-scale events that happen on very fast time scales. Therefore computer simulations based on molecular models are currently the easiest way to observe, for instance, the events taking place within a cell membrane. The way to do this is to build a molecular model of the target system and change the atomic positions step by step by solving their Newton's equations of motion from the forces acting on them. By repeating this process for a desired time, the evolution of the system can be observed.

The basic idea behind molecular dynamics is simple, however the simulations themselves may be difficult to process. The durations of the simulations have to be long enough for the desired event to reliably occur, which can take anything from nanoseconds to milliseconds depending on the event, while the time steps are limited by the characteristic times of atomic-scale events such as bond vibrations, which are of the order of femtoseconds. Therefore the used time steps must also be of the order of femtoseconds in order to simulate the system realistically. Along with a large system, this can result in a very time and resource consuming simulation. In order to speed the calculations up, approximations and cut-off radii for long range interactions are used and the simulated system is kept as small as possible. To prevent the problems arising on the edges of the simulation box, periodic boundary conditions are used, in which the used box is replicated in every direction to form an infinite lattice. As a particle crosses the boundary, its image appears on the other side of the original box. The simulation box must also be large enough to prevent the particles from interacting with their copies. In order to prevent the temperature and the pressure of the system from drifting, a thermostat and a barostat have to be implemented. Temperature coupling is usually done by weakly coupling the system to an external heat bath and the pressure coupling by varying the size of the simulation box when necessary.

The molecular dynamics simulations in this study were performed using GROMACS [85, 86], which is an engine for performing molecular dynamics simulations and energy minimization. The basics of the molecular dynamics simulation processes are discussed in this chapter, beginning with the simulation process itself,

followed by the algorithm used for molecular dynamics. A short review of the biologically relevant force fields consisting of bonded and non-bonded interactions between atoms is given along with a short introduction to the v-rescale–thermostat and the Parrinello–Rahman barostat used in temperature and pressure coupling. The algorithms used for bond constraints are introduced and the mechanics of umbrella sampling, which is used to calculate the free energy differences along a reaction path, is also briefly reviewed along with its error analysis methods. The user manual for GROMACS [87] gives a good view into the theory of molecular dynamics, so most of the topics discussed in this chapter are based on it.

4.1 The Simulation Process

In order to simulate a system, the topologies of the molecules have to be created and the acting force fields have to be listed. The topology describes the molecular structure of the system by including a list of all the atoms in the system and of the interactions between these atoms. The atom types are characteristically listed in the topology of the system. Different atom types exist for a single atom as its parameters depend on its bonding and location in molecules. For instance, a carbon atom in a ring structure has different parameters than one in a hydrocarbon chain due to its different partial charge and interactions with nearby atoms. If a certain atom type is missing, a new one can be created and its parameters either approximated from an existing atom type, calculated quantum mechanically, or obtained via trial and error through varying their values and comparing the simulation results with experiments. The new atom type must then be appended to the topology. The bond parameters and atom parameters for both bonded and non-bonded interactions for different atom types are also found in the force field. Again, missing bond types can be manually created. The topology file also includes possible position restraints which can be used to hold certain parts of the system in place. The Hamiltonian operator for each atom is calculated from the given parameters.

In addition to the topology, the coordinates of single atoms in the molecular structure are also required. The coordinates are basically given as a list of all the atoms in the system and their respective locations in the simulation box. The coordinate file can also be visualized and analyzed with ,e.g., VMD (Visual Molecular Dynamics), which is a software for visualizing biomolecular systems [70, 88].

Once the topology and the atom coordinates are in order, they can be processed into a run input file, which contains all the necessary parameters for molecular dynamics simulations. They are entered in the input file with a molecular dynamics parameter file in which, e.g., the type and length of the simulation and all related simulation parameters are given. The input file can then be run, and once finished, the simulation program will give a trajectory file of the locations and velocities of

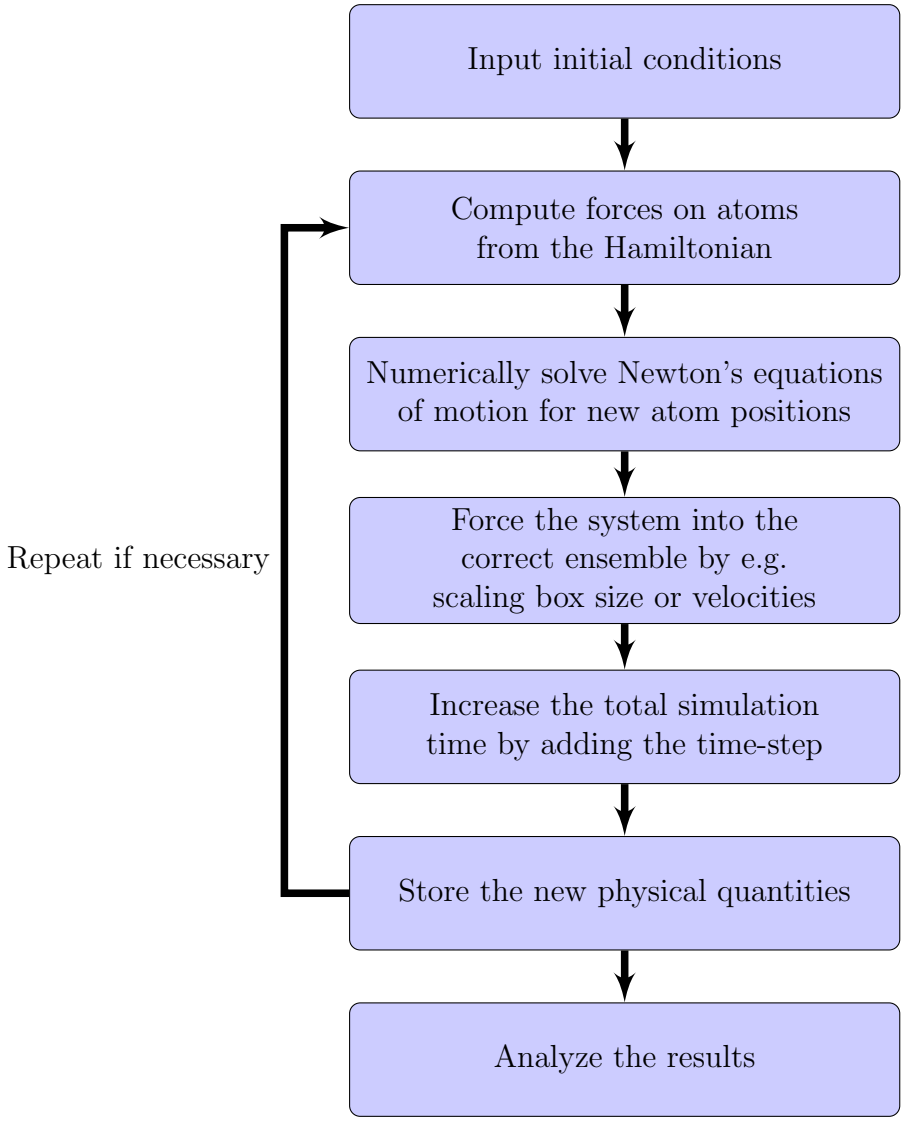

Figure 4.1: A flow chart of the molecular dynamics algorithm.

the particles in the system at different time intervals as an output. In GROMACS, visualizing the previous coordinate file along with this trajectory file allows the user to follow the evolution of the system as time passes. The trajectory file can also be used for further analysis on the dynamics of the system, such as measuring structural changes of proteins or calculating system energetics.

Before conducting any molecular dynamics simulations, it is important to first minimize the energy of the system. This is done by conducting energy minimization runs that discard thermal motion by setting the temperature to zero and equilibrate the molecular structures to the lowest energy at absolute zero by using algorithms such as the steepest descents method or conjugate gradient minimization. Setting the temperature to absolute zero also removes the effect that entropy has on the total energy. Running energy minimization to relax the original system is important, as

conducting MD simulations on a non-equilibrated system often causes it to blow up due to unstabilized structures.

The global algorithm for molecular dynamics simulations is shown in Fig. 4.1. By repeating the previous steps after analyzing the results of each step and modifying the parameters if necessary, one should eventually obtain the desired results from the simulation process.

4.2 Computational Methods

Molecular dynamics (MD) simulations are based on solving the Newton's equations of motion for a system of N interacting atoms, where the forces are negative derivatives of a potential energy function $V(\mathbf{r}_1, ..., \mathbf{r}_N)$

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = \mathbf{F}_i = -\frac{\partial V}{\partial \mathbf{r}_i}, \tag{4.1}$$

where m_i and $\mathbf{r}_i$ are the mass and the position of particle i and $\mathbf{F}_i$ is the force acting on it. The potential energy function is included along with related parameters in the force field that is applied to the system. If the force field and the positions of the particles are known, the forces acting on each particle can be calculated. Then numerical integration can be used to calculate the positions and the velocities of the particles after a certain time step. The numerical integration method for Newton's equations of motion used by GROMACS is the so called leap-frog algorithm [89].

The leap-frog algorithm is a third order integration method with similarities to the Verlet algorithm. Its major advantages in mechanics problems include time-reversibility and subsequently conservation of the energy of the system. The leap-frog equations read

$$\mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right) = \mathbf{v}_i \left(t - \frac{1}{2}\Delta t \right) + \frac{\Delta t}{m_i} \mathbf{F}_i(t), \tag{4.2}$$

$$\mathbf{r}_i (t + \Delta t) = \mathbf{r}_i (t) + \Delta t \mathbf{v}_i \left(t + \frac{1}{2}\Delta t \right), \tag{4.3}$$

where $\mathbf{v}_i$ and $\mathbf{r}_i$ are the velocity and the position of the particle i and Δt is the length of the time step. As can be seen from the equations, the leap-frog algorithm never calculates the positions at the same point in time as the velocities, which gives the leap-frog algorithm an important quality of time-reversibility. The obtained coordinates and velocities are saved into a trajectory file at predetermined time intervals allowing the analysis of the particles' trajectories.

4.3 Force Fields

Various types of force fields can be utilised in GROMACS. The one used in this study was the Optimized Potentials for Liquid Simulations (OPLS) force field using an all-atom (AA) representation, which includes even the nonpolar hydrogen atoms [90]. Although computationally slower than the united-atom (UA) model, which includes the hydrogen atoms linked to carbon atoms only in the carbon atoms' parameters, it provides much higher accuracy and allows more flexibility in torsional energetics and charge distributions [90]. Therefore given that enough computational power is available, it is preferred for systems that include hydrocarbons if high accuracy is desired.

A force field comprises the interaction potentials between different particles of the system. They can be divided into two categories depending on whether the particles are connected by a chemical bond or if the particles only affect each other through long range interactions. The bonded interactions of OPLS-AA constitute of covalent bond-stretching, angle-bending and proper or improper dihedrals. The non-bonded interactions include Lennard-Jones potentials, which include the van der Waals forces and steric repulsion, and Coulomb potentials. The total potential function in OPLS can be obtained by adding the bonded and non-bonded potentials together:

$$V_{\text{total}} = V_{\text{bonded}} + V_{\text{non-bonded}} \tag{4.4}$$

$$\begin{aligned} &= \sum_{\text{bonds}} V_b\left(r_{ij}\right) + \sum_{\text{angles}} V_a\left(\theta_{ijk}\right) + \sum_{\text{dihedrals}} V_d\left(\phi_{ijkl}\right) \\ &\quad + \sum_{i=1}^{N} \sum_{j=i+1}^{N} V_{\text{LJ}}\left(r_{ij}\right) + \sum_{i=1}^{N} \sum_{j=i+1}^{N} V_{\text{c}}\left(r_{ij}\right). \end{aligned} \tag{4.5}$$

The components of the sums are further explained in the following subchapters. All of them include certain bond characteristic parameters, which are found listed in the data files of the employed force fields.

4.3.1 Bonded Interactions

As mentioned above, bonded interactions are not exclusively pair interactions as they also include 3-body and 4-body interactions through chemically bonded atoms. The bond between two atoms should not be considered as a rigid rod, but rather as a stiff spring, which can stretch on occasion. The bonded interactions described below are used in OPLS force fields.

The most basic interaction between two bonded atoms is called bond stretching, which occurs when the energy in the bond changes. It is represented by a harmonic

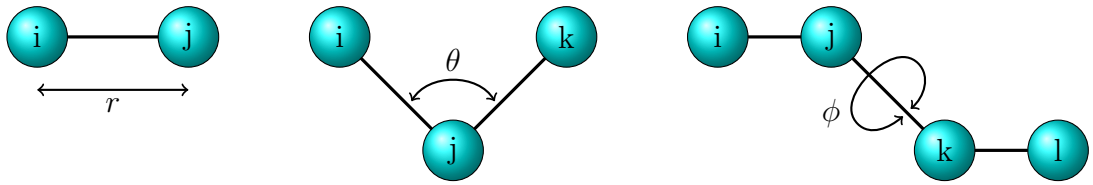

Figure 4.2: The bonded interactions of OPLS force fields are illustrated in this figure. Bond stretching occurs with two bonded atoms i and j at a distance r. Angle bending can be observed on a triplet of connected atoms i–j–k as a function of the angle θ. The dihedral angles are defined by the torsion angles ϕ between two planes ijk and jkl defined by a set of four connected atoms i–j–k–l.

potential

$$V_b\left(r_{ij}\right) = \frac{1}{2}k_{ij}^r\left(r_{ij} - r_{ij}^0\right)^2, \tag{4.6}$$

where i and j are the two bonded atoms, k_{ij}^r the force constant of the bond, r_{ij} the distance between the particles, and r_{ij}^0 the reference distance, which is the distance where the energy of the bond is the lowest.

As the length of a bond can change, so can the angle between two bonds due to interactions on the triplet of connected atoms $i-j-k$. This angle-bending is also represented by a harmonic potential

$$V_a\left(\theta_{ijk}\right) = \frac{1}{2}k_{ijk}^\theta\left(\theta_{ijk} - \theta_{ijk}^0\right)^2, \tag{4.7}$$

where θ_{ijk} is the angle between the two bonds with the atom j in the center, k_{ijk} the force constant of the angle, and θ_{ijk}^0 the reference angle.

A dihedral angle refers to the angle between two planes formed by four connected atoms $i-j-k-l$. For proper dihedrals a zero angle in OPLS force fields corresponds to the *cis* configuration where the atoms i and l are on the same side of the bond. Likewise a 180° angle in OPLS force fields corresponds to the *trans* configuration where the atoms i and l are on opposite sides of the bond. For other force fields, the angles can be described the other way around. There are two ways to describe proper dihedrals in OPLS force fields. They can be of periodic type

$$V_d\left(\phi_{ijkl}\right) = k_\phi\left(1 + \cos\left(n\phi_{ijkl} - \phi_s\right)\right), \tag{4.8}$$

where n is the periodicity, ϕ_{ijkl} the angle between the planes ijk and jkl, ϕ_s the reference angle and k_ϕ the force constant. They can also be described by a Ryckaert–Belleman's (RB) function commonly used in alkane chains found in phospholipid tails

$$V_{rb}(\phi_{ijkl}) = \sum_{n=0}^{5} C_n\left(\cos\left(\psi\right)\right)^n, \tag{4.9}$$

where $C_n\,(n = 0, ..., 5)$ are Ryckaert–Belleman torsion parameters. The OPLS force

fields use a modified form of the RB function, which utilises Fourier series. The purpose of improper dihedrals is to force planar groups to stay planar or to prevent molecules from becoming their mirror images. Improper dihedrals are either of periodic type, in which case the potential function is identical to the proper periodic dihedral, or of harmonic type, where the harmonic potential is

$$V_{id}\left(\xi_{ijkl}\right) = \frac{1}{2} k_{\xi}\left(\xi_{ijkl} - \xi_0\right)^2, \tag{4.10}$$

where ξ_{ijkl} is the angle between the planes ijk and jkl, ξ_0 the reference angle and k_{ξ} the force constant.

4.3.2 Non-bonded Interactions

The non-bonded interactions contain a repulsion term and a dispersion term, which are combined in the Lennard-Jones (LJ) interaction, and a Coulomb term. The value of the Lennard-Jones potential depends heavily on the distance r_{ij} between the two interacting atoms, because it contains the repulsive r_{ij}^{-12} term and the attractive r_{ij}^{-6} term. Lennard-Jones potential can be written as

$$V_{\mathrm{LJ}}(r_{ij}) = 4\epsilon_{ij}\left(\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right), \tag{4.11}$$

where ϵ_{ij} is the mimimum energy of the potential well and σ_{ij} the van der Waals distance, which defines the point where the potential turns from repulsive to attractive [91].

The Coulombic interaction between two charged particles can be calculated from

$$V_{\mathrm{c}}(r_{ij}) = \frac{1}{4\pi\epsilon_0}\frac{q_i q_j}{\epsilon_r r_{ij}}, \tag{4.12}$$

where q_i and q_j are the charges of the particles in question, ϵ_0 is the permittivity of the vacuum, ϵ_r the relative permittivity of the medium, and r_{ij} the distance between the two particles.

Lennard-Jones and Coulombic interactions are long range interactions. Thus, in order to speed calculations up, a certain cut-off radius is used to truncate non-bonded interactions between two atoms. The cut-off radius has to be limited though, because of the periodic boundary conditions of the system box, to prevent any molecule from interacting with itself. The longest allowed cut-off radius is half the shortest box vector. GROMACS uses a periodically updated neighbor list to keep track of molecules within the cut-off radius and applies the non-bonded interactions only to the listed pairs.

4.3.3 Long-range Interactions

Force fields are parametrized in order to allow certain approximations at long distances. For instance, van der Waals interactions decrease rapidly with distance, which allows us to use a cut-off radius, which limits the amount of van der Waals interactions that need to be considered. Too high a cut-off radius, however, can result in particles interacting with themselves due to the periodic boundary conditions and therefore some consideration must be put in selecting a proper cut-off radius.

Even though a cut-off radius is implemented to speed calculations up, long-range interactions decay too slowly to be neglected completely. Beyond the defined cut-off radius a method called Ewald summation is applied to calculate long-range interactions for electrostatics in periodic boundary conditions [92]. It is a technique for calculating the sum of long-range interactions between particles and all their infinite periodic images in an efficient way. The total electrostatic energy of N particles and all their periodic images can be written as

$$V = \frac{f}{2} \sum_{\mathbf{n}} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{\mathbf{r}_{ij,\mathbf{n}}}, \tag{4.13}$$

where $\mathbf{n} = (n_x, n_y, n_z)$ is the box vector and $\mathbf{r}_{ij,\mathbf{n}}$ the real distance between the charges. The terms with $i = j$ should be omitted when $\mathbf{n} = \mathbf{0}$ [92].

Because the sum is very slowly convergent, the Ewald summation method converts it into two quickly converging terms and a constant term

$$V = V_{\text{dir}} + V_{\text{rec}} + V_0, \tag{4.14}$$

where the real (direct) space sum can be written as

$$V_{\text{dir}} = \frac{f}{2} \sum_{i,j}^{N} \sum_{\mathbf{n}} q_i q_j \frac{\operatorname{erfc}(\alpha \mathbf{r}_{ij,\mathbf{n}})}{\mathbf{r}_{ij,\mathbf{n}}}, \tag{4.15}$$

the reciprocal sum as

$$V_{\text{rec}} = \frac{f}{2\pi V} \sum_{i,j}^{N} q_i q_j \sum_{\mathbf{m} \neq \mathbf{0}} \frac{\exp\left(-(\pi \mathbf{m}/\alpha)^2 + 2\pi i \mathbf{m} \cdot (\mathbf{r}_i - \mathbf{r}_j)\right)}{\mathbf{m}^2}, \tag{4.16}$$

and the constant term as

$$V_0 = -\frac{f\alpha}{\sqrt{\pi}} \sum_{i}^{N} q_i^2, \tag{4.17}$$

where α is a parameter used to determine the relative weight of the direct and reciprocal sums, V the volume of the simulation box, $\mathbf{m} = (m_x, m_y, m_z)$ a reciprocal-

space vector, and $\mathrm{erfc}\,(x)$ the complimentary error function

$$\mathrm{erfc}\,(x) = 1 - \frac{2}{\sqrt{\pi}} \int_0^x \exp\left(-u^2\right) \mathrm{d}u \tag{4.18}$$

which decreases monotonically as x increases [92]. These equations allow the use of a short cut-off radius of the order of 1 nm, where direct sums are calculated at ranges below the cut-off radius and reciprocal sums at ranges beyond the cut-off radius, when a high enough relative weight factor α is chosen. The computational cost, however, increases as $\mathcal{O}\left(N^2\right)$, and therefore Ewald summation in its current form is not viable for large systems.

A faster method called Particle-mesh Ewald (PME) was developed in order to improve the calculation rate of the reciprocal sum in large systems [93]. The direct sum is explicitly calculated by using cut-offs, but instead of summing the wave vectors directly in the reciprocal sum, it assigns the charges to a grid using either Lagrange interpolation or more accurate cardinal B-spline interpolation. A Fourier transform $\mathcal{F}\,(Q)$ is then applied to the charge grid using a 3D FFT algorithm and the reciprocal energy term is then obtained by a single sum over the grid in the reciprocal space

$$V_{\mathrm{rec}} = \frac{f}{2\pi V} \sum_{\mathbf{m}\neq\mathbf{0}} \frac{\exp\left(-\left(\pi\mathbf{m}/\alpha\right)^2\right)}{\mathbf{m}^2} \mathcal{F}\,(Q)\,(m)\,\mathcal{F}\,(Q)\,(-m)\,. \tag{4.19}$$

The potential energy at points of the grid is obtained by inverse transformation, and the forces on each atom is calculated by using the interpolation factors.

The computational cost of the PME algorithm increases as $\mathcal{O}\left(N \log N\right)$, making it significantly faster than Ewald summation. In addition to computational speed, the PME method also gains in accuracy, which can be further improved by adjusting the parameters [92].

4.4 Temperature and Pressure Coupling

In order to keep the system in a desired ensemble, controlling the pressure and the temperature is required to prevent them from drifting due to various computational errors or external forces. Temperature control is usually done by coupling the system to an external heat bath, the temperature of which is set at the desired temperature.

The Berendsen thermostat [94] uses an external heat bath set at temperature T_0 and changes the temperature T of the system exponentially towards the temperature of the heat bath as

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau_T}, \tag{4.20}$$

in which τ_T is the time constant used for exponential decay. The time constant

and the strength of the coupling can be varied depending on the needs of the user, resulting in an ensemble that is close to either microcanonical or canonical [95]. It is also very efficient in system equilibration and thus widely used.

An expansion of the Berendsen thermostat, the velocity rescaling method uses an additional stochastic term to ensure correct kinetic energy distribution in the system [96]. The stochastic term can be written as

$$\mathrm{d}K = (K_0 - K)\,\frac{\mathrm{d}t}{\tau_T} + 2\sqrt{\frac{KK_0}{N}}\,\frac{\mathrm{d}W}{\sqrt{\tau_T}}, \tag{4.21}$$

where K is the kinetic energy of the system, K_0 the target kinetic energy, N the number of degrees of freedom, and $\mathrm{d}W$ a Wiener noise term. Use of the v-rescale method produces a correct canonical ensemble with all the advantages of the Berendsen thermostat with addition of conservation of energy in the NVT ensemble.

The pressure of the system can be coupled in several ways. The Parrinello–Rahman barostat [97] forces the box vectors represented by the matrix $\mathbf{b}$ to follow the matrix equation of motion

$$\frac{\mathrm{d}\mathbf{b}^2}{\mathrm{d}t^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}\left(\mathbf{P} - \mathbf{P}_{\mathrm{ref}}\right), \tag{4.22}$$

thus controlling the pressure $\mathbf{P}$ by changing the volume V of the box. The strength of the coupling is determined by a matrix parameter $\mathbf{W}$, which forces the system towards the reference pressure $\mathbf{P}_{\mathrm{ref}}$. The matrix parameter $\mathbf{W}$ is defined as

$$\left(\mathbf{W}^{-1}\right)_{ij} = \frac{4\pi^2\beta_{ij}}{3\tau_p^2 L}, \tag{4.23}$$

where β_{ij} are the isothermal compressibilities, τ_p is the pressure time constant, and L the largest box matrix element. The pressure time constant now refers to the period of pressure fluctuations at equilibrium and is not equivalent to the time constant used in the Berendsen barostat or thermostat, thus it should be set at a 4–5 times greater value. The pressure is presented as a 3×3 tensor as the system may not be isotropic. The Parrinello–Rahman barostat allows the simulation box to change its size, prevents instantaneous changes in atom positions, and provides realistic fluctuations in pressure, making it a reliable choice when selecting pressure coupling methods. It is a barostat which allows accurate calculation of the thermodynamic properties of the system. For a system with the pressure far from equilibrium, however, use of the Parrinello–Rahman coupling may result in very large box oscillations that cause the system to blow up. In such cases, care must be taken when selecting the time constant.

4.5 Bond Constraints

The time step in classical molecular dynamics simulations is limited by bond oscillations with a high frequency and a low amplitude. By replacing the bond vibrations with constraints that represent the average physical behavior of the bonds, the time step can be increased by a factor of four [98]. These constraints are algorithms applied to the Newton's equations of motion that force both the bonds to stay at correct lengths and/or the bond angles at correct values after the integration of forces [98].

A common constraint algorithm is the Linear Constraint Solver algorithm (LINCS), which resets the bonds to their original lengths after each unconstrained update by calculating the new constrained positions from both the old positions and the new unconstrained positions [98]. The algorithm is stable, because the constraints themselves are reset instead of their derivatives. It is also three or four times faster than the SHAKE algorithm, which is another standard approach for molecular dynamics [99]. Unlike SHAKE, LINCS is a non-iterative method, which allows for parallelization of the algorithm over several processors, decreasing CPU time and improving load balance [98]. Both SHAKE and LINCS algorithms are time-reversible, because they reach convergence by multiple bond rotation corrections and by additional matrices in the expansion [98]. The LINCS algorithm works on such a mechanism that the constraint algorithm is no longer the limiting factor on the time step, meaning that we can use a time step that is as long as the physical conditions of the system permit [98].

A fourth of the total computational time in typical systems is spent on constraints of the solvent liquid [100]. Therefore a faster constraint algorithm called SETTLE is typically used for the large number of rigid water molecules in each system [100]. The SETTLE algorithm is an analytical solution of SHAKE, three to seven times faster than RATTLE, an extension of SHAKE [100]. In addition, it is non-iterative and accurate, but too complicated for larger molecules [100]. Therefore, it is well suitable for small tri-atomic molecules such as water modelled by three-point SPC or TIP3P models.

Typically a badly equilibrated system results in problems such as unnatural bond rotating or stretching in the LINCS or SETTLE algorithms due to large forces applied on the atoms. Unnatural bond rotating or unsettled water molecules may in turn cause the system to blow up. These problems can be avoided either by shortening the time step or by better preparing the system with additional energy minimization runs or revision of the system structure.

4.6 Umbrella Sampling

A major challenge in molecular dynamics is the calculation of free energy differences, which are the driving forces of every chemical process [101]. It is challenging, because entropy and free energy of a system can not be measured simply as averages of phase space coordinate functions of the system, but rather as functions related to the volume in the phase space, that is accessible to the system [102]. These thermal quantities can not thus be measured directly from the simulation, so an alternate measurement path must be found [102].

By the basic concepts of thermodynamics, the Helmholtz free energy difference ΔF between the initial and final configurations is equal to the amount of work W performed to the system if the parameters are changed infinitely slowly along the reaction path [103]. Usually the parameters are changed at a finite rate, which leads to the total work being greater than the free energy differences, so that

$$\langle W \rangle \geq \Delta F, \tag{4.24}$$

where $\langle W \rangle$ denotes the ensemble average of measurements of the total work [103]. The difference between the two variables is the dissipated work, which is associated with the increase of entropy in an irreversible process [103]. The inequality between work and the difference in free energy can be replaced by an equality derived by Jarzynski et al. [103],

$$\langle \exp\left(-\beta W\right) \rangle = \exp\left[-\beta \Delta F\right], \tag{4.25}$$

where β is the inverse temperature $\beta = 1/(k_{\mathrm{B}}T)$, in which k_{B} is the Boltzmann's constant and T the absolute temperature. This result is independent of both the reaction path and the rate at which the parameters are changed [103]. Solving the above equation for free energy gives

$$\Delta F = -1/\beta \ln \langle \exp\left(-\beta W\right) \rangle, \tag{4.26}$$

where the amount of work can be measured from MD simulations allowing the result to be implemented into calculations of differences in free energy [103].

One method of obtaining the change of free energy along the reaction coordinate with the help of the result above is to use biased molecular dynamics, so called umbrella sampling, which is a viable alternative to thermodynamic integration [101, 104]. Umbrella sampling is based on bias potentials w_i that drive a system from one thermodynamic state to another along a reaction coordinate ξ. The reaction path is divided into a series of windows and an MD simulation is performed at each window. The change in free energy at each window can be calculated from the sampled distribution of the system along the reaction coordinate. The windows are

set to overlap each other sufficiently so that the distributions can be combined to obtain the total free energy difference.

There are various ways to select the reaction coordinate ξ. It does not necessarily have to be selected along a certain axis as the reaction path might not be a straight line. It is also possible to measure the free energy differences of an angle between two bonds by selecting the angle in question as the reaction coordinate. Similarily, the reaction coordinate can be, for instance, the length of a bond or the distance between two atoms. In this thesis, however, the reaction coordinate is selected as a straight line between a phospholipid and rhodopsin or a straight line across the bilayer.

The probability distribution of an unbiased ergodic system in a small interval dξ around ξ can then be written as

$$P(\xi) = \frac{\int \delta[\xi(r) - \xi] \exp[-\beta E(r)] \, \mathrm{d}^N r}{\int \exp[-\beta E(r)] \, \mathrm{d}^N r}, \tag{4.27}$$

where $P(\xi)\mathrm{d}\xi$ can be interpreted as the probability distibution of the system, E the potential energy, and N the number of degrees of freedom of the system [101]. The free energy along the reaction coordinate can be calculated from the probability distribution:

$$A(\xi) = -\frac{1}{\beta} \ln P(\xi), \tag{4.28}$$

where $A(\xi)$ is also known as the potential of mean force (PMF) [101]. PMF can therefore be directly obtained from the MD simulations.

Umbrella sampling differs from the aforementioned methods by addition of a bias potential. These so called umbrella potentials $w_i(\xi)$ are usually harmonic and dependent only on the reaction coordinate, and they are applied to the system to ensure efficient sampling along the reaction coordinate [101]. Commonly used umbrella potentials are of type

$$w_i(\xi) = \frac{K_i}{2(\xi - \xi_i^C)^2}, \tag{4.29}$$

where K_i is a force constant that is used to restrain the system at a predetermined position ξ_i^C $(i = 1, ..., N_w)$ for N_w umbrella windows [101]. The force constants have to be chosen carefully to make the umbrella potentials high enough to drive the system over the potential barrier, but also small enough to create wide enough distributions [101].

By conducting an MD simulation on a biased ergodic system, a biased distribution

$$P_i^b(\xi) = \frac{\int \exp\{-\beta[E(r) + w_i(\xi'(r))]\} \delta[\xi'(r) - \xi] \, \mathrm{d}^N r}{\int \exp\{-\beta[E(r) + w_i(\xi'(r))]\} \mathrm{d}^N r} \tag{4.30}$$

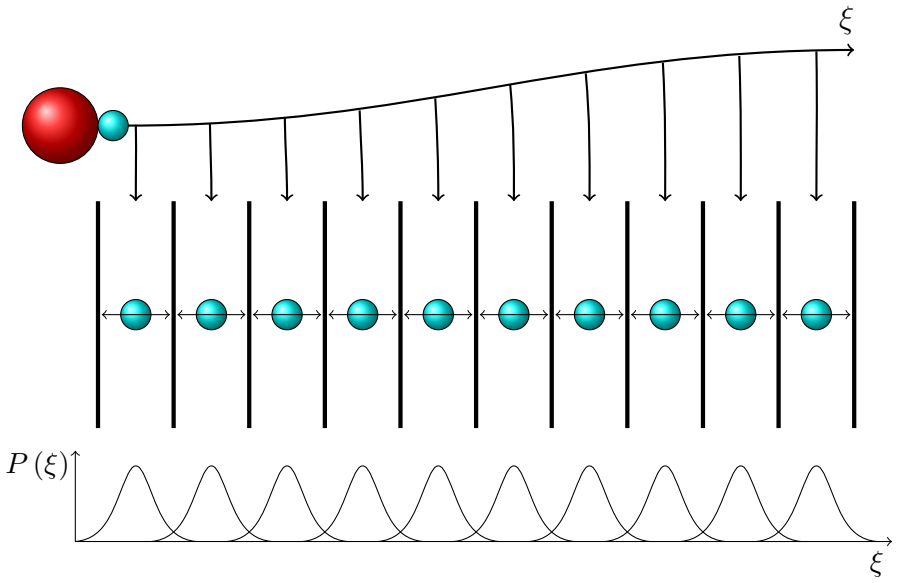

Figure 4.3: The basic idea behind umbrella sampling. The molecule represented by the cyan circle is pulled along the reaction coordinate ξ. A number of sampling windows are selected along the coordinate path in a way that sufficient overlap can be seen in the resulting histograms that show the probability $P(\xi)$ of the molecule being at a certain point along the reaction path. Independent MD simulations are then conducted within each sampling window. Sufficient overlap of the sampling windows ensures that the continuous energy function can be derived from the separate simulations.

can be obtained [101]. Because the umbrella potentials depend only on the reaction coordinate and the integration is performed over all degrees of freedom except ξ, a relation between the unbiased system in Eq. 4.27 and the biased system in Eq. 4.30 can be written as

$$P_i^u(\xi) = P_i^b(\xi) \exp[\beta w_i(\xi)] \langle \exp[-\beta w_i(\xi)] \rangle, \tag{4.31}$$

where $\langle \exp[-\beta w_i(\xi)] \rangle$ is the ensemble average of the system trajectories [101, 103]. As the biased distributions P_i^b can again be obtained from the MD simulations and the umbrella potentials $w_i(\xi)$ are given analytically, the unbiased free energies can now be calculated from

$$A_i(\xi) = -\frac{1}{\beta} \ln P_i^b(\xi) - w_i(\xi) + F_i, \tag{4.32}$$

where $F_i = -1/\beta \ln\langle \exp[-\beta w_i(\xi)] \rangle$ is obtained from Eq. 4.26 and is independent of the reaction path ξ [101, 103].

In order to calculate the global free energy $A(\xi)$, the potentials of mean force first have to be calculated for every window and then combined to a global PMF. Umbrella sampling simulations can be analyzed for instance with a weighted histogram analysis method (WHAM), which aims to minimize the statistical error by calcu-

lating the unbiased global distribution from a weighted average of the individual distributions of N_w umbrella windows. The weighted average can be written as

$$P^u(\xi) = \sum_i^{N_w} p_i(\xi) P_i^u(\xi), \tag{4.33}$$

where the weights p_i are selected to minimize the statistical error of $P^u(\xi)$ so that

$$\frac{\partial \sigma^2 (P^u(\xi))}{\partial p_i} = 0 \tag{4.34}$$

under the normalizing condition $\sum p_i = 1$ [101]. The calculations result in

$$p_i = \frac{a_i}{\sum_j a_j}, a_i(\xi) = N_i \exp\left[-\beta w_i(\xi) + \beta F_i\right], \tag{4.35}$$

where N_i is the number of steps calculated for window i and F_i for each window can be obtained from

$$\exp(-\beta F_i) = \int P^u(\xi) \exp\left[-\beta w_i(\xi)\right] \mathrm{d}\xi, \tag{4.36}$$

where $P^u(\xi)$ is the global unbiased distribution [101, 103]. After iterating these equations to solve the two unknowns $P^u(\xi)$ and F_i, the free energy for the whole reaction path can be calculated from Eq. 4.28.

Since umbrella sampling is used to measure the PMF of the molecule in question and PMF describes the probability of the molecule being in a particular position along the reaction coordinate, it can be used for instance to measure the energetically favorable positions of a protein on a membrane bilayer by pulling the protein along a coordinate ξ on the surface of the layer. In this thesis, umbrella sampling is used to measure the free energy requirements for a phospholipid to enter the interior of the putative phospholipid flippase, and subsequently the free energy requirements for the lipid to translocate to the other leaflet with the help of rhodopsin. By measuring the free energy curve, the possible flipping path for the phospholipids can be deduced.

4.6.1 Weighted Histogram Analysis

A standard technique for computing potentials of mean force is the weighted histogram analysis method (WHAM) first introduced by Kumar et al. [105]. Hub et al. implemented a tool for GROMACS called `g_wham` [106], which iteratively calculates the potentials of mean force for a number of umbrella windows and combines them to form the global free energy of the system. In addition to the global free energy, `g_wham` also estimates the statistical errors using bootstrap analysis. The

two equations iterated by WHAM are obtained from Eq. 4.33 and Eq. 4.36 and they read

$$P^u(\xi) = \frac{\sum_i^{N_w} g_i^{-1} h_i(\xi)}{\sum_j^{N_w} n_j g_j^{-1} \exp\left[-\beta\left(w_i(\xi) - F_j\right)\right]} \tag{4.37}$$

and

$$\exp\left(-\beta F_i\right) = \int P^u(\xi) \exp\left[-\beta w_i(\xi)\right] \mathrm{d}\xi, \tag{4.38}$$

where n_j is the total number of data points in the histogram h_i [106]. The statistical inefficiency g_i is defined by $g_i = 1 + 2\tau_i$, in which τ_i is the integrated autocorrelation time of the umbrella window i [106]. The two unknown qualities that are solved are the free energy constants F_j and the unbiased distribution $P^u(\xi)$ [106]. The number of iterations required for convergence depend on the number of histograms and the height of the PMF barriers and are typically in the order of tens to thousands [106].

The bootstrapping method is based on hypothetical observations that are obtained from probability distributions of the umbrella histograms. A number of bootstrapped trajectories are created for each umbrella window, which again yield new histograms for that umbrella window. WHAM can then be applied to the bootstrapped histograms in order to compute a bootstrapped PMF $W_b(\xi)$. By repeating this process for each bootstrapped trajectory, a large number of N_b bootstrapped PMFs $W_{b,k}(\xi)\,(k = 1, ..., N_b)$ can be obtained, and by calculating the standard deviation of these bootstrapped PMFs via

$$\sigma_{\mathrm{PMF}} = \left[(N_b - 1)^{-1} \sum_{k=1}^{N_b} \left(W_{b,k}(\xi) - \langle W_b(\xi)\rangle\right)^2\right]^{1/2}, \tag{4.39}$$

the uncertainty of the original PMF can be calculated and used as an error estimate [106]. The term $\langle W_b(\xi)\rangle = N_b^{-1} \sum_{i=k}^{N_b} W_{b,k}(\xi)$ denotes the average value of the bootstrapped PMFs at position ξ [106]. The error estimates can be considered accurate if the umbrella histograms have converged sufficiently.

By simulating each umbrella window multiple times from independent initial conditions, hypothetical trajectories need not be used, and several complete histograms can be bootstrapped resulting in a more accurate error estimate in cases of long autocorrelation times [106]. Usually bootstrapping of hypothetical trajectories gives a reliable enough error estimate and there is no need to carry out additional independent simulations [106].

5. MOLECULAR DYNAMICS SIMULATIONS OF RHODOPSIN IN A LIPID BILAYER

Before performing any simulations, a molecular model of the system first has to be built. The topologies of the protein and the phospholipids have to be constructed and their parameters matched with the selected force field. This part usually includes some trial and error as ready parameters for uncommon molecules may originally exist for a different force field. The built bilayer structure then has to be properly hydrated and equilibrated before performing any simulations. Selection of a correct ensemble and proper coupling-methods for temperature and pressure is also critical for realistic results.

The properties of the simulated system is explained in this chapter and the employed simulation parameters and methods are given in detail so that the simulations can be reproduced later. The steered MD methods used for lipid pulling are also explained along with the ideas behind them.

5.1 The System

The simulated system consisted of a lipid bilayer with a mixed lipid composition and the rhodopsin protein. This membrane was surrounded by a large number of water molecules in approximately a $(10\ \mathrm{nm})^3$ sized box with periodic boundary conditions. The size of the system was chosen such that the protein did not interact with itself.

The system was constructed by adding a total of 196 POPC molecules, 98 SSM molecules and 98 cholesterol molecules evenly in the two leaflets. The bilayer structure was then hydrated with a total of 28427 water molecules and properly equilibrated using energy minimization algorithms. Afterwards, the modified crystal structure of rhodopsin was added into the centre of the system and the overlapping lipids were removed. The resulting lipid composition included 162 POPC molecules, 81 SSM molecules and 92 cholesterol molecules with a total of 27559 water molecules in the solute. The salt concentration was set to a physiological level of 150 mmol/l by addition of 75 NaCl molecules to the solute along with three additional Na^+-ions to counter the charge imbalance created by rhodopsin. The system was again energy-minimized with the steepest descents method in order to remove all nonphysical interactions left from the addition of rhodopsin.

Our model of rhodopsin was constructed using the model of bovine rhodopsin by

Okada et al. [66], which was downloaded from the PDB database under the identifier 1U19. The downloaded crystal structure included two rhodopsin molecules and was constructed for the united-atom model. The second rhodopsin molecule was removed from the model in order to simplify the system as much as possible. The model was also converted to an all-atom representation for higher accuracy. This was done by adding the hydrogen atoms to their proper places with the GROMACS tool `pdb2gmx`.

The topology of the retinal molecule of rhodopsin was constructed from scratch and built to have a 11-*cis* conformation. RESP ESP charge Derive (R.E.D.) [107], which is a software for quantum mechanically calculating charge values for molecular fragments, was used to obtain the partial charges of the atoms in the retinal molecule. The newly constructed retinal was then added to the lysine residue 296 of rhodopsin through a Schiff-base-linkage to induce correct folding of the protein as mentioned for instance in Ref. [3]. The force field parameters for the Schiff-base were approximated from parameters of an imine functional group obtained from an article by Moyna et al. [108] and converted to the OPLS-AA format.

Two palmitoyl chains were also added to the rhodopsin at cysteine residues 322 and 323. The topologies of the palmitoyl chains were taken from the topology of POPC. The palmitoyl chains were connected to the sulfhydryl groups of the cysteine residues through thioester linkages and are believed to act as anchors between rhodopsin and the bilayer. Some connection between the palmitoyl chain anchors and scrambling activity has also been found in hPLSCR1, which might be beneficial for our study [21]. The force field parameters for the thioester linkage were obtained from an article by Nagy et al. [109] and converted to the OPLS-AA format.

A bilayer without rhodopsin was also constructed for comparing the free energy of flip–flop in a pure bilayer to the free energy of the rhodopsin-mediated flip–flop process. For quicker simulations, the pure bilayer was smaller in size, but had a similar lipid composition consisting of 64 POPC molecules, 32 SSM molecules and 32 cholesterol molecules hydrated with a total of 3868 water molecules. The salt concentration was set to a physiological level by addition of 10 NaCl molecules. The box size was approximately $6 \times 6 \times 7$ nm^3. The system was equilibrated for approximately 20 ns before any umbrella sampling simulations were conducted.

5.2 Simulation Parameters

The structure of the mixed bilayer with rhodopsin embedded in it was energy-minimized in order to equilibrate the system. Afterwards the system was simulated for approximately a total of 200 ns in order to see how the composition of the bilayer affects the conformation and the crystal structure of rhodopsin. The retinal chromophore was added to the system and some adjustments to its parameters were made at the early stages of the 200 ns simulation.

All molecular dynamics simulations were performed with the GROMACS software package 4.5.5 [85, 86]. The energy minimization runs were mainly performed using the steepest descent algorithm with LINCS constraints on all bonds. The conjugate gradient algorithm was used to balance the system with the newly added retinal for a better equilibration. All MD simulations were conducted with the leap-frog algorithm for integrating Newton's equations of motion. The time steps used in all simulations were 2 fs with the exception of 1 fs time steps during equilibration of some imbalanced systems.

The system was simulated in the NPT ensemble. The temperature of the system was coupled using the v-rescale method with separate heat baths for the membrane lipids, the protein, and for the solute. The reference temperatures were set at the normal human body temperature of 310.0 K with time constants of 0.1 ps. The pressure of the system was coupled using the semi-isotropic Parrinello–Rahman barostat. The pressures in the direction of the plane of the membrane were coupled separately with its normal vector using a reference pressure of 1 bar. The time constants for pressure coupling were set at 0.5 ps and the compressibilities at 4.5×10^{-5} bar^{-1}.

The cut-off radii for the non-bonded interactions were set at 1.0 nm, beyond which the Particle-Mesh Ewald electrostatics were used. The same cut-off radius was also used for the neighbour list and the Lennard-Jones interactions. Bonds were constrained using the LINCS algorithm with the exception of water molecules, which were constrained using the faster SETTLE algorithm. Water molecules were modelled using the accurate three-point model (TIP3P) discussed for instance in Ref. [110].

Steered MD simulations were used to move the phospholipids. These lipid pulling simulations were conducted by applying a force to the lipid headgroup, which moved the lipid either towards a certain molecule or in a certain direction at a constant rate. Several lipids around the protein were selected for pulling. In the first part of the simulations, the lipid headgroup was selected as the first pull group and rhodopsin as the second pull group. The first pull group was then pulled towards the center of mass of the second group on the membrane plane (xy-plane). The second part of the simulations consisted of pulling the phospholipid located next to, or within, rhodopsin through the membrane along the membrane normal direction (z-axis). In these cases only one pull group was used. The applied force constant for pulling was 1000 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ in all parts of the simulations. The pull rates varied from 0.3 nm/ns to 1.0 nm/ns. For the system with the pure bilayer, the steered MD simulations only consisted of pulling a phospholipid through the membrane along the z-axis by its headgroup. The steered MD mechanics that were used are further explained in the results section.

6. RESULTS AND DISCUSSION

In this chapter the results obtained in this thesis are shown and discussed. The main purpose of this thesis was to study phospholipid flip–flop in the presence of rhodopsin using molecular dynamics simulations. The aims were to establish the main flipping mechanics of rhodopsin and to calculate the free energy curve of rhodopsin-mediated flip–flop of POPC. The obtained results would then be used to confirm the hypothesis that rhodopsin is a phospholipid flippase, or possibly even a scramblase. The free energy curves presented were obtained by umbrella sampling and the statistical error estimates by weighted histogram analysis, using the GROMACS tool `g_wham`.

The first part of the results are obtained by pulling the phospholipid towards the center of rhodopsin with a goal in mind to see how close to the center the phospholipid can naturally travel. This information was then used to deduce the most likely flipping route for phospholipids. The six investigated routes are compared by examining the pulling forces required to move the phospholipid close to the protein center. One of the routes is chosen as an example and umbrella sampling is conducted to obtain the free energy differences along it.

In the second part of the results, a POPC molecule is pulled through a protein-free bilayer in order to establish the free energy difference between the bilayer center and the equilibrium position of phospholipids. The data was gathered to act as a comparison for rhodopsin-mediated flip–flop.

The third part of the results are obtained by examining the most likely route for lipid translocation and by pulling the phospholipid through the membrane along that specific route. The energy barrier of the translocation route in question is obtained by umbrella sampling. Whether phospholipid flip–flop along the route in question can occur by thermal fluctuations alone or with the help of external energy sources such as ATP is also discussed. The results are compared to previous experimental studies and mathematical approximations are also made to complement the results.

The fourth part of the results is used to determine whether the simulated system is stable. The stability is examined by calculating the root mean square fluctuations of the protein backbone and the root mean square deviations of all amino acid residues of rhodopsin. The data is also used to analyze whether any specific changes in rhodopsin structure could help in the lipid translocation process.

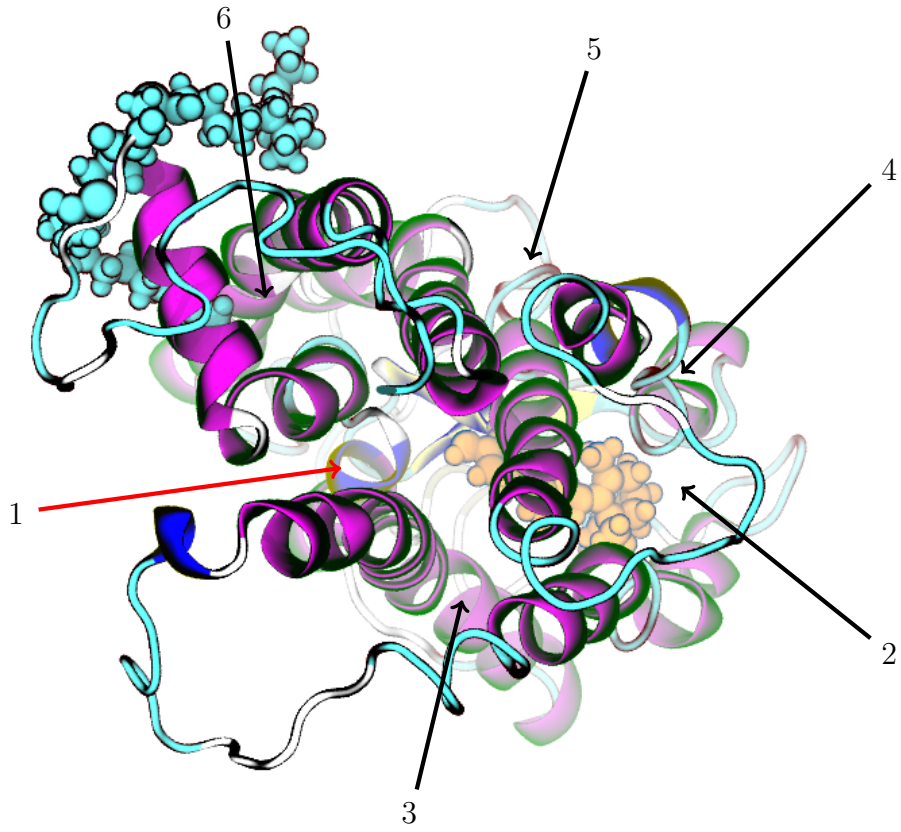

Figure 6.1: Rhodopsin from above. The six arrows show the pulling routes used for the six separate POPC molecules. All phospholipids were pulled from their headgroup towards the center of mass of rhodopsin in the xy-plane. The most promising route is shown by the red arrow on the left side.

6.1 The Flipping Routes of Rhodopsin

The seven transmembrane alpha-helices of rhodopsin form a total of six distinct gaps between each neighbouring helices. The most likely routes the phospholipids would take while flipping across the membrane are thought to be somehow linked to these gaps. In order to further examine the flipping routes, steered MD simulations were conducted on six separate POPC molecules, which were pulled one at a time towards the center of rhodopsin in order to determine the energetically most favorable route. A diagram of this process is shown in Fig. 6.1.

The pulling simulations were conducted using two different pull rates. The faster pull rate was set at 1.0 nm/ns and the slower pull rate at 0.3 nm/ns. The pull forces acting on the phospholipid headgroup during both pulls were compared to see whether the pulling rate has any notable effect on the route of the phospholipid. The pulling forces for all six pulling routes are shown in Fig. 6.2.

The first lipid was the only one to enter the protein interior through a gap between two alpha-helices. The pulling force increases quite rapidly though due to repulsive

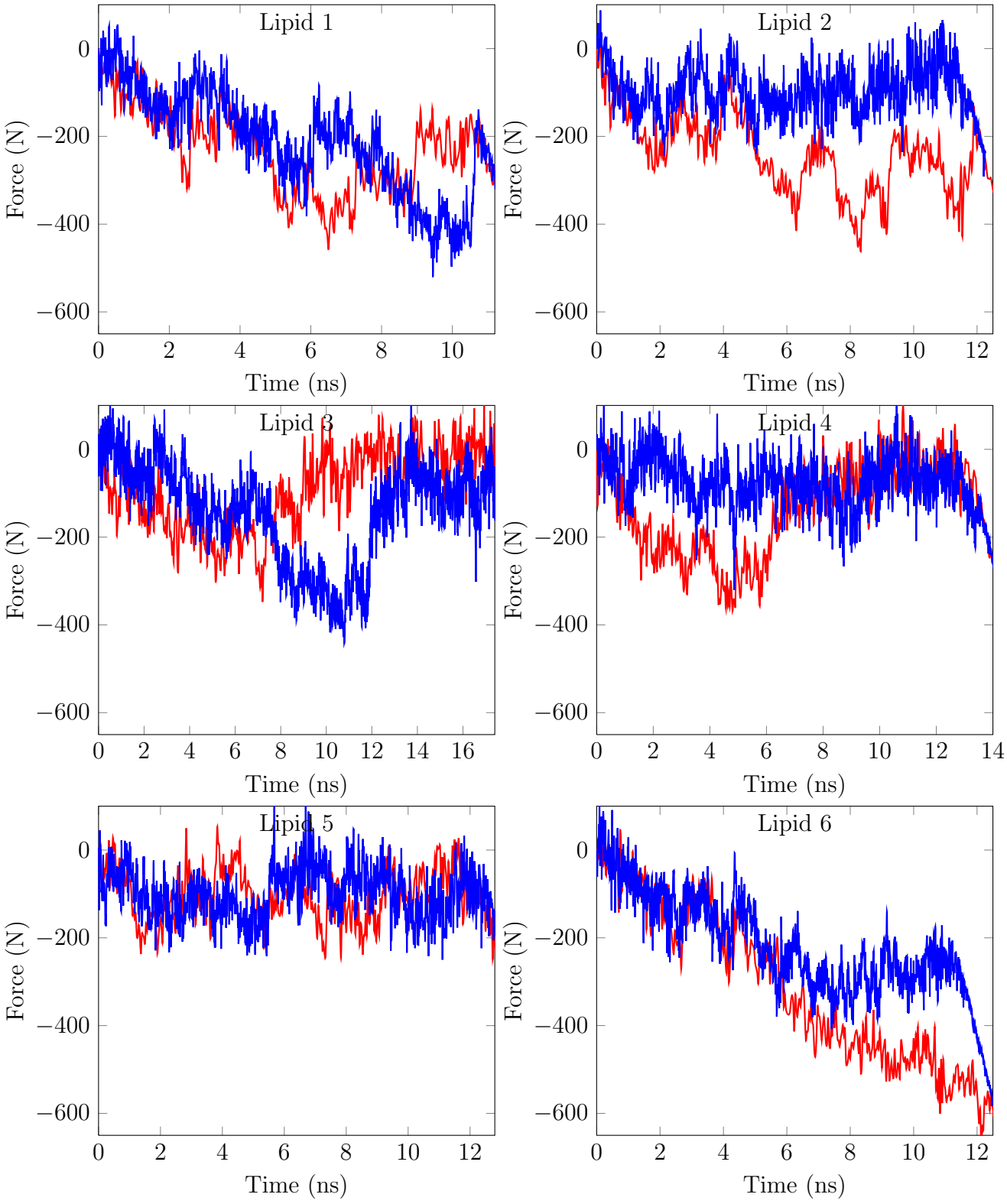

Figure 6.2: Pulling force acting on phospholipid headgroups during the six pulls towards rhodopsin center. The lipid numbering corresponds to the numbering in Fig. 6.1. The red curve corresponds to a pull rate of 1 nm/ns and the blue curve to a pull rate of 0.3 nm/ns. The time scale of the faster pull has been scaled by a factor of 10/3 for easier comparison with the slower pull rate.

interactions with two protein loop structures, where the jumps in the pulling force correspond to them. The second lipid shows a smooth trajectory with the slower pull rate, but the lipid reaches the protein center from above with the only remaining way across the membrane through the protein channel. The third lipid interacts quite strongly with the protein loop before entering the transmembrane channel from above. The jump in the pulling force corresponds to the moment the phospholipid is freed from its interaction with the loop. The fourth and fifth lipids also enter the protein interior from above rather than either of the two possible gaps in the nearby alpha-helices. The sixth lipid had to pass a loop structure in order to reach the protein center resulting in a constantly increasing pulling force.

No notable differences between the two pulling rates are seen in the graphs. The slower pull rate results in a much smoother path as expected, but in both cases, some pulling force was required to move the lipid instead of finding a process, where it would have traveled along a downhill path. The routes taken by the phospholipid were the same in both cases.

By examining the trajectories of the six POPC molecules and the graphs showing the applied pulling forces on each molecule, some of the routes were dismissed as highly unlikely due to odd trajectories or unnaturally high forces. The remaining routes showed many similarities and the trajectory leading to the widest gap was chosen as an example route. The selection was done as the pulling trajectory looked the most promising as the route was the only one where the phospholipid was pulled into the rhodopsin interior along a direct path and not through the water phase. Low pulling force thus did not necessarily indicate a viable route. The resulting trajectories indicate that the headgroup entering the transmembrane channel with the tails remaining in the membrane interior is highly unlikely.

Umbrella sampling was then conducted on the selected route with a total of 30 umbrella windows picked at 0.1 nm intervals from the original location of the POPC molecule to the center of rhodopsin. The umbrella windows were simulated for a total of 40 ns of which the first 20 ns were used for equilibration and discarded from the free energy calculations. The phospholipid was constrained to the middle of the umbrella window by a harmonic umbrella potential with a force constant of 1000 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ in both cases. The constraints were only applied along the reaction coordinate and the lipid was free to move in the other directions. The obtained free energy graph with its statistical error estimates obtained by `g_wham` is shown in Fig. 6.3. The red circle in the figure shows the estimated radius of rhodopsin, and the blue line the path along which the lipid was pulled to the rhodopsin interior.

The free energy curve becomes the steeper the closer to rhodopsin the POPC molecule gets, and no local minima can be seen at least on the pulling route in question. This is supported by MD simulations, which showed that a POPC

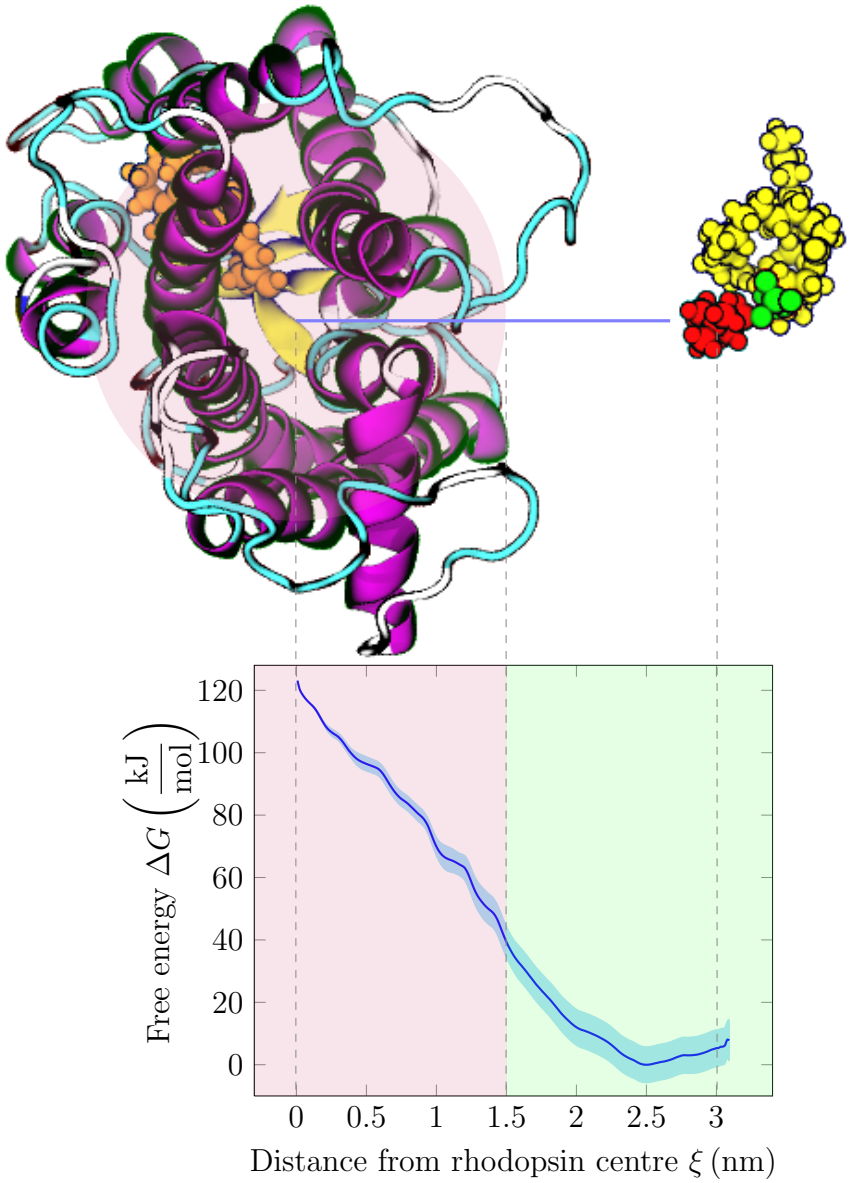

Figure 6.3: Free energy of POPC movement towards rhodopsin. The data was gathered from 30 umbrella windows along the route highlighted in Fig 6.1. and indicated by the blue line. The reaction coordinate ξ defines the distance from the center of rhodopsin.

molecule that was pulled close to the rhodopsin centre spontaneously traveled back to the bilayer–water interface once the restraints on it were removed. The radius of rhodopsin can be estimated to be about 1.5 nm, below which the required free energy increases at such a pace that it is very unlikely for the lipid to naturally reach the center of rhodopsin. The cut-off radius for long range interactions has been set at 1 nm, which can be seen from the relatively flat free energy profile at distances above 2 nm.

From the free energy graph it can be deduced that the flipping path of phospholipids most likely lies either on the surface of the protein or with the head group in the transmembrane channel. This is supported by MD simulations conducted on rhodopsin in our model membrane in which phospholipids were frequently found on

the surface of rhodopsin at a distance of about 1.5 nm from the center, but never in its interior. However, pulling simulations indicated that it is very difficult for the lipid headgroup to stay in the transmembrane channel. When the phospholipid was pulled through the membrane with pulling force centered on the lipid headgroup, the lipid either dived completely into the transmembrane channel or disengaged from the protein and remained on the surface. These simulation results were then used in determining subsequent directions taken in the thesis.

Despite the steep change of free energy close to the protein, the trajectories showed that phospholipids frequently visited the surface of rhodopsin during the course of the 200 ns simulation. The lateral diffusion of lipids occurs on a rapid scale and the average lipid can diffuse the length of a large cell (2 μm) in about one second [1]. Thus it is reasonable to expect that lipids can often be found on the surface of rhodopsin despite the 20–30 kJ/mol difference in free energy.

6.2 Free Energy of POPC Flip–flop in a Pure Bilayer

For comparison with the rhodopsin-mediated flip–flop process, a POPC molecule was pulled through a protein-free bilayer in order to obtain the free energy difference between the membrane center and a lipid leaflet. The pulling simulations were conducted by steered MD applied on the lipid headgroup along the z-axis. The POPC molecule was pulled through the bilayer at a pull rate of 0.5 nm/ns.

Umbrella sampling with a total of 21 umbrella windows at 0.2 nm intervals were used in obtaining the free energy curve. The umbrella windows were equilibrated for 10 ns before conducting a 15 ns simulation from which the data was obtained. The phospholipid was constrained in the middle of the window by a harmonic umbrella potential using force constants of 2000 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ for the equilibration period and 750 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ for rest of the simulation. The constraints on the lipid were only applied along the reaction coordinate. Because the two leaflets had the same lipid composition, it was assumed that the free energy curve is symmetric respective to the bilayer center and thus the free energy profile was mirrored at the middle. The obtained free energy curve and the statistical error estimates obtained by `g_wham` can be seen in Fig. 6.4.

The free energy profile looks similar to the ones obtained for pure bilayers with different compositions in previous studies [34, 35]. The PMF increases steeply as the phospholipid enters the membrane interior until a peak is reached in the middle of the membrane. The PMF also increases as the phospholipid is pulled towards the water phase although this is not shown in the figure. The height of the energy barrier for a protein-free bilayer seems to be at around 130 kJ/mol, which is high enough to induce a low intrinsic flip–flop rate.

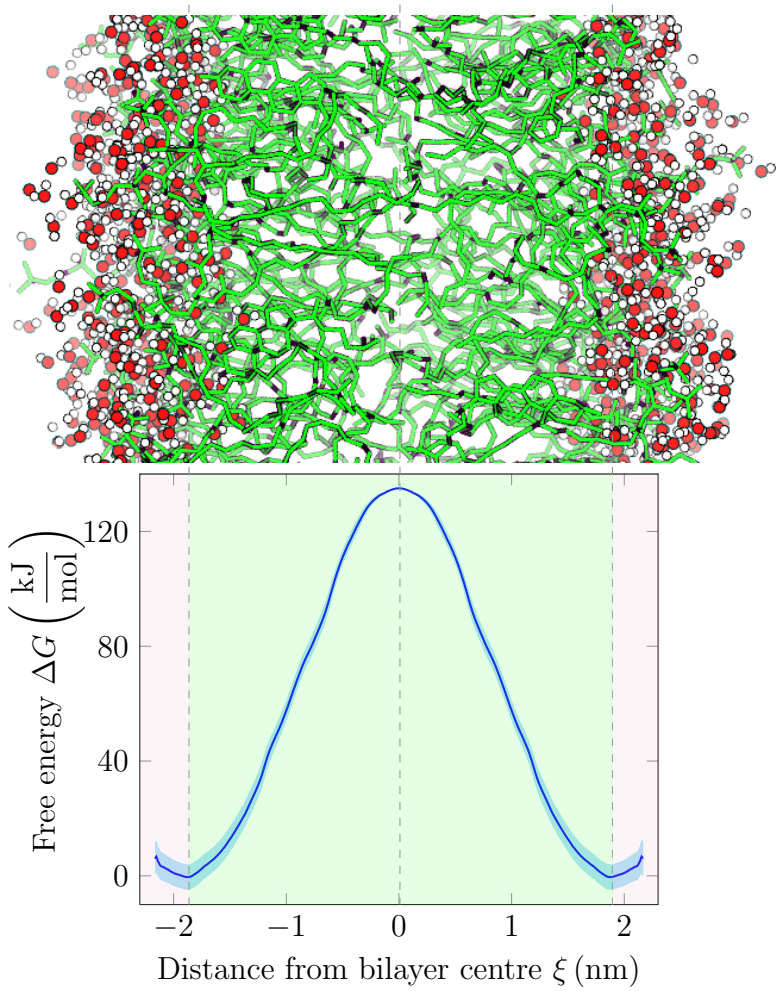

Figure 6.4: Free energy of phospholipid flip–flop in a protein-free membrane illustrated in the figure above the free energy graph. The reaction coordinate ξ defines the distance from the center of the membrane.

6.3 Free Energy of Rhodopsin-mediated Flip–flop

The main goal of this thesis was to determine the free energy of protein-mediated phospholipid translocation through which the flip–flop enhancing nature of rhodopsin can be deduced. The free energy calculations were conducted using umbrella sampling, which has been shown to be effective in previous studies [34, 35, 111] in researching lipid flip–flop in protein-free bilayers.

The results shown in the previous section indicated that the most likely flipping path is either along the surface of the protein or with the lipid completely enclosed by the protein. Trial simulations were conducted on both cases using short umbrella sampling windows. The path through the transmembrane channel showed an unnaturally high difference (of the order of hundreds of kJ/mol) in free energy after short (of the order of 10 ns) umbrella sampling simulations and was thus discarded. It is, however, entirely possible that the equilibration times with the lipid inside the

transmembrane channel should be of the order of hundreds of nanoseconds in which case large amounts of resources would be required to obtain reliable data. Such resources for further analysis were not currently available for us and thus the route along the protein surface was chosen for further study.

The free energy graph had a tendency to be uphill towards whichever direction the lipid was previously pulled, likely due to residue forces from the pulling or insufficient equilibration times. Therefore, in order to obtain more reliable data, the pulling was conducted towards both directions from several properly equilibrated umbrella windows along the original pulling path. Subsequent umbrella windows were then chosen in turn from these pulls in opposite directions with the idea of two neighboring windows canceling the residue forces of each other.

A total of 33 umbrella windows were used to obtain the free energy curve. These windows were chosen in a way that they were more densely packed in the middle of the membrane, where the changes in free energy are likely to be more sudden. The sparse sections also had sufficient overlap of umbrella windows. Each umbrella window was first equilibrated for 20 ns to allow the lipid to stabilize its position around the desired spot and to discard the effects that the changes in protein conformation might have on the free energy differences. After the equilibration period, a 20 ns simulation run was conducted from which the free energy curve was obtained. The force constant used for constraining the phospholipid in the middle of the membrane was set to 4000 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ during the equilibration period and to 1000 $\frac{\text{kJ}}{\text{mol}\times\text{nm}^2}$ during the simulation run. The constraints on the lipid were only applied along the reaction coordinate. The results of the free energy difference measurements and the statistical error obtained from `g_wham` are shown in Fig. 6.5 along with a figure of rhodopsin showing the surface along which the phospholipid traveled.

The observed flip–flop process occurs along the surface of rhodopsin along the route shown in Fig. 6.5. The corresponding lipid would thus be lipid number 5 in Fig 6.1. The selection was done due to high hydrophobicities of the nearby alpha-helices. The phospholipid is pulled into the membrane interior with tails first, and after a while it reorients itself in a way that its headgroup faces the rhodopsin while the tails float freely in the membrane interior. Finally the phospholipid enters the other leaflet, with its headgroup standing up. Minor interactions can be observed between the phospholipid tails and rhodopsin during the flipping process, causing the phospholipid tails to occasionally point towards the protein. Selected frames of the protein-mediated flip–flop process are shown in Fig. 3.2.

The polar headgroup faces a free energy barrier upon entering the membrane interior. The maximum of the barrier corresponds to the part of the flip–flop process where the phospholipid headgroup turns to face the other leaflet. The lipid then prefers to enter the second leaflet as can be seen from the free energy curve.

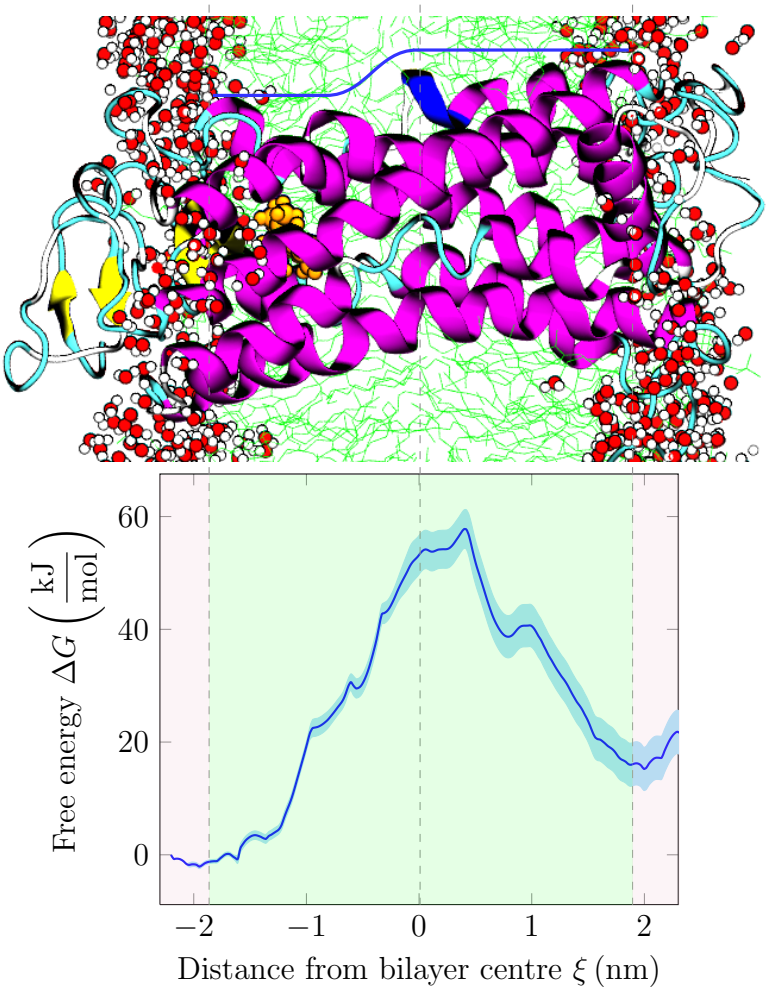

Figure 6.5: Free energy of phospholipid flip–flop along the surface of rhodopsin. The flipping route is indicated by the blue line above the figure of rhodopsin. The reaction coordinate ξ defines the distance from the center of the membrane along the z-axis.

The free energy barrier seems to be a little different for the two leaflets. Translocation from the cytosolic side of the membrane to the extracellular side seems to be more favorable for POPC than translocation in the opposite direction. This is likely due to the asymmetrical structure of rhodopsin, which might favor the outward flip–flop process. Some of the energy difference between the two water phases might also be due to some error in the simulation process.

The free energy curve has a peak near the middle of the membrane at around 45–60 kJ/mol (17–23 k_BT at 310 K temperature). Thermal fluctuations are of the order of several $k_\mathrm{B}T$ and thus are not enough to induce rhodopsin-mediated flip–flop alone. Since the amount of energy obtained from a single ATP molecule is 20 $k_\mathrm{B}T$, or 51.5 kJ/mol at 310 K temperature, the energy obtained from ATP coupled with thermal fluctuations is enough to push the phospholipid over the free energy barrier and from one leaflet into the other. Comparison with the result obtained for the pure bilayer in section 6.2 shows that the free energy barrier is lowered by over a half

due to the presence of rhodopsin, indicating that rhodopsin acts as a phospholipid flippase.

The height of the energy barrier can also be approximated mathematically by comparing the experimentally measured flip–flop times of a pure bilayer and a bilayer containing opsin. Tieleman et al. measured the free energy difference for a pure bilayer at around 80 kJ/mol using umbrella sampling and estimated the respective flip–flop times to be at around 10^5 seconds [34]. Respectively Menon et al. experimentally measured an upper bound of about 10 seconds for opsin-mediated flip–flop times [3]. A relation between the flip–flop rates k and the free energy barrier can be found from the Arrhenius equation

$$k = A \exp\left(-\frac{\Delta G}{RT}\right), \tag{6.1}$$

where A is the prefactor specific for a process, measured by the number of collisions leading to the process, and ΔG can be considered as the height of the free energy barrier [112].

Since the flip–flop rates are inversely proportional to the flip–flop times, we can compare the two Arrhenius equations for opsin-mediated flip–flop and flip–flop in a pure bilayer. By inserting the estimated flip–flop times, the free energy barrier for the pure bilayer measured by Tieleman et al., and the value of RT at the human body temperature of 310 K into Eq. 6.1, and by cancelling out the reaction specific prefactors, we get

$$\frac{10s}{10^5 s} = \exp\left(\frac{-80\text{kJ/mol} + \Delta G_{\text{opsin}}}{2.577\text{kJ/mol}}\right). \tag{6.2}$$

Solving the equation gives us $\Delta G_{\text{opsin}} \approx 56$ kJ/mol, which is a very crude approximation for the height of the energy barrier, but one that is very close to the values we obtained from umbrella sampling. However, the assumption that the prefactors for the two reactions are equal, is not quite correct. The lipid has to be located next to the rhodopsin surface in order for the rhodopsin-mediated reaction to occur making the respective prefactor smaller. A small difference in the prefactors, however, does not result in a high difference for ΔG, since a logarithm is taken from their ratio. The obtained energy from the calculation should thus be considered an upper limit for the height of the free energy barrier, but one that is very close to the real value.

The obtained results indicate that rhodopsin is a phospholipid flippase as it significantly lowers the free energy barrier of the membrane. The results from molecular dynamics simulations are backed up by the mathematical approximations. The examined flipping route was only one of many possible routes and thus additional information about the flipping process might be found by examining the other routes.

Also, small deviations in lipid positioning along the reaction path were observed to result in large differences in the free energy profile, which indicates that the translocation path with the lowest free energy difference can be extremely difficult to find. Thus, finding the flipping route with the lowest free energy profile requires extreme precision.

6.4 The Stability of the System

When simulating a system, an important question to be resolved is whether the system in hand is stable and thus realistic. The stability of the system can be resolved by analysing the deviation of the protein backbone and the root mean square fluctuations of the atoms in the protein. This was done using the GROMACS tools `g_rms` and `g_rmsf`, which calculate the root mean square (RMS) of protein backbone deviation and the root mean square fluctuation (RMSF) of protein residues during the course of the MD simulation. RMSD can thus be used to measure the amount of structural changes in the protein and RMSF to measure the amount of fluctuations in the positions of the amino acid residues. RMSD and RMSF of the protein were obtained from the first 200 ns of stabilizing simulations of the mixed composition bilayer with rhodopsin embedded in it.

The root mean square deviation is simply calculated as the deviation between the position of the particle $\mathbf{r}_i(t)$ and a reference position $\mathbf{r}_i^{\text{ref}}$ at each time step by the equation

$$\text{RMSD}(t) = \sqrt{\frac{1}{M}\sum_{i=1}^{N} m_i \|\mathbf{r}_i(t) - \mathbf{r}_i^{\text{ref}}\|^2}, \tag{6.3}$$

where M is the total mass of the protein, m_i the mass of the particle i, and N the total number of atoms in the protein [113]. The reference positions are typically taken from the starting structure of the protein as deviation from the original structure is usually desired.

As can be seen from the RMSD curve in Fig. 6.6, the original crystal structure of rhodopsin immediately begins to change once entered into the bilayer. The original crystal was grown by vapor diffusion using the hanging drop method at 283 K temperature and flash-frozen in liquid nitrogen to preserve the structure for x-ray diffraction [66]. Thus some amount of deviation from the original crystal structure is expected once the protein is entered into a bilayer system. The backbone deviation stabilizes quite rapidly and the protein structure can be considered stable by the end of the equilibration simulation. The leveling of the RMSD curve indicates that the used system is stable and the used protein and force field parameters correct. Incorrect parameters would likely have resulted in protein unfolding, which would clearly be seen in the RMSD curve.

The root mean square fluctuation of each residue is calculated as the deviation between the (center of mass) position of the residue $\mathbf{r}_i(t_j)$ and a reference position $\mathbf{r}_i^{\text{ref}}$ over a certain time t [87]. RMSF can be calculated via the equation

$$\text{RMSF}_i = \sqrt{\frac{1}{t}\sum_{t_j=1}^{t}\left(\mathbf{r}_i(t_j) - \mathbf{r}_i^{\text{ref}}\right)^2}, \tag{6.4}$$

where i corresponds to the the residue in question. The reference position is typically selected as the time-averaged position of the same particle.

The RMSF curve in Fig. 6.7 shows considerable fluctuations in certain amino acid residues. The residues in question are the loops located outside the bilayer, which are in direct contact with the water phase. Residue fluctuation occurs due to thermal movement of atoms, which is naturally higher in the water phase. In such cases, where flimsy loops are in contact with the water phase, substantial residue fluctuation is expected. No notable fluctuation can, however, be observed on the residues located inside the bilayer, which indicates that the retinal chromophore added to Lys296 had reasonable parameters and did not induce changes in the nearby residues. The lack of fluctuation could also indicate that no such conformational changes that could facilitate ATP-independent flip–flop occurred by the addition of the retinal. The observations by Menon et al. also indicated that the flippase activity of opsin does not require the 11-*cis*-retinal [3].

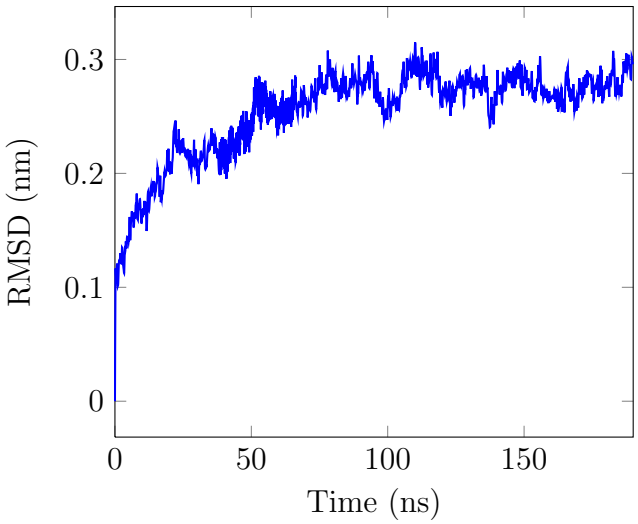

Figure 6.6: Root mean square deviation of rhodopsin backbone. The data was collected from approximately 200 ns of MD simulations of rhodopsin in the mixed composition bilayer. The deviation stabilizes around the middle of the simulation after which the system can be considered well-equilibrated.

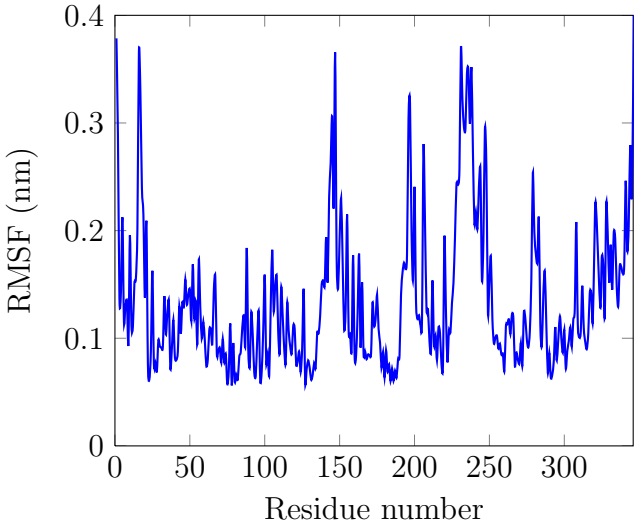

Figure 6.7: Root mean square fluctuation of the amino acid residues of rhodopsin. The data was collected from approximately 200 ns of MD simulations of rhodopsin in the mixed composition bilayer. All residues showing considerable fluctuations are loop structures in contact with the water phase where such behavior is expected.

7. CONCLUSIONS

The aim of this thesis was to study transbilayer lipid motion (flip–flop) facilitated by a transmembrane flippase protein with the use of molecular dynamics simulations. The MD simulations performed in this thesis were the first that investigated the free energy differences of lipid translocation in the presence of a suggested flippase protein. Previous MD simulations that utilise umbrella sampling to obtain free energy differences for phospholipid movement include lipid translocation in a protein-free membrane with various compositions, and lipid translocation in the presence of a transmembrane peptide chain.

The free energy graphs for phospholipid movement towards the protein and across the membrane were obtained by the use of umbrella sampling. The system was equilibrated for a total of 200 ns before further simulations were conducted. The RMSD graph of protein backbone indicated that the used force fields and parameters of the constructed system were appropriate. Each umbrella window was equilibrated for an additional 20 ns in order to stabilize the lipid position on the protein surface after which the umbrella sampling data was obtained from 20 ns simulation runs.

The investigated flipping route was selected to be on the surface of the protein after various unsuccessful steered MD attempts to hold the lipid headgroup in the transmembrane channel of rhodopsin while pulling the phospholipid across the membrane. The route in which the whole phospholipid is located in the transmembrane channel was discarded due to its long equilibration time requirement.

The results showed a free energy barrier of about 45–60 kJ/mol for the rhodopsin-mediated flip–flop process, which is considerably lower than the 130 kJ/mol barrier observed for a protein-free membrane. The results thus confirm the flippase properties of rhodopsin as a phospholipid translocator across the membrane. No significant conformational changes were observed in the rhodopsin structure before the flipping process.

Further studies could be focused on the two translocation routes disregarded in this thesis, as further proof of the scramblase activity of rhodopsin might be found there. The conserved waters found in the transmembrane channel of rhodopsin could prove to be a favorable translocation route for the hydrophilic PC headgroup, but large amounts of computational resources will probably be needed to sufficiently equilibrate each umbrella window with the lipid in the transmembrane channel.

In addition, the potential role of Ca^{2+}-ions in the flipping process could be investigated. A high concentration of Ca^{2+}-ions in the cytosol could induce conformational changes in the rhodopsin structure near a calcium-binding site. This in turn might open up a suitable route for flip–flop with a low free energy barrier.

This thesis shows that umbrella sampling is a viable way to measure free energy differences for phospholipid translocation in the presence of a transmembrane protein. It is also shown that steered molecular dynamics simulations are an efficient way to examine the possible flipping routes of phospholipids and the flippase mechanisms of transmembrane proteins. Molecular dynamics could thus be used to identify new flippase proteins once educated guesses of their identities have been made. All in all, this thesis shows that the use of molecular dynamics works efficiently on the topic at hand making it an excellent method of further pursuing the flip–flop phenomenon.

BIBLIOGRAPHY

[1] **Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P.**, *Molecular biology of the cell*. Garland Science, New York, USA, 2008.

[2] **Sanyal, S. and Menon, A.**, Flipping lipids: why and what's the reason for? *ACS Chemical Biology*, 4(11): 895–909, 2009.

[3] **Menon, I., Huber, T., Sanyal, S., Banerjee, S., Barré, P., Canis, S., Warren, J., Hwa, J., Sakmar, T., and Menon, A.**, Opsin is a phospholipid flippase. *Current Biology*, 21(2): 149–153, 2011.

[4] **Graham, T.**, Flippases and vesicle-mediated protein transport. *Trends in Cell Biology*, 14(12): 670–677, 2004.

[5] **Pomorski, T., Holthuis, J., Herrmann, A., and van Meer, G.**, Tracking down lipid flippases and their biological functions. *Journal of Cell Science*, 117(6): 805–813, 2004.

[6] **Sharom, F.**, Flipping and flopping – lipids on the move. *IUBMB Life*, 63(9): 736–746, 2011.

[7] **Kol, M., de Kruijff, B., and de Kroon, A.**, Phospholipid flip-flop in biogenic membranes: what is needed to connect opposite sides. *Seminars in Cell & Developmental Biology*, 13(3): 163–170, 2002.

[8] **Kol, M., de Kroon, A., Killian, J., and de Kruijff, B.**, Transbilayer movement of phospholipids in biogenic membranes. *Biochemistry*, 43(10): 2673–2681, 2004.

[9] **Williamson, P.**, Phospholipid transport: sighting a new face of an old friend. *Current Biology*, 21(4): 168–169, 2011.

[10] **Hart, H., Hart, D., Craine, L., and Hadad, C.**, *Organic chemistry: a short course*. Houghton Mifflin Company, New York, USA, 2011.

[11] **Lodish, H., Berk, A., Zipursky, S., Matsudaira, P., Baltimore, D., and Darnell, J.**, *Molecular cell biology*. W.H. Freeman and Company, New York, USA, 2003.

[12] **Lehninger, A., Nelson, D., and Cox, M.**, *Principles of biochemistry*. W.H. Freeman and Company, New York, USA, 2008.

[13] **Heino, J. and Vuento, M.**, *Biokemian ja solubiologian perusteet*. WSOY Oppimateriaalit, Helsinki, 2007.

[14] **Cooper, G. and Hausman, R.**, *The cell – a molecular approach*. Sinauer Associates, Inc., Sunderland, USA, 2000.

[15] **Niemelä, P., Hyvönen, M., and Vattulainen, I.**, Structure and dynamics of sphingomyelin bilayer: insight gained through systematic comparison to phosphatidylcholine. *Biophysical Journal*, 87(5): 2976–2989, 2004.

[16] **Hannun, Y. and Obeid, L.**, Principles of bioactive lipid signalling: lessons from sphingolipids. *Nature Reviews: Molecular Cell Biology*, 9(2): 139–150, 2008.

[17] **De Meyer, F. and Smit, B.**, Effect of cholesterol on the structure of a phospholipid bilayer. *Proceedings of the National Academy of Sciences*, 106(10): 3654–3658, 2009.

[18] **Slotte, J.**, Sphingomyelin–cholesterol interactions in biological and model membranes. *Chemistry and Physics of Lipids*, 102(1–2): 13–27, 1999.

[19] **Nelson, P., Radosavljevic, M., and Bromberg, S.**, *Biological physics*. W.H. Freeman and Company, New York, USA, 2004.

[20] **Bretscher, M.**, Asymmetrical lipid bilayer structure for biological membranes. *Nature*, 236(61): 11–12, 1972.

[21] **Contreras, F., Sánchez-Magraner, L., Alonso, A., and Goñi, F.**, Transbilayer (flip – flop) lipid motion and lipid scrambling in membranes. *FEBS Letters*, 584(9): 1779–1786, 2010.

[22] **Alberts, B.**, Encyclopedia Britannica Online: the cell membrane, Available at `http://www.britannica.com`. Accessed on September 24, 2012.

[23] **Spence, M. and Johnson, I.**, *The molecular probes handbook: a guide to fluorescent probes and labeling technologies*. Life Technologies Corporation, Paisley, UK, 2010.

[24] **Balasubramanian, K. and Schroit, A.**, Aminophospholipid asymmetry: a matter of life and death. *Annual Review of Physiology*, 65(1): 701–734, 2003.

[25] **Falck, E., Patra, M., Karttunen, M., Hyvönen, M., and Vattulainen, I.**, Lessons of slicing membranes: interplay of packing, free area, and lateral diffusion in phospholipid/cholesterol bilayers. *Biophysical Journal*, 87(2): 1076–1091, 2004.

[26] **Korlach, J., Schwille, P., Webb, W., and Feigenson, G.**, Characterization of lipid bilayer phases by confocal microscopy and fluorescence correlation spectroscopy. *Proceedings of the National Academy of Sciences*, 96(15): 8461–8466, 1999.

[27] **Lingwood, D. and Simons, K.**, Lipid rafts as a membrane-organizing principle. *Science*, 327(5961): 46–50, 2010.

[28] **Ball, D., Hill, J., and Scott, R.**, *The basics of general, organic, and biological chemistry*. Barnes & Noble, Inc., New York, USA, 2011.

[29] **Branden, C. and Tooze, J.**, *Introduction to protein structure*. Garland, New York, USA, 1999.

[30] **Schlick, T.**, *Molecular modeling and simulation: an interdisciplinary guide*. Springer, New York, USA, 2010.

[31] **Rose, G., Glerasch, L., and Smith, J.**, Turns in peptides and proteins. *Advances in Protein Chemistry*, 37(1): 1–109, 1985.

[32] **Karp, G.**, *Cell and molecular biology: concepts and experiments*. John Wiley & Sons, New York, USA, 2009.

[33] **Forrest, L., Krämer, R., and Ziegler, C.**, The structural basis of secondary active transport mechanisms. *Biochimica et Biophysica Acta (BBA) – Bioenergetics*, 1807(2): 167–188, 2011.

[34] **Tieleman, D. and Marrink, S.**, Lipids out of equilibrium: Energetics of desorption and pore mediated flip-flop. *Journal of the American Chemical Society*, 128(38): 12462–12467, 2006.

[35] **Bennett, W., MacCallum, J., Hinner, M., Marrink, S., and Tieleman, D.**, Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *Journal of the American Chemical Society*, 131(35): 12714–12720, 2009.

[36] **Gurtovenko, A., Anwar, J., and Vattulainen, I.**, Defect-mediated trafficking across cell membranes: insights from in silico modeling. *Chemical Reviews*, 110(10): 6077–6103, 2010.

[37] **Gurtovenko, A. and Vattulainen, I.**, Molecular mechanism for lipid flip-flops. *The Journal of Physical Chemistry B*, 111(48): 13554–13559, 2007.

[38] **Kol, M., de Kroon, A., Rijkers, D., Killian, J., and de Kruijff, B.**, Membrane-spanning peptides induce phospholipid flop: a model for phospholipid translocation across the inner membrane of E. coli. *Biochemistry*, 40(35): 10500–10506, 2001.

[39] **John, K., Schreiber, S., Kubelt, J., Herrmann, A., and Müller, P.**, Transbilayer movement of phospholipids at the main phase transition of lipid membranes: implications for rapid flip–flop in biological membranes. *Biophysical Journal*, 83(6): 3315–3323, 2002.

[40] **Chojnacki, T. and Dallner, G.**, The biological role of dolichol. *Biochemical Journal*, 251(1): 1–9, 1988.

[41] **Paulusma, C. and Oude Elferink, R.**, P4 ATPases – the physiological relevance of lipid flipping transporters. *FEBS Letters*, 584(13): 2708–2716, 2010.

[42] **Paterson, J., Renkema, K., Burden, L., Halleck, M., Schlegel, R., Williamson, P., and Daleke, D.**, Lipid specific activation of the murine P4-ATPase Atp8a1 (ATPase II). *Biochemistry*, 45(16): 5367–5376, 2006.

[43] **Coleman, J., Kwok, M., and Molday, R.**, Localization, purification, and functional reconstitution of the P4-ATPase Atp8a2, a phosphatidylserine flippase in photoreceptor disc membranes. *Journal of Biological Chemistry*, 284(47): 32670–32679, 2009.

[44] **Muthusamy, B., Natarajan, P., Zhou, X., and Graham, T.**, Linking phospholipid flippases to vesicle-mediated protein transport. *Biochimica et Biophysica Acta (BBA) – Molecular and Cell Biology of Lipids*, 1791(7): 612–619, 2009.

[45] **Gomès, E., Jakobsen, M., Axelsen, K., Geisler, M., and Palmgren, M.**, Chilling tolerance in arabidopsis involves ALA1, a member of a new family of putative aminophospholipid translocases. *The Plant Cell Online*, 12(12): 2441–2453, 2000.

[46] **Nagao, K., Kimura, Y., Mastuo, M., and Ueda, K.**, Lipid outward translocation by ABC proteins. *FEBS Letters*, 584(13): 2717–2723, 2010.

[47] **Tsybovsky, Y., Molday, R., and Palczewski, K.**, The ATP-binding cassette transporter ABCA4: structural and functional properties and role in retinal disease. *Inflammation and Retinal Disease: Complement Biology and Pathology*, 703(1): 105–125, 2010.

[48] **Oude Elferink, R. and Paulusma, C.**, Function and pathophysiological importance of ABCB4 (MDR3 P-glycoprotein). *Pflügers Archiv – European Journal of Physiology*, 453(5): 601–610, 2007.

[49] **Eckford, P. and Sharom, F.**, The reconstituted P-glycoprotein multidrug transporter is a flippase for glucosylceramide and other simple glycosphingolipids. *Biochemical Journal*, 389(2): 517–526, 2005.

[50] **Huang, Z., Chang, X., Riordan, J., and Huang, Y.**, Fluorescent modified phosphatidylcholine floppase activity of reconstituted multidrug resistance-associated protein MRP1. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1660(1): 155–163, 2004.

[51] **Watkins, W. and Menon, A.**, Reconstitution of phospholipid flippase activity from E. coli inner membrane: a test of the protein translocon as a candidate flippase. *Biological Chemistry*, 383(9): 1435–1440, 2002.

[52] **Eckford, P. and Sharom, F.**, The reconstituted Escherichia coli MsbA protein displays lipid flippase activity. *Biochemical Journal*, 429(1): 195–203, 2010.

[53] **van Meer, G., Voelker, D., and Feigenson, G.**, Membrane lipids: where they are and how they behave. *Nature Reviews: Molecular Cell Biology*, 9(2): 112–124, 2008.

[54] **Bell, R., Ballas, L., and Coleman, R.**, Lipid topogenesis. *Journal of Lipid Research*, 22(3): 391–403, 1981.

[55] **Sheetz, M. and Singer, S.**, Biological membranes as bilayer couples. A molecular mechanism of drug-erythrocyte interactions. *Proceedings of the National Academy of Sciences*, 71(11): 4457–4461, 1974.

[56] **Lentz, B.**, Exposure of platelet membrane phosphatidylserine regulates blood coagulation. *Progress in Lipid Research*, 42(5): 423–438, 2003.

[57] **Devaux, P., Herrmann, A., Ohlwein, N., and Kozlov, M.**, How lipid flippases can modulate membrane structure. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1778(7): 1591–1600, 2008.

[58] **Williamson, P., Kulick, A., Zachowski, A., Schlegel, R.A., and Devaux, P.F.**, Calcium induces transbilayer redistribution of all major phospholipids in human erythrocytes. *Biochemistry*, 31(27): 6355–6360, 1992.

[59] **Wiedmer, T., Zhou, Q., Kwoh, D., and Sims, P.**, Identification of three new members of the phospholipid scramblase gene family. *Biochimica et Biophysica Acta (BBA) – Biomembranes*, 1467(1): 244–253, 2000.

[60] **Sahu, S., Gummadi, S., Manoj, N., and Aradhyam, G.**, Phospholipid scramblases: an overview. *Archives of Biochemistry and Biophysics*, 462(1): 103–114, 2007.

[61] **Sims, P. and Wiedmer, T.**, Unraveling the mysteries of phospholipid scrambling. *Thrombosis and Haemostasis*, 86(1): 266–275, 2001.

[62] **Stout, J., Zhou, Q., Wiedmer, T., and Sims, P.**, Change in conformation of plasma membrane phospholipid scramblase induced by occupancy of its Ca^{2+} binding site. *Biochemistry*, 37(42): 14860–14866, 1998.

[63] **Sahu, S., Aradhyam, G., and Gummadi, S.**, Calcium binding studies of peptides of human phospholipid scramblases 1 to 4 suggest that scramblases are new class of calcium binding proteins in the cell. *Biochimica et Biophysica Acta (BBA) – General Subjects*, 1790(10): 1274–1281, 2009.

[64] **Orrenius, S., Zhivotovsky, B., and Nicotera, P.**, Regulation of cell death: the calcium-apoptosis link. *Nature Reviews: Molecular Cell Biology*, 4(7): 552–565, 2003.

[65] **Liu, J., Dai, Q., Chen, J., Durrant, D., Freeman, A., Liu, T., Grossman, D., and Lee, R.**, Phospholipid scramblase 3 controls mitochondrial structure, function, and apoptotic response. *Molecular Cancer Research*, 1(12): 892–902, 2003.

[66] **Okada, T., Sugihara, M., Bondar, A., Elstner, M., Entel, P., and Buss, V.**, The retinal conformation and its environment in rhodopsin in light of a new 2.2 Å crystal structure. *Journal of Molecular Biology*, 342(2): 571–583, 2004.

[67] **Palczewski, K., Kumasaka, T., Hori, T., Behnke, C., Motoshima, H., Fox, B., Trong, I., Teller, D., Okada, T., and Stenkamp, R.**, Crystal structure of rhodopsin: A G protein-coupled receptor. *Science Signalling*, 289(5480): 739–745, 2000.

[68] **Angel, T., Chance, M., and Palczewski, K.**, Conserved waters mediate structural and functional activation of family A (rhodopsin-like) G protein-coupled receptors. *Proceedings of the National Academy of Sciences*, 106(21): 8555–8560, 2009.

[69] **Kristiansen, K.**, Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacology & Therapeutics*, 103(1): 21–80, 2004.

[70] **Humphrey, W., Dalke, A., and Schulten, K.**, VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14(1): 33–38, 1996.

[71] **Stone, J.**, *An efficient library for parallel ray tracing and animation*. Master of Science Thesis. Computer Science Department, University of Missouri–Rolla. 1998.

[72] **Zhao, J., Zhou, Q., Wiedmer, T., and Sims, P.**, Palmitoylation of phospholipid scramblase is required for normal function in promoting Ca^{2+}-activated transbilayer movement of membrane phospholipids. *Biochemistry*, 37(18): 6361–6366, 1998.

[73] **Olausson, B., Grossfield, A., Pitman, M., Brown, M., Feller, S., and Vogel, A.**, Molecular dynamics simulations reveal specific interactions of post-translational palmitoyl modifications with rhodopsin in membranes. *Journal of the American Chemical Society*, 134(9): 4324–4331, 2012.

[74] **Wiedmer, T., Zhao, J., Nanjundan, M., and Sims, P.**, Palmitoylation of phospholipid scramblase 1 controls its distribution between nucleus and plasma membrane. *Biochemistry*, 42(5): 1227–1233, 2003.

[75] **Agarwal, N., Nir, I., and Papermaster, D.**, Opsin synthesis and mRNA levels in dystrophic retinas devoid of outer segments in retinal degeneration slow (rds) mice. *The Journal of Neuroscience*, 10(10): 3275–3285, 1990.

[76] **Hubbard, R. and Wald, G.**, The mechanism of rhodopsin synthesis. *Proceedings of the National Academy of Sciences of the United States of America*, 37(2): 69–78, 1951.

[77] **Wald, G. and Steven, D.**, An experiment in human vitamin A -deficiency. *Proceedings of the National Academy of Sciences of the United States of America*, 25(7): 344–349, 1939.

[78] **Kerr, J., Wyllie, A., and Currie, A.**, Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British Journal of Cancer*, 26(4): 239–257, 1972.

[79] **Smrž, D., Lebduška, P., Dráberová, L., Korb, J., and Dráber, P.**, Engagement of phospholipid scramblase 1 in activated cells. *Journal of Biological Chemistry*, 283(16): 10904–10918, 2008.

[80] **Robertson, J., Fadeel, B., Zhivotovsky, B., and Orrenius, S.**, 'Centennial' Nobel Conference on apoptosis and human disease. *Cell Death and Differentiation*, 9(4): 468–475, 2002.

[81] **Fadeel, B., Gleiss, B., Högstrand, K., Chandra, J., Wiedmer, T., Sims, P., Henter, J., Orrenius, S., and Samali, A.**, Phosphatidylserine exposure during apoptosis is a cell-type-specific event and does not correlate with plasma membrane phospholipid scramblase expression. *Biochemical and Biophysical Research Communications*, 266(2): 504–511, 1999.

[82] **Lowe, S. and Lin, A.**, Apoptosis in cancer. *Carcinogenesis*, 21(3): 485–495, 2000.

[83] **Ryan, K., Phillips, A., and Vousden, K.**, Regulation and function of the p53 tumor suppressor protein. *Current Opinion in Cell Biology*, 13(3): 332–337, 2001.

[84] **Barinaga, M.**, Is apoptosis key in Alzheimer's disease? *Science*, 281(5381): 1303–1304, 1998.

[85] **van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., and Berendsen, H.**, GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, 26(16): 1701–1718, 2005.

[86] **Hess, B., Kutzner, C., van der Spoel, D., and Lindahl, E.**, GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3): 435–447, 2008.

[87] **van der Spoel, D., Lindahl, E., Hess, B., van Buuren, A., Apol, E., Meulenhoff, P., Tieleman, D., Sijbers, A., Feenstra, K., and van Drunen, R.**, Gromacs user manual version 4.5, 2010.

[88] **Hsin, J., Arkhipov, A., Yin, Y., Stone, J., and Schulten, K.**, Using VMD: an introductory tutorial. *Current Protocols in Bioinformatics*, 24(5): 1–48, 2008.

[89] **van Gunsteren, W. and Berendsen, H.**, A leap-frog algorithm for stochastic dynamics. *Molecular Simulation*, 1(3): 173–185, 1988.

[90] **Jorgensen, W., Maxwell, D., and Tirado-Rives, J.**, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45): 11225–11236, 1996.

[91] **Lim, T.**, The relationship between Lennard-Jones (12-6) and Morse potential functions. *Z. Naturforsch*, 58a(11): 615–617, 2003.

[92] **Toukmaji, A. and Board, J.**, Ewald summation techniques in perspective: a survey. *Computer Physics Communications*, 95(2): 73–92, 1996.

[93] **Darden, T., York, D., and Pedersen, L.**, Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*, 98(12): 10089–10092, 1993.

[94] **Berendsen, H., Postma, J., van Gunsteren, W., DiNola, A., and Haak, J.**, Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8): 3684–3690, 1984.

[95] **Morishita, T.**, Fluctuation formulas in molecular dynamics simulations with the weak coupling heat bath. *The Journal of Chemical Physics*, 113(8): 2976–2982, 2000.

[96] **Bussi, G., Donadio, D., and Parrinello, M.**, Canonical sampling through velocity-rescaling. *The Journal of Chemical Physics*, 126(1): 1–7, 2007.

[97] **Parrinello, M. and Rahman, A.**, Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12): 7182–7190, 1981.

[98] **Hess, B., Bekker, H., Berendsen, H., and Fraaije, J.**, LINCS: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12): 1463–1472, 1997.

[99] **Ryckaert, J., Ciccotti, G., and Berendsen, H.**, Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of *n*-alkanes. *Journal of Computational Physics*, 23(3): 327–341, 1977.

[100] **Miyamoto, S. and Kollman, P.**, SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8): 952–962, 1992.

[101] **Kästner, J.**, Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6): 932–942, 2011.

[102] **Frenkel, D. and Smit, B.**, *Understanding molecular simulation: from algorithms to applications*. Academic Press, Orlando, USA, 2002.

[103] **Jarzynski, C.**, Nonequilibrium equality for free energy differences. *Physical Review Letters*, 78(14): 2690–2693, 1997.

[104] **Lemkul, J.**, GROMACS tutorials: umbrella sampling. Available at `http://www.bevanlab.biochem.vt.edu/Pages/Personal/justin/gmx-tutorials/`. Accessed on October 17, 2012.

[105] **Kumar, S., Rosenberg, J., Bouzida, D., Swendsen, R., and Kollman, P.**, The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, 13(8): 1011–1021, 2004.

[106] **Hub, J., De Groot, B., and van Der Spoel, D.**, g_wham – a free weighted histogram analysis implementation including robust error and autocorrelation estimates. *Journal of Chemical Theory and Computation*, 6(12): 3713–3720, 2010.

[107] **Dupradeau, F., Cézard, C., Lelong, R., Stanislawiak, É., Pêcher, J., Delepine, J., and Cieplak, P.**, RE DD. B.: a database for RESP and ESP atomic charges, and force field libraries. *Nucleic Acids Research*, 36(suppl 1): D360–D367, 2008.

[108] **Moyna, G., Hernandez, G., Williams, H.J., Nachman, R.J., and Scott, A.I.**, Development of Weiner et al. force field parameters suitable for conformational studies of [1,4]-benzodiazepines and related compounds. *Journal of Chemical Information and Computer Sciences*, 37(5): 951–956, 1997.

[109] **Nagy, P.I., Tejada, F.R., Sarver, J.G., and Messer, W.S.**, Conformational analysis and derivation of molecular mechanics parameters for esters and thioesters. *The Journal of Physical Chemistry A*, 108(46): 10173–10185, 2004.

[110] **Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R., and Klein, M.**, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2): 926–935, 1983.

[111] **Sapay, N., Bennett, W., and Tieleman, D.**, Molecular simulations of lipid flip-flop in the presence of model transmembrane helices. *Biochemistry*, 49(35): 7665–7673, 2010.

[112] **Zumdahl, S.**, *Chemical principles*. Houghton Mifflin Company, New York, USA, 2002.

[113] **Maiorov, V. and Crippen, G.**, Size-independent comparison of protein three-dimensional structures. *Proteins: Structure, Function, and Bioinformatics*, 22(3): 273–283, 2004.