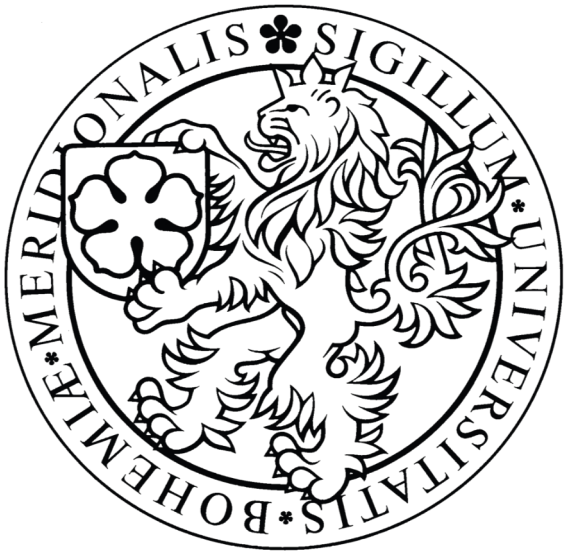
University of South Bohemia in České Budějovice
Faculty of Science



Molecular dynamics as a tool to study biological systems

Ph.D. Thesis

Mgr. Žofie Sovová

supervisor: doc. RNDr. Rüdiger H. Ettrich, Ph.D.

Institute of Nanobiology and Structural Biology of GCRC, Academy of Sciences of the Czech Republic
Faculty of Science, University of South Bohemia in České Budějovice

České Budějovice 2013

This thesis should be cited as:
Sovová, Ž., 2013: Molecular dynamics as a tool to study biological systems. University of South Bohemia, Faculty of Science, České Budějovice, Czech Republic, 94 pp.

Annotation
Molecular dynamics simulations are a theoretical method enabling to trace the movement of atoms within a system. The system studied is usually treated on the atomistic level, however its overall properties can be also described satisfactory if several atoms are handled as one particle (coarse-grained molecular dynamics).
This thesis presents molecular modeling and (coarse-grained) molecular dynamics as tools for the description of different biologically relevant systems. The coarse-grained force field parameters had to be developed prior to characterization of the thylakoid membrane from cyanobacterium *Synechocystis PCC6803*. Two different compositions of the membrane were studied in order to reveal differences in their behavior. The PsbI subunit of photosystem II was embedded into the thylakoid membrane and its behavior, both as an isolated protein and as a cluster of several units, was described.
The last system examined was the C-type lectin-like domain of NKR-P1, a surface receptor of natural killer cells. Attention was payed to its structural characterization.

Declaration [in Czech]
Prohlašuji, že svoji disertační práci jsem vypracovala samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v úpravě vzniklé vypuštěním vyznačených částí archivovaných Přírodovědeckou fakultou elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 15. 2. 2013 .....

Žofie Sovová

This thesis originates from a partnership of the Faculty of Science, University of South Bohemia and the Institute of Nanobiology and Structural Biology of the Global Change Research Center, Academy of Sciences of the Czech Republic

Financial support
This research was supported by HPC-EUROPA2 (project number 228398) with support from European Commission Capacity Area – Research Infrastructure Initiative, the University of South Bohemia (GAJU 170/2010/P) and the Ministry of Education of the Czech Republic (ME09062, Aktion64p1). Calculations were performed in the METACentrum computing facility provided under research intent MSM6383917201.

Děkuji všem, díky nimž jsem se dostala tam, kde jsem

Ďakujem vsedkým, vďaka ktorým som se dostala tam, kde som

Dank aan degenen die me brachten waar ik nu ben

List of papers and author's contribution

The presented thesis is based on following papers (listed chronologically)

Sovova Z., Kopecky jr. V., Pazderka T., Hofbauerova K., Rozbesky D., Vanek D., Bezouska K., Ettrich R. (2011): Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop involved in ligand specificity. J Mol Model 17; 1353 – 1370 (*IF: 1.8*)

Žofie Sovová performed the phylogenetic (incl. experiment design) and sequence analysis and designed and analyzed structural models. She participated in paper writing.

Kopecky jr V., Kohoutova J., Lapkouski M., Hofbauerova K., **Sovova Z.**, Ettrichova O., Gonzalez-Perez S., Dulebo A., Kaftan D., Kuta Smatanova I., Revuelta J.B., Arellano J.B., Carey J., Ettrich R. (2012): Raman spectroscopy adds complementary detail to the high-resolution X-ray crystal structure of photosynthetic PsbP from *Spinacia oleracea*. Plos One 7; e46694 (*IF: 4.1*)

Žofie Sovová performed homology modeling and molecular dynamics of both the proteins and analyzed its trajectories. She participated in paper writing.

Rozbesky D., **Sovova Z.**, Marcoux J., Man P., Ettrich P., Robinson C.V., Novak P. (2013): Structural model of lymphocyte receptor NKR-P1C revealed by mass spectrometry and molecular modeling. Anal Chem, 85: 3. 1597–1604 (IF: 5.9)

Žofie Sovová participated in modeling and analysis of the model of the receptor and in paper writing.

Lopez C.A., **Sovova Z.**, van Eerden F.J., de Vries A.H., Marrink S.J. (2013): Martini force field parameters for glycolipids. JCTC, accepted doi: 10.1021/ct3009655 (*IF: 5.2*)

Žofie Sovová developed the alternative model of force field parameters for thylakoid glycolipids.

Sovova Z., Ettrich R (2013): Characterization of the thylakoid membrane from cyanobacterium *Synechocystis PCC6803* by the means of united-atom and coarse-grained molecular dynamics. To be submitted

Žofie Sovová performed, designed and analyzed molecular dynamics simulations. She participated in the paper writing.

Sovova Z., Ettrich R. (2013): PsbI protein from photosystem II and its behaviour in the tylakoid membrane – a theoretical study. To be submitted

Žofie Sovová performed, designed and analyzed molecular dynamics simulations. She participated in the paper writing.

Contents

List of abbreviations

APL	area per lipid
CG	coarse-grain
COM	center of mass
CTLD	C-type lectin-like domain
DGDG	digalactosyl diacylglycerol
DP	dipalmitoyl
DS	distearoyl
LHCII	Light-harvesting complex II
LJ	Lennard-Jones
LP	sn-1-γ-linolenyl-sn-2-palmitoyl
lP	sn-1-linoleyl-sn-2-palmitoyl
LQ	liquid crystalline
MD	molecular dynamics
MGDG	monogalactosyl diacylglycerol
MSD	mean square displacement
NK	natural killer
NMR	nuclear magnetic resonance
PA	phosphatidic acid
PC	phosphatidyl choline
PE	phosphatidyl ethanolamine
PG	phosphatidyl glycerol
PI	phosphatidyl inositol
PME	particle mesh Ewald
PS	phosphatidyl serine
PSII	photosystem II
RDF	radial distribution function
RMSD	root mean square deviation
Sakurai's membrane	37% LPMG, 20% LPDG, 29% DPSQ, 14 lPPG
SQDG	sulfoquinovosyl diacylglycerol
Synechocystis	Synechocystis PCC6803
Wada's membrane	59% LPMG, 17% LPDG, 16% DPSQ, 8% lPPG

1.1 Molecular dynamics simulations

Molecular dynamics (MD) simulations are a theoretical method of statistical physics that reconstructs the mechanical motion of molecules.

It is based on solving the equation of motion for every particle of the system at every – discrete – time (for review see e.g. [Allen and Tildesley, 1987; van Gunsteren et al. 2006]). Force *F* acting onto a given particle can be according to the Newton's second law [Newton, 1687], written as:

$$\vec{F} = m \cdot \vec{a} = m \ddot{\vec{r}} \quad ,$$

where *m* stands for mass and *a* (that is the second derivative of the position vector) is acceleration.

The other way how to define a force was introduced by Lagrange and states that force is a negative gradient of the potential energy *U* of a system:

$$\vec{F} = -\nabla U(\vec{r}) \quad ,$$

The potential energy on the right side of the equation, that is in the field of MD simulations called a force field, is described in more detail in the next chapter.

Merging these two equation together we will get:

$$m\vec{a} = -\nabla U(\vec{r}) \quad ,$$

from which the acceleration can be expressed as:

$$\vec{a} = \frac{-\nabla U(\vec{r})}{m} \quad .$$

Knowing the position $\vec{r}(t)$ of a particle at a time *t*, and the position $\vec{r}(t-\delta t)$ of a particle at a time $t - \delta t$ and the acceleration at a time *t*, $\vec{a}(t)$ and considering the mass *m* of a particle to be constant within in time (which is true for molecules), the Verlet algorithm [Verlet, 1967] can be used to determine the position of the particle in the next timestep, $\vec{r}(t+\delta t)$. This algorithm (for its derivation see e.g. [Verlet, 1967; Allen and Tildesley, 1987; van Gunsteren et al. 2006]) states:

$$\vec{r}(t+\delta t) = 2\vec{r}(t) - \vec{r}(t-\delta t) + \delta t^2 \vec{a}(t) \quad .$$

It is important to note, that the velocity necessary for kinetic and total energy determination, is not present in the algorithm. Velocity can be expressed as [Allen and Tildesley, 1987]:

$$\vec{v}(t) = \frac{\vec{r}(t+\delta t) - \vec{r}(t-\delta t)}{2\delta t} \quad .$$

The position obtained by this algorithm is correct till the order of δt^4, while the velocity is correct only till order δt^2 [Allen and Tildesley, 1987].

To deal with the inaccuracy of velocities, the leap-frog algorithm [Hockney, 1972] was proposed to solve the equation of motion.

Knowing a position $\vec{r}(t)$ and an acceleration $\vec{a}(t)$ of a heavy point at a time *t* and velocity its $\vec{v}(t-(1/2)\delta t)$ at time *t-1/2δt*, a position at time a $t + \delta t$ can be determined as:

$$\vec{r}(t+\delta t) = \vec{r}(t) + \delta t \vec{v}(t+(1/2)\delta t)$$

and the velocity at time $t + 1/2\delta t$ is expressed as:

$$\vec{v}(t+(1/2)\delta t) = \vec{v}(t-(1/2)\delta t) + \delta t \vec{a}(t) \quad .$$

This means that velocities and positions of particle are not being determined at the same time, but the time when they are determined is mutually shifted for 1/2 δt. The time gap between the determination of the following positions resp. velocities is for both δt. The equation is solved for every particle and timestep δt of the simulation which enables us to step the motion of the simulated molecule. These equations are solved for all particles within the system simulated. This algorithm is used by Gromacs [van der Spoel et al., 2005; Hess et al., 2008], which I use in my work.

Apart from the force field *U*, the time step δt is the other crucial parameter in MD simulations. It tells, with which frequency the force acting onto particles is calculated. Obviously, a longer timestep enables performance of longer simulations. On the other hand, simulations with a shorter time step are more precise. The largest possible timestep is limited by the fastest bond vibrations in the given system (usually vibrations of hydrogens). Considering atoms as a particle, the generally used timestep is of 2 fs [van Gunsteren et al. 2006].

Considering the system being simulated from a thermodynamical point of view, we must be aware that it obeys the equation of state, which can be written in its simple form for the ideal gas. The equation states:

$pV = NkT$, where

p is pressure, V is the volume of the system, N is the number of particles in the system, T is the thermodynamic temperature and k = 1.38 $\exp^{-23}$ J/K is the Boltzmann constant. As during the simulation thermodynamic properties change, and we have only a definite precision, being able to save only a limited number of decimal places of each value, MD simulations are not able to reproduce the equation of state exactly. There must be some outer influence that takes care of this. Obviously, changes in the equation of state cannot be projected into the Boltzmann constant (as it is a constant) and MD simulations of biomacromolecules are usually considered to be of a constant number of particles.

There are two generally used ensembles in MD simulations: the isobaric-isothermal ensemble (NpT), where the volume of the simulation box may change and the isochoric-isothermal (NVT) ensemble, where pressure may vary. To keep the temperature resp. pressure constant, the system is once in the coupling time τ connected to a pressure resp. a temperature bath and the value of the given quantity is shifted to the desired value [Berendsen et al., 1984]. To do so, thermostats resp. barostats are used. The Berendsen barostat [Berendsen et al., 1984] is the most frequently used barostat in MD simulations. To keep the averaged temperature constant, Berendsen [Berendsen et al., 1984] or rescale-velocities [Bussi et al., 2007] thermostats are used. The advantage of the later one is that it keeps the kinetic energy of the system constant [van der Spoel et al., 2006].

Systems being simulated nowadays contain particles in the order of 10^5-10^6, which means they have a volume of about 1 μm^3 (all atom simulation). To rule out an effect of boundaries of this relatively (from a macroscopic point of view) small systems, periodic boundaries are used [Born and van Karmar, 1912]. The idea behind this technique is that the system being simulated is surrounded by its replicas. This means, that if a particle leaves the simulation box on one site it automatically enter it on the opposite site. To avoid the system from interactions with its periodical images, so-called minimum image convention is used, that allows the particle to interact only with the closest periodic image of every other particle.

There is the word “particle” may stand for objects of various size. Originally, each particle in a simulation represents one atom (atomistic force field resp. simulations). To reduce the number of particles, non-polar hydrogen started to be merged to adjacent heavy atom. This is referred as united-atom approach. The coarse-grained (CG) approach expects several heavy atoms to be treated as one particle, called a bead. The main reason for this is not only that the number of heavy points is considerably reduced but also the overall potential energy in CG simulations is smoother, which allows the use of longer time steps (up to 40 fs for Martini force field [Marrink et al., 2007] contrary to 2.5 fs atomistic resp. united atom simulations). This enables longer simulations (10 μs for CG vs. 250 ns for atomistic simulations). On the other hand, CG force fields are not able to provide such detailed information (such as which atoms makes hydrogen bonds) as the atomistic ones. We must be aware of that although these two – atomistic and CG – approaches use the same physics behind, they provide qualitatively different types of information, that answer different questions.

1.1.1 Force field

Potential energy in MD simulations is described by a so-called force field. This quantity is an important part of the input of any MD simulation and the extent to which the force field reflects the behavior of molecules influences the accuracy of the simulation. Force fields are considered to be the bottleneck of current MD simulations [van Gunsteren et al. 2006].

There have been many force fields introduced, differing in the compounds they are parametrized for and in the accuracy. Force fields can be divided, according to the size of particles they consider, into

all-atom, united-atom (the non-polar hydrogens are treated together with the heavy atom they are attached to) and coarse-grained (several atoms are treated as one heavy point) force fields. Polarizable force fields, such as PIPF, mimic the polarization of atoms. The force fields, that are able to capture the chemical bond formation, such as ReaxFF, are referred as reactive force fields.

Concerning the simulations of biomacromolecules, force fields OPLS (Optimized potentials for liquid simulations) [Jorgensen et al., 1996], Gromos (Groningen molecular simulation) [Oostenbrink et al., 2004], Amber (Assisted model building with energy refinement) [Cornell et al., 1995] and recently also CHARMM (Chemistry at Harvard molecular mechanics) [MacKerell at al., 1998] are generally used. The first three force field are discussed below.

Force field can be divided into a terms for a bonded U_{bonded} and a non-bonded interactions $U_{non\text{-}bonded}$:

$$U = \sum U_{bonded} + \sum U_{non-bonded} \quad .$$

Each of these terms can be split into several other terms. The bonded term contains individual terms for a bond stretching U_{bond}, an angle bending U_{angle} and both proper and improper dihedral angles torsion $U_{dihedral}$:

$$\sum U_{bonded} = \sum U_{bond} + \sum U_{angle} + \sum U_{dihedral}$$

Consideration of the bonded interactions makes only sense, when molecules are simulated. For simulations of the non-bonded atoms are these terms of the force field left out.

The non-bonded term can be split into the term for electrostatics and for van der Waals interactions.

$$\sum U_{non-bonded} = \sum U_{Coulombic} + \sum U_{van\,der\,Waals}$$

All terms are summed over all interaction of the given kind in the system. The non-bonded interactions are treated as pair-wise interactions.

There are plenty of various force fields for biomacromolecular simulations (for review see e.g. [Ponder and Case, 2003; MacKerell, 2004]) using different expression of individual potential energy terms. The following text is focused on the most frequently used atomistic/united atom force fields. These are OPLS, Gromos and Amber force fields.

For OPLS and Amber force fields the bond term U_{bond} is expressed as: $U_{bond} = K_b (r - r_0)^2$, while in Gromos this term is written as $U_{bond} = K_b (r^2 - r_0^2)^2$,where K_b is a force constant, r is the current length and r_0 is the equilibrium length for the given bond.

Similarly, a potential energy for the given bonded angle is described as $U_{angle} = K_a (\theta - \theta_0)^2$ for Amber and OPLS force field resp. as $U_{angle} = K_a (\cos\theta - \cos\theta_0)^2$ for Gromos force field. Like in the previous force field term, K_a stands for the force constant and θ and θ_0 are the current resp. the equilibrium value of the given angle. The force field terms for bonds and angles are relatively simple, as they are treated as functions with a single energetic minimum. The situation for the dihedral angles is a bit complicated, as these may have more than one minimum on their energy profile. Dihedral angle is usually written as a sum of several terms each defining a potential-energy curve minimum. Obviously, if there is only one minimum on the curve, only one terms is used. For Amber force field, the kinetic energy of the given dihedral angle is expressed as $U_{dihedral} = \sum K_d [1 + \cos(n\varphi - \gamma)]$, where K_d is a force constant, n is a periodicity, φ is the actual value of the given dihedral and γ is its phase. Gromos force field uses the potential energy term in the form of $U_{dihedral} = \sum 0.5 K_d [1 + \cos(\delta)\cos(m\varphi)]$, where K_d is a force constant, δ = ±1 is a phase shift, m is a multiplicity and φ is the actual value of the dihedral. OPLS force field uses a decomposition of the dihedral angle into the the Fourier series $U_{torsion} = \frac{1}{2}[K_{d1}[1 + \cos\phi] + K_{d2}[1 - \cos(2\phi)] + K_{d3}[1 + \cos(3\phi)] + K_{d4}[1 - \cos(4\phi)]]$, where K_{dn} are force constants, and φ is the value of the dihedral. While in Amber and OPLS improper dihedral angles are treated in the same manner as dihedral angles, there is a special term $U_{improper} = K_i [\xi - \xi_0]^2$ for improper dihedral angles in Gromos force field. Analogically to the previous, K_i stands for a force constant and ξ for the current value and ξ_0 for the equilibrium value

of the improper dihedral angle. To obtain the total bonded potential energy of the system, energies for all bonds, angles and (improper) dihedral angles in the system have to be summed.
As the OPLS force field is derived from the Amber force field, they are developed in the general philosophy and their individual therms are much closer to each other than to Gromos force field. The main difference is that, while the former deal with every molecule as a unique structure and so they have a wide range on atom types, Gromos force field uses a so-called building blocks philosophy, that expects that every larger molecule can be divided into some essential blocks that have the same properties for all the molecules. The other difference between Gromos and the other force fields is, that Gromos considers the free enthalpy of hydration and solvation in its parametrization. Further, the recent versions OPLS and Amber are all-atom force fields, while Gromos force field uses the united-atom approach.
The non-bonded potential energy term is composed of an electrostatic and a van der Waals terms.
The former potential $U_{coulomb}$ is obtained from the Coulomb's law:

$$\vec{F} = -\nabla U(\vec{r}) \Rightarrow U_{coulomb} = \frac{F_{coulomb}}{r} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \quad \text{, where}$$

F is a force, r is a positioning vector between two charged particles, that have charges of values q_i and q_j and ε_0 resp. ε_r are a vacuum resp. a relative permitivities. The total electrostatic component of the potential energy of the system is a sum of all the pair interactions within the system. In nature, a charge is not localized on one atom, but is spread over the molecule. The model of partial charges is used to describe this situation. According to this model, each atom have some fraction of the charge – either positive or negative – and the total charge of the molecule is in agreement with the charge, that molecule exhibits in experiments.
The van der Waals term is usually represented by a so-called Lennard-Jones (LJ) potential [Lennard-Jones, 1924] term:

$$U_{LJ} = \left[\frac{C_{12}}{r^{12}} - \frac{C_6}{r^6}\right] = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \quad \text{, where}$$

C_{12} stands for a constant of the repulsive part (overlapping of electron orbitals) of the potential and C_6 is a constant for attractive term of the potential. In the alternative formula, the LJ potential reaches zero at the point σ and the minimum value ε of the potential is reached at the point $2^{1/6}\sigma$. The total van der Waals energy is obtained as a sum of all the pair interactions within the system. In all the mentioned force fields, all heavy points interacting mutually via a bond or an angle interactions are excluded from the LJ interactions and for the outer members of a dihedral angles a special values for the constants of the LJ potentials are used.
Contrary to the bonded interactions, that are restricted to the given bond, angle or (improper) dihedral angle, the non-bonded interactions are long-ranged, which means that they are of an infinite range. Calculation of these forces in the full extent is computationally very expensive. That is the reason why they are at a certain distance from the given particle cut (cut-off radius). While for the LJ interaction, is the effect of a cut-off negligible, the electrostatic interactions have to be treated behind this radius [Allen and Tildesley, 1987]. (The former decreases with r^{-6}, while the later decreases much slower with r^{-1}.) The most frequently used techniques to do so are the particle-mesh Ewalds (PME), with which Amber and OPLS force fields are parametrized, and the reaction field used for Gromos force field parametrization [Allen and Tildesley, 1987].
The variables in the force field equations are to be determined by a process called parametrization. The main aim of the force field parametrization is to obtain the potential energy function, that is in agreement with the reality. Atomistic (united-atom) force-fields are usually parametrized for small molecules representing the function group being parametrized. Properties of the given group are determined either experimentally (e.g. x-ray crystalography, NMR spectroscopy) or via quantum mechanics calculations [van Gunsteren et al., 2006]. Coarse-grained (CG) simulations may be parametrized based on atomistic simulations [Hinner at al., 2009] and/or directly confronted with the experimental values of the quantity to be examined.

As the CG approach is relatively new in the field of biomolecular MD simulations, no generally used force fields exists. Rather, there are many force fields differing not only in how many atoms are treated as one bead but also in their accuracy. Only the Martini force field [Marrink et al., 2007], that is used further, is described here. It treats usually four heavy atoms as one bead and uses the same potential energy terms like Gromos force field. There are only the first neighbors excluded from the LJ interactions (no special LJ potential for the 1-3 interactions is defined) and instead of the usage of the partial charges, an unitary charge is located on a bead.

1.2 Energy minimization methods

Prior to running a MD simulation, the initial structure that is being simulated must be close to an (local) energetic minimum to avoid a rise of too big forces within the system. Otherwise the molecule could collapse in the early stages of the MD simulation. To do so, the potential energy function of the molecule(s) to be simulated is minimized. From the methods implemented in Gromacs software, the steepest descent method is generally considered to be less accurate but on the other hand the fastest one. Both, conjugated gradients and BFGS methods are more accurate but also more computationally demanding.

1.3 Homology modeling

Homology modeling is a theoretical method for a protein structure prediction based on the assumption that proteins with the same or similar primary structure (amino acid sequence) are of the same or similar secondary (local 3D structure) resp. tertiary (global structure of protein monomer) structure [Chothia and Lesk, 1986]. However, it was shown that this prediction is not always true. [Kabsch and Sander, 1985]

To be able to make a homology model of a given protein, we have to have a so-called template structure with known 3D structure (crystal or NMR structures are used), that have a sequence homology to the protein target of at least about 30%. Sometimes structures with even smaller homology are used, but here special attention must be payed to the model development and validation [Baker and Sali, 2001]. After the template structure identification, a sequence alignment, that is further used as the input for a modeling code, is made.

In the restrained based homology modeling, that is used in this thesis, various restraints on a stereochemical properties (such as bond distances, angles value) of the molecule to be modeled are obtained from the template structure and they are implemented into the target structure. These restrains are compared with the database of probability distributions given value obtained from a comparison of the homologous structures. In the last step, an objective function combined from restraints obtained as described above and CHARMM (Chemistry at HARvard Macromolecular Mechanics) potential energy (i.e. force field) [MacKerell et al., 1998] is computed and minimized [Sali and Blundell, 1993].

Experience shows that it is useful to calculate more homology models and choose the best one with respect to the quality of its stereochemical parameters.

2.1 Biological membranes

Biological membranes (further membranes) are considered to be the most common structures in cells [Israelachvili et al., 1980]. Although their most important function is to separate cells (and their compartments) from the outer environment, they are also involved in the transfer of molecules inside and outside of the cell, in the energy transduction, nerve conduction, cell signaling and biosynthesis. Membrane vesicles are used to transfer a hydrophobic molecules from one part of the cell to another and to release a cell waste into the environment [Israelachvili et al., 1980].

Membranes are made of naturally occurring hydrophobic molecules called lipids [Vodrazka; 2002; Hauser and Poupart, 2005]. These can be according to their structure divided into several categories. Biologically relevant lipids are mainly glycerolipids and sphingolipids [Hauser and Poupart, 2005]. The former are derived from glycerol to which (in the majority of the membrane lipids) two fatty acids and a polar headgroup are esterified. Based on the nature of their headgroups, glycerolipids can be further classified either as phospholipids (containing a phosphate group) or as glycolipids that have a sugar moiety bound to the glycerol. Sphingolipids are derived from a sphingosine (2-amino-4-octadecene-1,3-diol), to what a fatty acid and a headgroup unit are substitued. If there is a hydrogen bound at the headgroup position the molecule is called a ceramide [Hauser and Poupart, 2005].

The other important component of especially mammalian membranes is cholesterol (2,15-dimethyl-14-(1,5-dimethylhexyl)tetracyclo[8.7.0.0^{2,7}.0^{11,15}] heptacos-7-en-5-ol), that makes up to 30 % of the membranes. Apart from its many biological functions, it is involved in raft formation and influences some membrane properties (such as fluidity) [Yeagle, 2005].

The presence of different lipids is typical for various membranes. Phospholipids, such as phosphatidyl choline (lecithin, PC), phosphatidyl ethanolamine (PE), phosphatidyl inositol (PI), phosphatidyl serine (PS) and phosphatidic acid (PA) (for review see e.g. [van Meer et al., 2008; Korn, 1969]) are typical components of the majority of eukaryotic membranes. The most common fatty acids attached to these lipids are palmitic (16:0), stearic (18:0), oleic (18:1(9)), linoleic (18:2(9,12)) and an arachidonic (20:4(5,8,11,13)) acids [White, 1973]. Phosphatidyl ethanolamine (PE), phosphatidylglycerol (PG) and cardiolipin (1,3-bis(sn-3'-phosphatidyl)-sn-glycerol) are the most prominent lipids in the bacterial plasma membrane, (summarized in [Epand et al., 2007]). The most frequently occuring fatty acids in *E.coli*, as an example of Gram-negative bacteria, are palmitic, palmitoleic and oleic acids [Morein et al., 1996]. A plasma membrane of Gram-positive bacteria, such as *S. haemolyticus*, contains apart from stearic (18:0) and arachidic (20:0) acids also branched fatty acids, like iso-C15:0, anteiso-C15:0, in relatively rich abundance (> 10%) [Nielsen et al., 2005]. An isoprenoid hydrocarbons joined via ether (not ester like in the others) bond to glycerol are typical for archaeal membranes [Patel and Sprott; 2005]. Another membrane of unusual composition is a thylakoid membrane of photosynthetic organisms, that is composed mainly of glycolipids (for review see e.g. [Dormann, 2007; Douce and Joyard, 1990; Gounaris et al., 1986]). As the subject of my study, the thylakoid membrane is described in more detail below.

Driven by their tendency to decrease their energy, lipid molecules dissolved in water (and other polar solutions) tend to aggregate [Singer and Nicolson, 1972; Israelachvili et al., 1977; reviewed in Israelachvili et al., 1980; Cullis et al., 1986]. The form of the resultant aggregate depends on the shape of molecules that makes it (for derivation how does the shape of the molecule influences the free energy of a molecular aggregation see e.g. [Israelachvili et al., 1980]). Generally spoken, lipids can be of two major shapes: either they are rectangular or of a (truncated) cone in shape. If rectangular, resultant aggregates are planar lamellae. To prevent an exposure of the hydrophobic region of lipids into a polar solution, two of these sheets join in a fashion, that the tail regions are sandwiched between the hydrophylic headgroups, that are in contact with the solution and so they make a bilayer. This behavior is typical for e.g. PE. If a planar sheet would exists, lipids on its edge would be exposed to water, which is energetically unfavorable for them. For this reason, the majority of lipids being found in the high abundance in a biologically relevant membranes, such as

PC, sphingomyelin, PG or DGDG, has the headgroup region slightly bigger than the tail region. This enables bilayer curvature, resulting in the formation of closed features, called vesicles.
Putting (inverted) cones next to each other results into curving of the obtained sheet, with the shorter size of the cone in the center of the aggregate. For cones, the resultant aggregate, called a micelle, is of a spherical shape, while the truncated cones are more likely to form a rod-like features, referred as a hexagonal phase (H_I). Obviously, for headgroups being the shorter part of the cone, the tails point out to the environment, forming a so-called inverted hexagonal phase (H_{II}) (for more details see e.g.[Israelachvili et al., 1980; Gruner, 2005]). Trying to avoid an exposure of their tails into polar solution, individual rods of lipids in the inverted hexagonal phase aggregate, forming the cubic phase (for review see e.g. [Lewis and McElhaney, 2005, Lindblom and Rilfors, 1989]). Lipids with a single tail and gangliosides usually form micelles and a hexagonal phase, while monoglycolipids and unsaturated PE are found in the inverted hexagonal phase (H_{II}) [Israelachvili et al., 1977; Tilcock, 1986].
It is important to note, that the fatty acid composition and the nature of a solution influence the mode of lipid aggregation. An unsaturation of fatty acids introduces a kink into their structure, that increases a volume of the hydrophobic part of a lipid. Mainly divalent ions may attract charged headgroups together resulting in a decrease of the size of lipid's headgroup. The other important factor that influences the relative size of the headgroup is a level of hydration of the membranes. It is generally known, that water molecules enter the headgroup region of the membrane herby increasing its area (for review see e.g. [Israelachvili et al., 1980; Tilcock, 1986]).
Lipids try to have the smallest possible energy not only in membranes of one component but also in mixed systems. As various lipids have different properties influencing in which stage they have minimal energy, lipids of the same kind may in the mixed membranes group together, forming so-called domains. If cholesterol is part of the domain, the feature is referred to as a raft. Domains are usually formed when the membrane is formed of lipids having significantly different size of headgroups or different length of tails. If some membrane components are charged, domain formation is induced in the presence of divalent ions (for review see e.g. [Israelachvili et al., 1980]).
Membranes can be found in the two phases: the gel phase (L_β) representing a frozen stage of the matter and in a liquid-crystalline (LQ) phase (L_α), that is an equivalent to a fluid [Cullis; 1986]. Lipids in the gel phase are more closely packed than in the LQ phase resulting into a smaller area per lipid (APL) and a higher membrane thickness. This is connected with more ordered packing of their tails. A lipid diffusion slows down in the gel phase [Lewis, 2005]. The phase transition between the gel and the LQ phase, sometimes referred to as the main phase transition, can be induced by changes in the temperature, pressure or by different hydration and pH of a solution [Lewis, 2005]. Biologically active membranes are found in the LQ phase [Lewis, 2005].
As the change in fluidity of the membranes is connected with the temperature variations, organisms had to find some mechanism to keep their membranes of similar fluidity at various ranges of temperatures. The generally accepted homeoviscous adaptation theory (for review see e.g. [Hazel, 1995]) explains, that the fluidity is regulated by increasing the extend of fatty acid saturation with growing temperature.
Cellular membranes are composed not only of lipids. Other major components are proteins. The generally accepted mosaic model of the biological membrane was proposed in 1972 by Singer and Nicolson [Singer and Nicolson, 1972]. It claims that the biological membranes are lipidal bilayers permeated by proteins and other molecules such as sterols. These molecules may either penetrate through the whole membrane or may interact only with some part of the membrane. The mass ratio of protein to lipid may be higher than one (for review see e.g [Knor, 1969] and references therein). In the case of e.g. Gram-positive bacteria, it reaches up to the value of 4. For the thylakoid membrane from pea (*Pisum sativum*) this proportion is 1.6 [Chapman et al., 1983].
Membranes are not static but exhibit several movements. Apart from the standard movement of any molecules (vibration of bonds, angles and dihedral angles), lipids can rotate around their longest

axes (with the correlation time of 10^{-8} s) and freely diffuse in the plane of the bilayer. This movement is called lateral diffusion and values of the lateral diffusion coefficient of lipids in the LQ are in the order of 10^{-8} cm^2/s. Eventually, lipids can jump from one leaflet of the bilayer to the other one. This motion with a correlation time in the order of 10^4 – 10^5 s (for bilayers and vesicles) is known as flip-flop motion (for review on lipid motion see e.g. [Gawrich, 2005]). Not only lipids, but also protein and other compounds move within a membrane [Kirchhoff et al., 2008; Gounaris et al., 1986].
Biological membranes are referred to be asymmetric (for review see e.g. [op den Kamp, 1979; Singer and Nicolson, 1972]), which means that they have different proteins attached to the opposed sides of the membrane or that they have a different number of lipids of a given kind in the individual leaflets. The effort to reach symmetry is one of the driving mechanisms of the flip-flop motion [op den Kamp, 1979]. It was also proven experimentally that there are lipids preferring the position in an outer membrane leaflet (PG, PC) and others (PE, PS) that can be more likely found in an inner one (for review see e.g. [op den Kamp, 1979]). This means, that lipids with smaller headgroups are the ones, that are more likely on the inner side of the biological membranes.

2.1.1 Thylakoid membrane

A thylakoid membrane (sometimes referred as a photosynthetic membrane) is an interface on which the light phase of photosynthesis occurs [Dormann, 2007; Douce and Joyard, 1990]. It is found in cyanobacteria and chloroplasts of algae and higher plants, that are according to the generally accepted endosymbiotic theory [Mereschkowsky, 1905] believed to have the origin in a cyanobacterium having been engulfed by an eukaryotic cell. During its development chloroplasts have lost many of their original functions and had specialized mainly on energy production via photosynthesis.
To increase its energy income, cyanobacteria and chloroplasts try to have the largest possible volume of the thylakoid membrane, to be able to accommodate the highest possible amount of photosynthetic proteins.
A thylakoid membrane is arranged in a different way in cyanobacteria and in chloroplasts. In the former it makes a sheet-like structures that at several places converge to a plasma membrane [van de Meene et al., 2006]. A plasma membrane and a thylakoid membrane are hypothesized [Nickelsen et al., 2010] to be joined by a so-called Prat-defined membrane (there is a high abundance of the Prat protein in this region), where the early stages of assembly of the thylakoid-membrane-bound protein complexes occur and where chlorophyll synthesis starts.
In eukaryotic species, there is a thylakoid membrane located in the specialized organella – chloroplast. The thylakoid membrane is in chloroplasts arranged into a sack-like features called a thylakoids, that are gathered into piles named grana. Individual granas are linked into a continuous feature by a stromal thylakoid, referred as lamella. An internal space within the thylakoid is called lumen, an outer space of a thylakoid is referred as stroma (for review see e.g. [Daum and Kuhlbrandt, 2011]). Different photosynthetic supercomplexes prefer different regions of the thylakoid membrane of a chloroplast. Namely photosystem II is associated with the granal part of

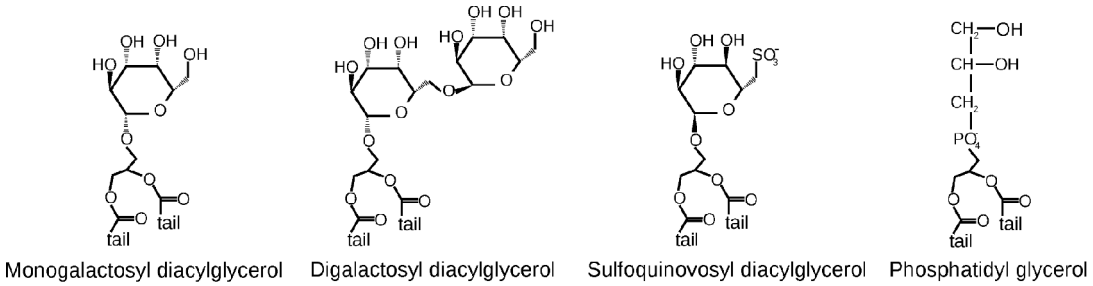


Figure 2.1: *Formula of thylakoid lipids.*

organism	t [°C]	MGDG	DGDG	SQDG	PG	PI	PE	PC	others	ref.
Arabidopsis thaliana	23/18	47.4 ± 1.2	16.5 ± 0.8	1.2 ± 0.4	9.6 ± 0.7	1.8 ± 0.5	9.0 ± 0.3	15.4 ± 1.5		Hartel et al., 2001
		51.3 ± 4.0	16.0 ± 2.1	2.4 ± 0.5	8.1 ± 1.2		9.3 ± 0.9	12.8 ± 0.8		Dormann et al., 1995
Anabaena siamensis	18	49.6 ± 2.2	24.7 ± 1.6	12.3 ± 1.1	13.4 ± 1.0					Choulub et al., 2003
	40	52.6 ± 1.5	23.1 ± 1.5	11.0 ± 0.9	13.3 ± 1.1					Choulub et al., 2003
Synechococcus PCC7942	25	54	16	11.0 ± 0.9	19					Gombos et al., 1997
	35	54	16	12.3 ± 1.1	18					Gombos et al., 1997
	34	57	13	9	21					Murata et al., 1992
Nerium oleander	20	51 ± 1	29 ± 4	6 ± 1	6 ± 1	1.2 ± 0.1	3 ± 2	3 ± 2		Orr and Raison, 1987
	45	52 ± 2	30 ± 2	6 ± 2	5 ± 1	1.3 ± 0.8	2 ± 1	5 ± 1		Orr and Raison, 1987
Pisum sativum	7/4	40	26	10	11			8	5	Chapman et al., 1983
	17/14	42	26	9	11			7	6	Chapman et al., 1983
Synechocystis PCC6803	22	54	18	15	13					Wada and Murata, 1989
	34	59	17	16	8					Wada and Murata, 1989
	30	37.4 ± 1.8	20.0 ± 1.6	28.9 ± 0.6	13.7 ± 0.8					Sakurai et al., 2007
T. vulcanus	55	43.5 ± 1.5	25.6 ± 0.2	24.8 ± 1.1	6.1 ± 0.6					Sakurai et al., 2007
Spinacia oleracea		40.1 ± 0.7	32.1 ± 0.8	15.2 ± 0.0	12.6 ± 0.1					Sakurai et al., 2007
		38	29	18	11	1		3		Nishihara et al., 1980
Oryza sp.		48.4 ± 0.1	32.9 ± 0.1	13.9 ± 0.3	4.8 ± 0.3					Sakurai et al., 2007
Mastigocladus laminosus	34	51	14	15	20					Murata et al., 1992
Plactenema boryanum	28	51	21	13	15					Murata et al., 1992
Nostoc muscorum	28	63	18	8	11					Murata et al., 1992
Synechococcus PCC7002	22	57	9	11	23					Murata et al., 1992
	34	58	12	19	13					Murata et al., 1992
Spirulina platensis	32	47	16	17	20					Murata et al., 1992
Synechococcus PCC6714	34	44	23	23	10					Murata et al., 1992
Tolypothrix tenuis	30	60	16	9	15					Murata et al., 1992
Avena sativa		61 ± 3	28 ± 2	3 ± 1	6 ± 1			2 ± 1		Giroud and Siegenthaler, 1988
C. reinhardtii		50 ± 1.5	18 ± 0.7	11 ± 0.3	9 ± 0.3				10 ± 0.4	Janero and Barrnett, 1981
Glycine max		39	28	14	11	2		2	3	Nishihara et al., 1980
Oryza sative		44	24	17	8	1		3	3	Nishihara et al., 1980
Zea mays		40	30	14	10	1		1	4	Nishihara et al., 1980

Table 2.1: *Composition of thylakoid membranes from various organisms. Values referred to for* C. reinhardtii *as phospholipid are used for PG, Nishihara et al. in their publication refer PA as others.*

the membrane, unless it undergoes a turn-over [Baena-Gonzalez and Aro, 2002], when it joins a photosystem I and an ATPase in the stroma. A cytochrome *b6f* does not have any preferable region of the thylakoid membrane (for review see [Dekker and Boekema, 2005; Gounaris et al., 1986]) although it is capable to participate in the cyclic electron transfer only if located in the grana [Barber, 1985].

The thylakoid membrane composition differs from those of an animal and a bacterial plasma membranes not only in glycolipids being their major component but also in the high fraction of polyunstaurated fatty acids [e.g. Dormann, 2007; Douce and Joyard, 1990; Gounaris et al., 1986]. Concerning headgroups, thylakoid membranes are composed mainly of glycolipids, namely of an electrically neutral 1,2-di-*O*-acyl-3-*O*-β-D-galactopyranosyl-*sn*-glycerol (shortly monogalactosyl diacylglycerol, MGDG) and 1,2-di-*O*-acyl-3-*O*-(6'-O-α-D-galactopyranosyl-β-D-galactopyranosyl)-*sn*-glycerol (shortly digalactosyl diacylglycerol, DGDG) and of at physiological pH negatively charged sulfolipid 1,2-di-*O*-acyl-3-*O*-(6'-deoxy-6'-sulfo-α-D-glucopyranosyl)-*sn*-glycerol (shortly sulfoquinovosyl diacylglycerol, SQDG) and phospholipid 1,2-diacyl-sn-glycero-3-phospho-3-(1'-sn-glycerol) (shortly phosphatidyl glycerol, PG), for formula see figure 2.1. Thylakoid membranes of some eukaryotes (like arabidopsis, pea, chlamydomonas), may contain a minor amount of other phospholipids such as PC and PE (reviewed by e.g. [Dormann, 2007 Gounaris et al., 1986]). I am not aware of any reports of the presence of the later phospolipids in the thylakoid membrane of cyanobactaria. Thylakoid membrane compositions of some organisms is reviewed in table 2.1.

As galactolipids are the major component only of membranes participating in oxygenic photosynthesis, Dormann [Dormann, 2007] stresses their necessity in this process (others glycomembranes do not have a galactose as the major component [Dormann, 2007]).

Apart from the thylakoid membrane both cyanobacteria and chloroplasts have also a plasma membrane/inner envelope, which has a similar composition as the thylakoid membrane. The outermost envelope of the chloroplast is simply referred to as an outer membrane and although rich in galactolipids, there is a significant amount of PC present in this membrane. The other difference between these two membranes is that it is DGDG and not MGDG, what is the major lipid in the outer envelope. The fractions of PG and SQDG in this membrane are similar to those in the thylakoid and inner membranes (reviewed by [Douce and Joyard, 1990]). Contrary to this, cyanobacterial cell walls (i.e. their outer membrane) contain a thin layer of peptidoglycan, which makes cyanobacteria to be a Gram-negative bacteria [Palinska, 2008].

Concerning the major fatty acid of thylakoid lipids, eukaryotic photoautotrophs can be divided into 2 groups. In the first one of a so-called “prokaryotic-like organisms”, where e.g. spinach and fresh-water algae belong, there is the first carbon of glycerol of MGDG associated with a 18:3(9,12,15) acid, while a 16:3(7,10, 13) acid is bound to the second carbon of a glycerol. In the second group of “eukaryotic-like organisms” (e.g. pea) there is a 18:3(9,12,15) acid attached to both the 1st and the 2nd carbons of a glycerol in MGDG [Heinz, 1977].

Four groups of cyanobacteria can be found with respect to their mode of fatty acid binding to the thylakoid lipids. The only common feature for all the groups is a 16:0 acid bound to the 2nd carbon of a glycerol of SQDG and PG. This acid is found on the 2nd carbon of a glycerol of MGDG and DGDG in all the groups and can be altered by a 16:1(9) acid in the first and second group. Acids bound to the 1st carbon of a glycerol are usually unique for all the groups and lipid. It was shown, that DGDG is synthetized by attaching one more galactose residue to the already existing MGDG [e.g. Sato and Murata, 1982] and so it is not surprising that these 2 lipids have the same fatty acid composition in all the cyanobacterial groups. For the first group (e.g. *Synechococcus lividus*), 16:1 and 18:1 fatty acids are found on the 1st carbon of the glycerol of MGDG and DGDG, 16:0 and 18:1 on SQDG and 18:0 and 18:1 are the major fatty acids found to the first carbon of the glycerol of PG. In e.g *Nostoc muscorum* as representatives of the second group, all lipids bind 18:1, 18:2 and 18:3α acids on sn-1 of aglycerol. For SQDG, a palmitic acid (16:0) can be located at first carbon of a glycerol as well. The third (e.g. *Spirulina platensis*) and fourth (*Synechocystis* PCC6803) groups

have the same major fatty acids, that are in the fourth group enriched by a minor fractions of an other fatty acid. The common fatty acids for these two groups are 18:1, 18:2 and 18:3γ acids for MGDG and DGDG, 16:0, 18:1 and 18:2 acids for SQDG and finally 18:1 and 18:2 acids for PG. The fourth group binds also 18:3α acid to the first carbon of the glycerol of SQDG and PG and 18:4 on the first carbon of the glycerol of MGDG and DGDG [Murata et al., 1992; Kenyon et al., 1972]. All the double bonds in the thylakoid membranes are in *cis*-conformation, with the exception for 16:1 acid on PG of eukaryotes, that is *trans* (reviewed by e.g.[Douce and Joyard, 1990; Gounaris et al., 1986]).

It is generally known, that proteins make the major mass fraction of membranes (reviewed in e.g. [Korn, 1969; Guidotti, 1972]). In thylakoid membranes the lipid-to-protein weigh fraction reaches the value of approx. 0.6 [Chapman et al., 1983].

An other interesting issue concerning the membrane composition is its lateral asymmetry (i.e. asymmetrical distribution of an individual lipid species among the membrane leaflets). Different methods were used to clarify this for higher plants (for review see [Gounaris et al., 1986; Webb and Green, 1991]). As the lipid composition of the thylakoid membranes is similar among all the species using oxygenic photosynthesis, we may expect, that the other species will have the similar lipid distributions. Studies [Rawyler and Siegenthaker, 1985; Rawyler et al, 1987; Sundby and Larsson, 1985; Rawyler and Siegenthaker, 1981; Unitt and Harwood, 1985; Siegenthaler and Giroud, 1986] are in agreement for an assymetrical distributions of MGDG and PG, where ≈60%, resp. ≈70% of the lipids was found on the outer (i.e. stromal) side of the membranes. Two [Rawyler and Siegenthaker, 1985; Rawyler et al., 1987] of three studies [Sundby and Larsson, 1985] are in agreement, that ≈ 85% of DGDG is on the inner (lumenal) side of thylakoid membranes. Only the study by Sundby and Larsson [Sundby and Larsson, 1985] found out that ≈60% of DGDG is at the stromal side of the membrane. As the lateral distribution of SQDG is difficult to be determined experimentally (reviewed by [Webb and Green, 1991]), its distribution among membrane leaflets was calculated from the distributions of the others lipids and gives a value of ≈75% of SQDG on the inner (lumenal) site [Webb and Green, 1991; Rawyler et al., 1987]. Siegenthaler and Giroud [Siegenthaler and Giroud, 1986] determined that 65 ± 10 % of PC is on the outer (stromal) side of the thylakoid membrane.

I was not able to find any information concerning the asymmetric distribution of lipids in the cyanobacterial thylakoid membranes.

Generally, there are very little experimental data available on the behavior of the pure thylakoid glycolipids, what is the information necessary for the validation of the force field parameters for molecular dynamics simulations. In fact, only the phase behavior (incl. the phase transition temperatures) is known (reviewed by [Webb and Green, 1991], more recent measurements for SQDG [Matsumoto et al., 2005]). All the lipids can be found either in a gel or a LQ phase [Shipley et al., 1973; Sen et al., 1981; Matsumoto et al., 2005]. The existence of metastable gel phases, such as a ripple phase, was reported [Sen et al., 1981] for distearoyl mono- and di- galactosyl diacylglycerol. The fully saturated MGDG tends to adopt an inverted hexagonal phase at the low level of hydration (< 10%) [Shipley et al., 1973].

The main phase transition (i.e. the phase transition between a gel and a LQ phase) was determined for di-stearoyl glycolipids at 343 K for MGDG, at 325 K for DGDG [Sen et al., 1983] and at 328 K for SQDG [Matsumoto et al., 2005].

As the second major lipid of the bacterial membranes [Morein et al., 1996; Nielsen et al., 2005], at physiological pH anionic PG is much better studied than the others cyanobacterial thylakoid lipids. With respect to its phase behavior (and so other properties connected with the shape of the molecule) it is considered to be an anionic equivalent of a PC [Tilcock, 1986; Koynova, 1997]. At the physiological pH, PG with the increasing temperature transfers from a lamellar gel phase to a lamellar LQ phase via metastable gel phases, such as a ripple phase. At a high divalent ion concentration, PG forms an inverted hexagonal phase (reviewed in [Tilcock, 1986]).

2.2 Thylakoid membrane – aims

Up to now there were only a few atomistic MD simulations of glyceroglycolipids reported (e.g. [Rog et al., 2007; Rodgers et al., 2012]. These simulations examined monoglycolipid either in pure membranes or mixed with some phospholipid. Parameters for glycolipids in these simulations were derived by mixing parameters for the sugar residues with standard parameters for lipid tails and glycerol.

As mentioned in previous section membranes composed of glycolipids are expected to be of different physical and chemical properties than the widely studied phospholipidal membranes. Apart being relatively abundant in bacterial cell walls, glycoglycerolipids make photosynthetically active membranes. As photosynthesis is widely studied, it is more than welcome to have a model of the membrane in which the photosynthetic supercomplexes are embedded. This membrane enables molecular dynamics studies of photosynthesis. Another challenge in building a model of thylakoid membrane lies in the fact that to my knowledge no other membrane composed of four lipids with different headgroups was ever published.

To be able to study the membrane by the means of coarse-grained molecular dynamics, first its parameters for MARTINI force field had to be developed. The first problem to be faced is the very little experimental data available for the thylakoid membrane lipids. For this reason it was decided that the parametrization will be based on the atomistic simulations of these lipids. As PG is a relatively abundant lipid in biological membranes that behaves in many ways similarly to PC (apart of PG is charged) its force field parameters were obtained by adaptation of PC parameters and were not developed *de novo*.

Having the parameters prepared the consequent aim of my work was to characterize the behavior of the membrane. As the main interest was in observation of long-time-scale effects – we hypothesize that PsbI protein of PSII may be involved in domain formation – a CG approach to describe the system is the only choice to reach relevant time scales and see these effect. A model of the membrane in atomistic resolution was then prepared for further work with proteins of PSII. The small atomistic patches of both compositions were characterized as well and their properties were compared with those obtained by CG simulations.

The mesophylic cyanobacterium *Synechocystis PCC6803* (further only Synechocystis) was chosen as the species on which the study will be performed as a lot of studies of PSII – mainly in pre-crystal era – were done with this organism.

As there are two different compositions of the thylakoid membrane of Synechocystis reported [Sakurai et al., 2007; Wada and Murata, 1989] that differ mainly in the content of charged lipids (43% resp. 23%) for which the membrane of Sakurai's composition is enriched on expense of MGDG of the Wada's membrane. Membranes with both compositions were constructed and characterized to explore differences between them. The most abundant combination of the fatty acid tails was used for each lipid. This means, that sn-1-γ-linolenyl-sn-2-palmitoyl tails were used for MGDG and DGDG, two palmitoyl tails were used for SQDG and sn-1-linoleyl-sn-2-palmitoyl tails were used for PG.

2.3 Methods

Coarse-grained Martini [Marrink et at., 2007] force field parameters are based on atomistic simulations of distearoyl glycolipids, kindly provided by Alex H. de Vries, (MD Group, University of Groningen, the Netherlands). These simulations use the Gromos 45A4 [Lins and Hunenberger, 2005] force field for the sugar part of lipids and the Gromos 53A6 [Oostenbrink et al., 2004] force field for the rest of the molecules. Simulations ran with a 2fs timestep for 100 ns. The last 50 ns of the trajectories were used for parametrization.

The method proposed by Hinner et al. [Hinner et al., 2009] was used for the parametrization of thylakoid lipids MGDG, DGDG and SQDG. This method is based on the atomistic simulation being converted into the CG representation by the script *g_fg2cg*, that can be downloaded from the Martini official web pages. This script calculates the position of a center of mass of atoms defined to form the given bead over the whole trajectory. This center of mass is then treated as a bead. Probability distributions of bond lengths, angle and dihedral angles values are calculated. CG parameters are set to reproduce these distributions (mainly equilibrium values, height of the distribution is not that important) in highest possible agreement.

The CG simulations for the development of force field parameters were performed on membranes of 128 lipids at full hydration (at least 14 water beads per lipid). Simulations ran with a timestep of 20 fs for at least 100 ns. Prior to the proper production runs, systems underwent 2000 steps of steepest descent energy minimization. The standard Martini run setting was used for the MD simulations. This means that long-range interactions were cut at a 1.2 nm distance from the bead, the shift algorithm was used to preserve grouping of charged particles at the cut-off distance. Pressure was kept semiisotropically at 1 atm by the Berendsen barostat with a compressibility $3e^{-5}$ bar^{-1}, that was connected to the pressure bath once in a ps. The temperature was coupled by the Berendsen thermostat connecting to a temperature bath separately for lipids and solution once in 0.5 ps. The reference temperature was set 5 K above the experimentally determined phase transition temperature of the given lipid. This means, that the values 353 K for MGDG, 335 K for DGDG and 338 K for SQDG were used. The last frame of the atomistic simulation mapped into the CG representation was used as the initial structure for the CG simulations. This structure was hydrated and if necessary (SQDG) neutralized. All simulations and the analyses were performed in Gromacs 4.5.5 software [Hess et al., 2008]. The probability distributions for bond lengths and for values of angles and dihedral angles were than calculated and compared with those obtained from the atomistic simulations. If the CG distributions were not in agreement with the atomistic ones, the force field parameters were adjusted. The whole cycle was than repeated.

Martini force field parameters were developed with the aim to preserve the timestep of at least 20 fs. In the later stage of the parametrization, major attention was paid to the correct reproduction of the area per lipid (APL) and the membrane thickness. APL is calculated by dividing the area of the plane of the membrane by the number of lipids in the monolayer. Membrane thickness is determine as the distance between maxima in the electron density profile of a solvated membrane. The so-called smoothed electron density (each lipid devided into headgroup, glycerol, tail1 and tail2 parts and on each bead the average number of electrons was assigned) was used to compute electron density profiles. When the exact number of electrons is assigned to the given bead, unnatural humps appear on the electron density curves. These humps are the result of grouping electrons unequally among beads.

The phase transition temperatures were determined as proposed by Marrink et al. [Marrink et al., 2005]. This method first couples one part of the membrane in the way to be in a gel phase, while the other part is equilibrated into a liquid-crystalline phase. Then the whole membrane is coupled at the same temperature, and after its equilibration it is inspected, which phase is preferred. This may be checked either visually, or by comparing properties of the membrane adopting the two phases with these obtained from a membrane in one phase. The phase transition temperature is determined as an interval between the highest temperature of the membrane in the gel phase and the lowest

temperature of membrane in the LQ phase. Phase transition goes together with a rapid change of membrane properties. Bilayers of 512 lipids and hydration of at least 20 water beads per lipid (if necessary neutralized) were used for these type of analysis in the simulations of unmixed membranes. These membranes were built by the *insane* script, kindly provided by Tsjerk Wassernaar, (MD Group, University of Groningen, the Netherlands), and energetically minimized by a procedure composed of 1000 steps of steepest descent energy minimization, and a series of 1000 steps of NVT simulation followed by NpT MD simulations with timesteps of 1ns, 2ns, 5ns and 10ns. The temperatures on which the two phases of the membranes were equilibrated, were determined empirically. These temperatures (first for a LQ phase, the second for a gel) are 353 K resp. 273 K for MGDG, 335 K resp. 235 K for DGDG and 368 K resp. 268 K for SQDG. Water was coupled to the average temperature of the gel and the LQ temperature. As for DGDG this average is close to the freezing temperature of water, the room temperature of 300 K was used. To determine the phase transition temperature of thylakoid membranes, bilayers of 200 lipids hydrated at about 15 water beads per lipid were built as described for the systems used for the phase transition temperature determination, duplicated and one of the bilayers was shifted in the x-direction and merged with the previous one (to have the parts of the membranes coupled at the different temperatures ofthe identical composition). After another minimization cycle, one part of the membranes was coupled at 310 K and the second one at 210 K. Than the systems were interpolated to the phase transition temperature as described above. The standard Martini simulation settings were used.

As distearoyl (DS) glycolipids are not natural for the thylakoid membrane, bilayers composed of sn-1-γ-linolenyl-sn-2-palmitoyl (LP) lipids were simulated for all four lipids (MGDG, DGDG, SQDG, PG), to see, how does the change of the fatty acid composition influences a membrane behavior. As SQDG occurs in the thylakoid membrane with two palmitoyl tails esterified and PG as sn-1-linoleyl-sn-2-palmitoyl (lP) PG, membranes of these lipids were examined as well. All these simulations were performed at 310 K, that is a temperature 5 K above the optimal growth temperature of Synechocystis. This rise of the temperature was done, because of MGDG and DGDG show the phase transition temperature shifted towards a gel in comparison with the experimental data. As DSSQDG is in a gel phase at 310 K, this molecule was characterized at

purpose	temp [K]	MGDG	DGDG	SQDG	PG	NA+	W	hydration	tails
parameters	353/283	128					1849	14.4	DS
	335/235		128				2630	20.5	
	338/268/368			128		128	2630	20.5	
phase	273/353	512					10240	20	
	235/335		512				10478	20.5	
	268/368			512		512	10346	20.2	
properties	310	128					2152	16.8	LP
	310		128				1926	15	
	310			128		128	1896	14.8	
	310				128	128	1974	15.4	
	310			128		128	2268	17.7	DP
	310				128	128	1972	15.4	lP
TM small	310	118	34	32	16	48	3068	15.3	TM
	310	72	40	56	32	88	3364	16.8	
TM phase	310/208	240	64	64	32	96	5746	14.4	
	310/208	144	80	112	64	176	6810	17	
TM big	310	1416	408	384	192	576	36812	15.3	
	310	864	480	672	384	1056	40380	16.8	
TM atomistic	310	120	32	32	16	48	8007	40	
	310	72	40	56	32	88	11501	57	

Table 2.2: *List of membrane simulations performed. TM stands for thylakoid membrane (lipid combination: LP for MGDG and DGDG, DP for SQDG and lP for PG).*

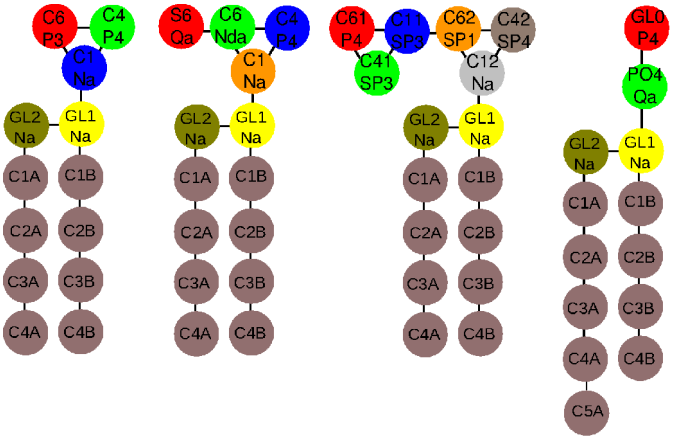
368 K to have an obvious LQ behavior. DSSQDG and DPSQDG are of the same architecture in Martini force field. All the membranes were composed of 128 lipids hydrated at at least 14.8 water beads per lipid. The input structures were built by the script *insane*, followed by the above mentioned energy minimization. Simulations with the standard Martini simulation setting ran for 1.2 μs with the timestep of 30 fs. Semiisotropic pressure coupling was used. Analysis of the membrane behavior were performed with the standard tools of Gromacs. Order parameters were calculated by the *do-order* script, downloaded from the Martini web pages (25 frames between 0.5 and 1.2 μs).

As already mentioned in the introduction, there are two different membrane compositions reported in the literature for the thylakoid membrane of Synechocystis. Both of them were characterized by the CG MD simulations, and differences in their structure and behavior were characterized, with the aim to tell, which one has a higher probability to reflect reality. Membranes of 200 lipids were built by the *insane* script. Simulation parameters are the same as for the previous simulations. They ran for 10 μs and the last 1.2 μs were used for the analysis.

Huge membranes of 2400 lipids were built by merging 12 of the last frames of the small thylakoid membranes together into a grid of 3 × 4 frames. The simulation of the membrane of Wada's composition ran for 10 μs and the one of Sakurai's composition for 3 μs, both with a timestep of 30 fs. The standard Martini simulation settings were used.

All the simulations performed are summarized in table 2.2, for the CG representation of tails see figure 2.2.

The atomistic simulations were performed for thylakoid membranes of both Wada's and Sakurai's composition. The initial configurations of these systems were prepared manually in Yasara [Krieger et al., 2004], hereby 25 lipids were randomly put together with their molecular surfaces not touching each other. These blocks were duplicated 4times and positioned next to each other and finally these monolayers were duplicated and arranged in the bilayer manner. The resultant composition of the system referred as Wada´s membrane contains 120 LPMGDG molecules, 32 LPDGDG and DPSQDG lipids and 16 lPPG residues. The second membrane (Sakuari`s composition) contains 72 LPMGDG, 40 LPDGDG, 56 DPQDG and 32 lPPG. The systems were then fully hydrated and neutralized by the standard means of Gromacs 4.0.7 package [van der Spoel et al., 2005], in which their behavior was examined. The above mentioned Gromos force field was used. Electrostatic interactions were cut at 1.4 nm, beyond what the reaction field approximation with the relative permittivity of 62 bar^{-1} was used. The neighbor list was updated every 5 steps, neighbors were searched to the distance of 0.9 nm and the periodic boundary conditions were used. The temperature of the system was hold at 310 K by connecting the membrane and solution separately to the temperature bath (Berendsen thermostat) every 0.5 ps. Prior to the production simulation with a timestep of 2 fs, and pressure of 1 bar maintained semiisotropically by the Berendsen barostat with a compressibility of 4.6 e^{-05} bar^{-1} and connected once a picosecond, the system underwent 1000 step of the steepest descent energy minimization The following relaxation



tail	glycolipids	PG
16:0	C1-C1-C1-C1	C1-C1-C1-C1-C1
18:0	C1-C1-C1-C1	C1-C1-C1-C1-C1
18:3(6,9,12)	C4-C4-C4-C1	C1-C4-C4-C4-C1
18:2(9,12)	C1-C4-C4-C1	C1-C1-C4-C4-C1

Figure 2.2: *CG representation of molecules used. Bead types for various fatty acids are shown in the table above (from glycerol to the end of the tail).*

simulations use the same settings, only the time constant for the pressure bath connecting (τ_p) and the timestep differs. The first simulation with the timestep of 0.5 fs ran for 0.5 ns with the τ_p = 2.5 ps. Next 3 simulations with the timesteps of 1 fs for 1 ns with τ_p = 4 ps, for 5 ns with τ_p = 2 ps and for 5 ns with τ_p = 1 ps, respectively. The last relaxation run with a timestep of 2 fs with τ_p = 2 ps for 5 ns. The production run with a timestep of 2 fs for 150 ns. The last 60 ns of the production simulation were used for the membrane characterization by the standard means of Gromacs.

2.4 Results and discussion

2.4.1 Martini force field parametrization

The main aim of this part of the thesis was to develop Martini force field parameters for the headgroups of glycolipids MGDG, DGDG and SQDG that are stable at the timesteps of at least 20 fs (timestep for the proteins in the Martini force field). Standard Martini tail parameters (the DS tails were mapped as four C1 beads) were used for the rest of the molecules.

At the early stages of the procedure several different mapping schemes for each molecule were tested and the ones having probability distributions for bond lengths and for angle sizes closest to the unimodal distributions were used. The mapping schemes and the bead types used in the parametrization are shown in figure 2.3. The bead designed as GL1 is not a standard Martini GL1 bead, as the 3rd carbon of a glycerol is not included. For this reason, properties of the GL1 – GL2 bond (and partly of the angles including this bond) differ from standard Martini parameters.

While capturing the probability distributions for bond lengths and angle sizes was relatively easy, obtaining the representation of the CG dihedral angles to be in the agreement with the atomistic distributions turned to be impossible at the desired timesteps.

To reproduce probability distributions of the size of dihedral angles, several things were tried:

- various bonds between non-bonding atoms were introduced
- various angles between non-bonded atoms were introduced
- improper dihedral angles were introduced
- dihedral angles were turned into angles by joining the middle beads into a dummy bead and there was an angle parameters introduced for the angle: bead1 – dummy bead – bead4. (The tricky thing of this technique is to define the dummy bead to be in the middle of the inner beads of the dihedral angle. This is complicated, as the distance between these beads changes (bond vibration). I usually solved this by defining a bond between one of the inner beads of the dihedral angle and the dummy bead. If a constraint was used instead of the bond, the system got frozen). This technique leads to the relatively good representation of the dihedral angles, that are symmetrical with respect to the 0°. To shift the asymmetry, a non-bonded angle of 3 beads making the dihedral angle is to be defined.
- altered 1-3 and 1-4 non-bonded interactions were introduced (In the Martini force field there are only the first neighbors excluded from the non-bonded interactions and all other non-bonded interactions are handled in an unique manner. Analogically to the Gromos force field, whose parametrization strategy Martini uses, I tried to introduce the reduced non-bonded interactions

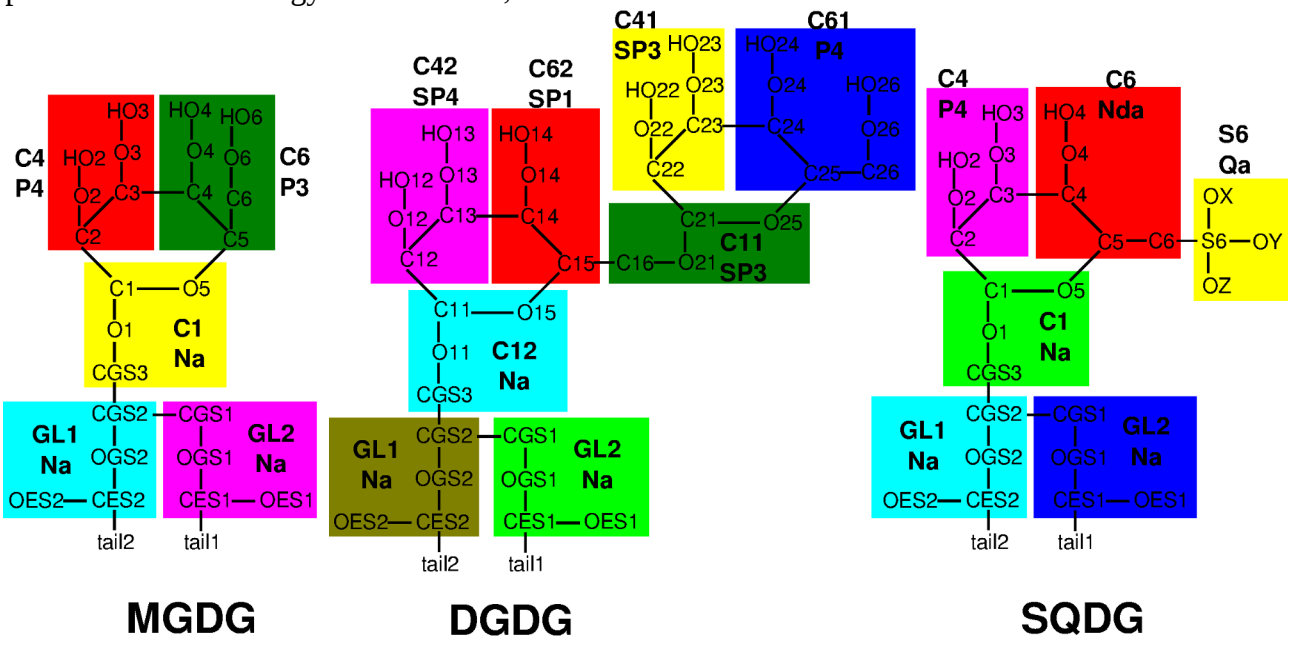


Figure 2.3: *Final mapping used for glycolipids MGDG, DGDG and SQDG, incl. bead names and types for Martini force field.*

between the side atoms of an angle or a dihedral angle. In the later case, the 1-3 non-bonded interactions were excluded. The level of the reduction tested reached 10 % to 80 % of the original value.)

- different definitions of the dihedral angle force field term were used, although neither this enabled to perform the simulations with the desired timestep.

A break-through in the parametrization came after resigning to the effort to capture the dihedral angle probability distributions and to try to reproduce the membrane thickness and APL obtained from the atomistic simulations at the given temperature. To stabilize the simulations, several high bonds and angles force constants were decreased. Doing this it was checked that the equilibrium value of the given CG parameter is after decreasing the force constant in agreement with the value obtained by the atomistic simulations. To adjust the thickness and APL of the membrane, the equilibrium value of an angle – mainly of the angle made by the last bead of the headgroup and the glycerol beads - were altered or its force constant was changed. Mainly in the case of SQDG, it appeared that the extend of water entering the headgroup region of the membrane influences the APL. To change this interaction, the bead type of (usually one) headgroup bead was changed to be of one grade more/less attractive with respect to water beads. In the case of DGDG, it was necessary to increase the soaking of the headgroup region of the membrane. To achieve this, some of the headgroup beads were turned into small bead type. Testing the different positions and numbers of small beads revealed that it is the number of small beads and not their position, that influences the APL of the molecule. The small bead types were then put onto the beads that are made of less atoms than those that are of normal type. The resultant parameters are shown in table 2.3.

The membrane thickness and APL for LQ and gel (the later was not the subject of the parametrization) phases of the DS glycolipids and their comparison with the values obtained by the atomistic simulations are shown in table 2.4. The phase transition temperatures for the CG

glycolipid	bonds	R_b(nm)	K_b (kJ/mol nm2)	angles	θ_0(deg)	K_a (kJ/mol)	exclusions
MGDG	C4 – C6	0.363	38000	C1 – C4 – C6	63	22	C1 – GL2
	C1 – C6	0.366	15000	C4 – C1 – C6	63	17	C4 – GL1
	C1 – C4	0.326	44000	C6 – C1 – GL1	139	28	C6 – GL1
	C1 – GL1	0.360	3000	C1 – C6 – C4	53	30	
	GL1 – GL2	0.414	4000	C4 – C1 – GL1	133	25	
				C1 – GL1 – GL2	81	70	
DGDG	C41 – C61	0.364	35000	C11 – C41 – C61	48	30	C12 – GL2
	C11 – C61	0.274	35000	C61 – C11 – C62	120	30	
	C11 – C41	0.312	16000	C41 – C11 – C62	156	40	
	C11 – C62	0.369	15000	C11 – C62 – C42	145	25	
	C42 – C62	0.315	35000	C11 – C62 – C12	106	35	
	C12 – C62	0.346	25000	C62 – C42 – C12	65	25	
	C12 – C42	0.325	35000	C62 – C12 – C42	56	35	
	C12 – GL1	0.372	3900	C62 – C12 – GL1	151	30	
	GL1 – GL2	0.414	4500	C42 – C12 – GL1	123	35	
				C12 – GL1 – GL2	77	40	
SQDG	C6 – S6	0.360	25000	S6 – C6 – C4	126	60	S6 – C4
	C4 – C6	0.250	33000	S6 – C6 – C1	75	50	S6 – C1
	C1 – C6	0.340	35000	C6 – C1 – GL1	165	30	C1 – GL2
	C1 – C4	0.290	34000	C4 – C1 – GL1	124	35	
	C1 – GL1	0.370	5600	C1 – GL1 – GL2	85	60	
	GL1 – GL2	0.385	4900				

Table 2.3: *Newly developed Martini force field parameters for glycolipids MGDG, DGDG and SQDG.*

lipid	phase	simulation	APL [nm^2]	thick [nm]	temp [K]	transition (CG)
MGDG	*gel*	CG	0.456	5.0	283	306 – 310
		AA	0.480	4.8	313	
	liquid	CG	0.627	4.0	353	
		AA	0.629	3.9	353	
DGDG	*gel*	CG	0.500	4.6	245	290 – 295
		AA	0.560	4.2	295	
	liquid	CG	0.660	4.1	335	
		AA	0.670	4.2	335	
SQDG	*gel*	CG	0.462	4.9	268	328 – 333
		AA	0.520	4.8	318	
	liquid	CG	0.570	4.4	338	
		CG	0.596	4.5	368	
		AA	0.570	4.6	338	

Table 2.4: *Comparison of APL and membrane thickness for distearoyl glycolipids membranes in gel and LQ phase for atomistic (kindly provided by Alex H. de Vries, University of Groningen, the Netherlands) and CG simulations. The phase transition temperatures in the lower field for each lipid are experimentally determined values [Sen et al., 1983; Matsumoto et al., 2005].*

simulations are shown there as well. Concerning the gel phase, the CG simulations are very below the phase transition temperature and so the system is frozen. As can be seen in the table, the properties obtained by the CG simulations are in very good agreement with those obtained by the atomistic ones. As the 18-carbon tails are inbetween of being mapped into 4 and 5 beads, both mappings were tried. For MGDG and DGDG the mapping of 4bead-tails was closer to the atomistic simulations, while for SQDG the atomistic value is in between of the CG values obtained with 4 and 5 beads (data not shown). To keep the consistency of all the models, the 4bead tail is used for SQDG.

The phase transition temperatures obtained by CG simulations were compared with experimental values. As can be seen in table 2.4, SQDG represents the phase transition temperature in agreement with the experiment, while for MGDG and DGDG the phase transition temperature of the model is of about 30 K below experimental values.

An occurrence of the phase change in the given temperature interval is supported by the APLs and membrane thicknesses for both, gel and LQ phase in table 2.5. For all cases it can be seen that the membrane thickness is of 0.5 nm higher in the gel than the LQ phase while the APL in the gel phase is of about 85% of the LQ value. This table also shows, that SQDG was parametrized much closer to the phase transition temperature than the other two lipids and so rising the temperature will effect its properties much more than for MGDG and DGDG as the result of further disordering the tails.

The smoothed electron density profiles by which the membrane thickness was determined, can be seen in figure 2.4 for both, gel and LQ phase. All diagrams show that the tail region of the membranes and the water regions are separated by the headgroup region of lipids, where the polar headgroups are mixed with water. Ions are associated mainly with the headgroup region of the membranes, although they diffuse into water. The positions of lipid headgroups, glycerol link and the tail region show, that membranes are assembled as expected, i.e. as a bilayer. The curves for membranes in the gel phase exhibit features of a frozen system.

lipid	phase	APL [nm^2]	thick. [nm]	temp [K]
MGDG	*gel*	0.470	4.8	306
	liquid	0.567	4.3	310
DGDG	*gel*	0.520	4.6	290
	liquid	0.590	4.2	295
SQDG	*gel*	0.479	4.8	328
	liquid	0.564	4.3	333

Table 2.5: *Comparison of APL and membrane thickness right above and below the phase transition temperature.*

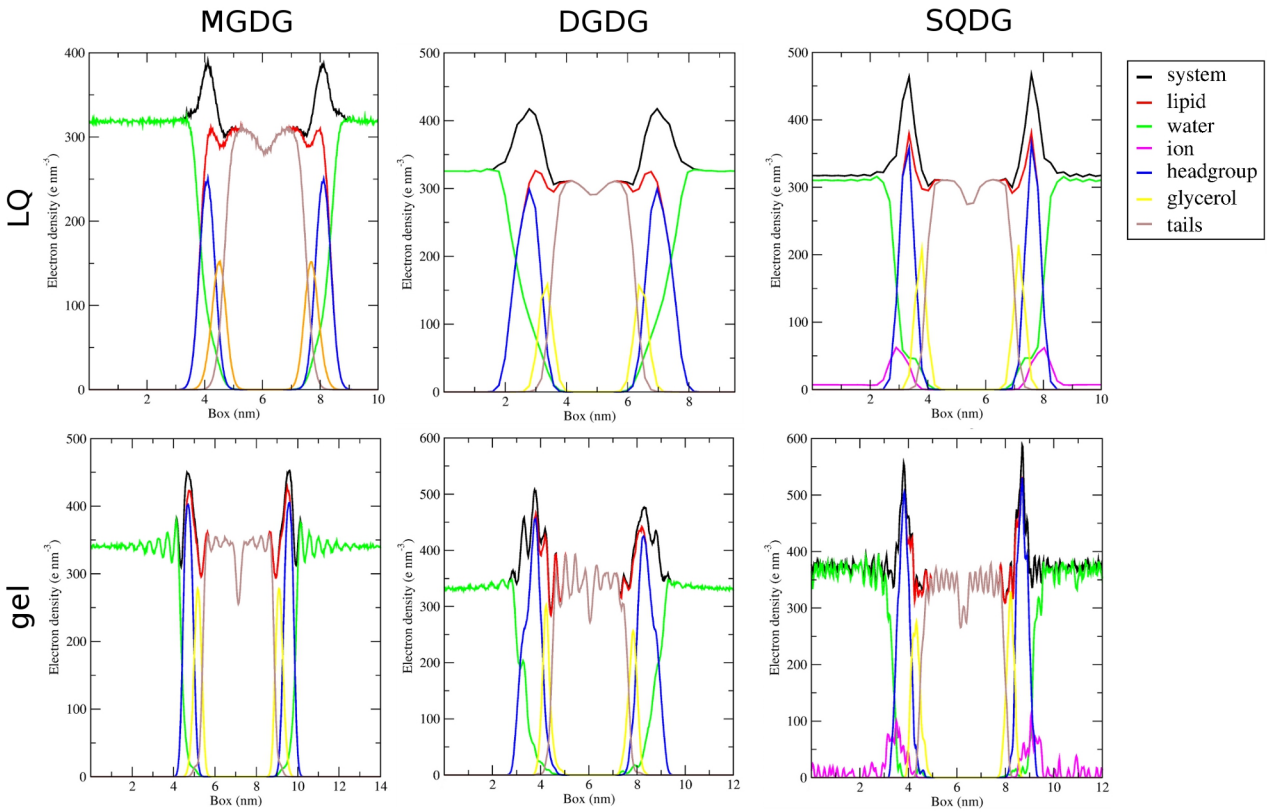
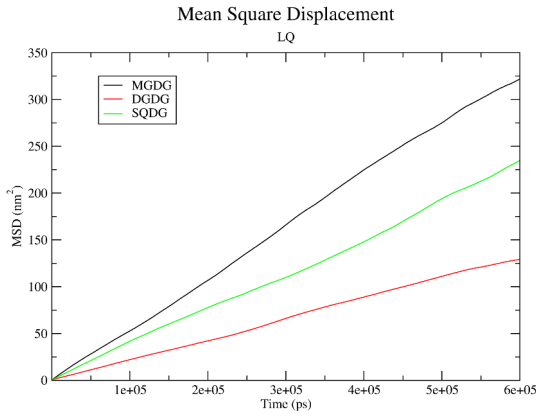


Figure 2.4: *Smoothed electron density profiles for the membranes in LQ and gel phases. These curves were used for the membrane thickness determination.*

Order parameters (in the CG simulations calculated as an angle between the given bond and the z-axis of the simulation box) shown in table 2.6 are another confirmation of the existence of two different phases in membranes. Their values tell, that at temperatures used for gel simulations there is SQDG more ordered than MGDG and that DGDG is the the less ordered molecule. For the LQ phase, MGDG and DGDG are of similar disorder, while SQDG is more ordered. This may be linked with the smallest APL of SQDG at the temperature used for the calculations (APL = 0.596 nm^2, thickness = 4.5 nm at 368 K) and so there is not enough space for tails to fully extend.
Lateral diffusion coefficients for DS galactolipids in the LQ phase are shown in table 2.7 and the belonging mean square displacement (MSD) curves can be seen in figure 2.5. MSD curves for membranes in the gel phase were also determined, but MSD here is that low that lipids hardly move for more than their own diameter and so it did not make sense to determine the lateral diffusion coefficients. This can be considered as another proof that the gel membranes are frozen.
Values of the lateral diffusion coefficient of membranes in the LQ phase are of 5 order higher than the experimental values for phospoglycerolipids and of 2 orders higher than the values obtained by atomistic simulations of phosphoglycerolipids. It was impossible to determine the lateral diffusion coefficients from the atomistic simulations of glycoglycerolipids, that were used for the CG force

		C1A-C2A	C2A-C3A	C3A-C4A	C1B-C2B	C2B-C3B	C3B-C4B
MGDG	*gel (283)*	0.890	0.915	0.885	0.873	0.899	0.875
	LQ (353)	0.419	0.347	0.218	0.444	0.375	0.235
DGDG	*gel (235)*	0.758	0.725	0.660	0.749	0.717	0.662
	LQ (335)	0.458	0.361	0.227	0.470	0.358	0.220
SQDG	*gel (268)*	0.904	0.897	0.839	0.940	0.892	0.847
	LQ (368)	0.582	0.527	0.355	0.589	0.516	0.371

Table 2.6: *Order parameters for lipid tails. Numbers in brackets in the column with the membrane phase are the temperatures at which the given simulation was performed.*



MGDG	DGDG	SQDG
1.4e-04	6.2e-05	8.4e-05

Table 2.7: *Lateral diffusion coefficients (in cm^2/s) for membranes in LQ phase. It was impossible to determine the lateral diffusion coefficients for membranes in the gel phase.*

Figure 2.5: *MSD curves that served for the determination of lateral diffusion coefficients.*

field parametrization [A.H. de Vries, personal communication]. This speeding up of lateral diffusion in CG simulations may be caused by smoothing of the potential energy surface of molecules when mapped onto the CG level. When neighboring molecules may pass along each other with a smaller chance of being – at least for some while – trapped into the energetic minimum of the surface of the other molecule (their surface is smoother) and so their mutual approaching is faster. This can also explain why lateral diffusion in atomistic simulations is faster than the one occurring in nature (and so measured). Analogically, we may state that this is caused by atomistic force fields making the molecular potential energy surface in simulation smoother than it would be in the reality.

As can be seen in table 2.7, MGDG is the molecule with the fastest lateral diffusion, while DGDG is the slowest one. This can be explained by DGDG having two polar headgroups that tend to interact with water and other polar components of the system such as other molecules. This interaction is the reason, why the diffusion is slowed down. In case of SQDG, whose lateral diffusion is closer to those of DGDG than MGDG, we must consider the electrostatic interactions of SQDG and counterions that hold the molecules together and result in slowing diffusion down.

When preparing the publication about the CG parameters, there was a request to unify the parameters of the glycerol beads. I tried to use standard Martini glycerol parameters (for the bond and angle between glycerol beads and the last bead of the headgroup) and to unify the glycerol parameters for all three molecules (as shown in table 2.3, they differs a bit) to use an averaged parameters for bonds r_0 = 0.414 nm and K_b = 4500 kJ/mol nm and the angle parameters θ_0 = 80° and K_a = 60 kJ/mol. The reason, why the parameters shown above, differ from the standard Martini values is that in the CG representation used here the 3rd carbon of a glycerol is mapped into the first headgroup bead and not in the the second glycerol bead as in the standard Martini.

As can be seen in table 2.8, there is almost no effect of these changes onto the membrane thickness, while there are some little changes in the APL of less than 4 % of the original value and so these changes can be claimed as negligible. In the resultant paper, only the unification of the GL1 – GL2 bond was used (unified for the above mentioned value r_0 = 0.414 nm and K_b = 4500 kJ/mol nm, which is the original values for DGDG).

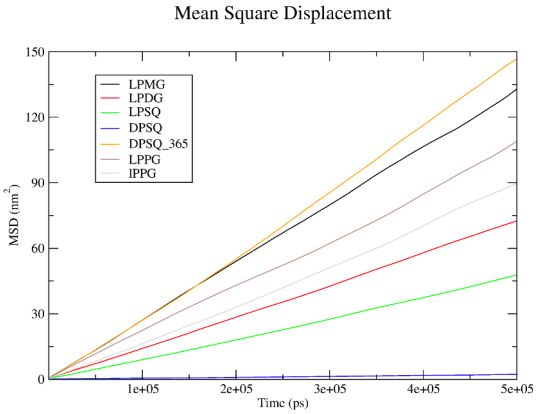
		original	unified	martini	temp [K]
MGDG	*APL [nm²]*	0.627	0.616	0.612	353
	thickness [nm]	4.0	4.0	4.0	
DGDG	*APL [nm²]*	0.660	0.672	0.658	335
	thickness [nm]	4.1	4.0	4.1	
SQDG	*APL [nm²]*	0.570	0.574	0.565	338
	thickness [nm]	4.4	4.4	4.5	

Table 2.8: *Effect of Martini parameters unification onto APL and membrane thickness.*

2.4.2 Small membranes of lipid present in thylakoid membranes

Pure membranes of lipids present in the thylakoid membrane were characterized to be able to compare, how does mixing of these lipids into the thylakoid membrane influence properties of every lipid. If not mentioned otherwise, all simulations were performed at 310 K.

Resultant characteristics of these membranes are shown in table 2.9, and the MSD curves in figure 2.6. Electron density profiles (data not shown) exhibit the typical arrangement of bilayers with ions associated with the headgroup region of lipids, where water molecules also enter. Visual inspection of simulations reveal that LPMG and LPDG are in LQ phase, while DPSQ is in the gel phase. Not even the introduction of 3 unsaturated bonds (system designed as LPSQ) turned the SQDG membrane into the LQ phase, although it approaches the phase transition temperature and some disturbances in ordered membrane structure appear. A similar situation occurs for PG, where the saturation of one bond increases the phase transition temperature and the system approaches the phase transition temperature from the LQ site.



	APL [nm²/lipid]	thickness [nm]	lat. diffusion e-05 [cm²/s]	phase
LPMG	0.684	3.5	6.5	LQ
LPDG	0.719	3.7	3.5	LQ
LPSQ	0.572	4.4	2.4	gel*
DPSQ	0.483	4.7		gel
DPSQ_365	0.601	4.3	7.4	LQ
LPPG	0.701	4.0	4.7	LQ
lPPG	0.610	4.4	4.5	LQ*

Figure 2.6: *MSD curves that served for the determination of lateral diffusion coefficients in table 2.9.*

Table 2.9: *Characteristics of membranes of thylakoid lipids with various tails at 310 K. Asterisk stands for membrane close to the phase transition temperature. Lateral diffusion coefficient for DPSQ at 310 K could not be determined.*

Looking at the other characteristics of the membranes, it is necessary to consider in which phase the membrane is and how far from the phase transition temperature it was characterized. Concerning membranes in the perfect LQ phase (i.e. LPMG, LPDG, DPSQ at 365 K and LPPG) we will find out that the electrically neutral membranes have a higher APL on expense of a smaller membrane thickness. For DPSQ (charged) the lack of unsaturated tails may participate in the decrease of its APL, see above. The lateral diffusion coefficient is the highest for DPSQ, which is influenced by higher temperature at which this simulation was performed in comparison with the others. LPDG shows the slowest lateral diffusion probably because of its big headgroup. LPMG moves faster than LPPG. For PG apart from the LPPG also the lPPG membrane was characterized. A lack of one double bond in the first tail of lPPG increases the phase transition temperature and shifts it towards the temperature at which the simulation is performed (310 K). The lPPG membrane is then not in the perfect LQ phase, decreasing APL and increasing its thickness. The lateral diffusion coefficient stays similar. A similar situation happens for DPSQ where the introduction of even three double bonds and prolonging its tail on the 1st carbon of a glycerol does not shift the membrane to the LQ phase at 310 K. It only increases the APL and the lateral diffusion coefficient and decreases the membrane thickness.

The order parameter shown in table 2.10 reflects, that the membranes are in different phases. All of them but LPDG show the general trend, that the further the bead from the headgroup, the less ordered the tails are. In case of LPDG we may suspect that its huge (in comparison with other lipids) headgroup may interact with the tails and disorder them. LPMG and LPDG show very

	C1A – C2A	C2A – C3A	C3A – C4A	C4A – C5A	C1B – C2B	C2B – C3B	C3B – C4B
LPMG	0.499	0.337	0.180		0.180	0.156	0.103
LPDG	0.485	0.326	0.162		0.094	0.147	0.103
LPSQ	0.656	0.587	0.410		0.544	0.549	0.427
DPSQ	0.872	0.871	0.808		0.876	0.864	0.788
DPSQ_365	0.579	0.505	0.356		0.577	0.504	0.346
LPPG	0.330	0.100	0.076	0.043	0.498	0.416	0.268
lPPG	0.584	0.475	0.225	0.089	0.552	0.474	0.332

Table 2.10: *Order parameters of characterized membranes. Tails on the sn-1 carbon are designed as A.*

similar values of the order parameters. The higher order of the tails of LPSQ is caused by the gel phase of the membrane. The smaller order of LPPG than of LPMG and LPDG could be explained by the different ε value of a non-bonded interaction between the last headgroup bead and the tail beads. Higher order of lPPG than LPPG reflect not only the higher degree of unsaturation of the later lipid but also its further distance from the phase transition temperature.

2.4.3 Thylakoid membrane – coarse-grained

All the simulations of both, small and big systems, show bilayers in LQ phase as can be seen from the electron density profiles shown in figure 2.7 and some frames of the simulations shown in figure 2.8. As expected, solvent enters the headgroup region of membranes with ions preferably associated here. The bilayer aggregation of lipids is supported by an arrangement of lipid headgroups on the outermost edge of the membrane, followed by the glycerol region and finally with the tail region sandwiched inbetween. The tails of both leaflets of the membranes make one continuous domain. It is not surprising that the individual regions of lipids overlap. The individual peaks of a given region of the membrane are within the small simulations much clearer defined that in the big ones. This may be caused by an undulation of the membrane that is more obvious in bigger than smaller membranes. When curving, the normal of the membrane is not parallel with the z-axis of the simulation cell (with respect to which the electron density profile of the system is

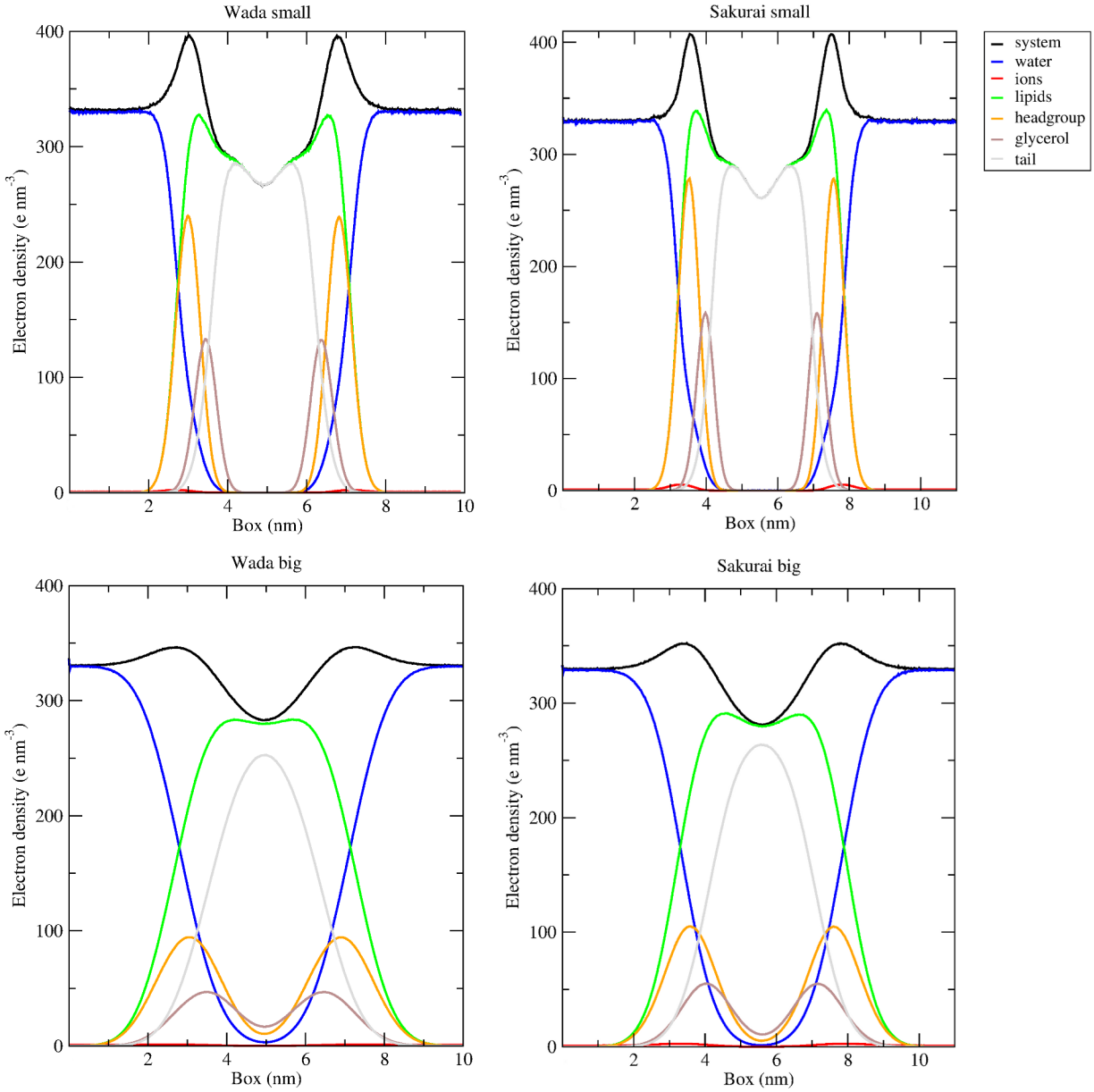


Figure 2.7: *Electron density profiles of the large and small thylakoid membranes of Wada's and Sakurai's composition. The widening of the curves for the big membranes is likely caused by the membrane undulation. The red curve of ions is much lower than the other ones, as the number of electrons of sodium ion (10) was assigned to the NA bead while the other beads have usually more than 30 electrons.*

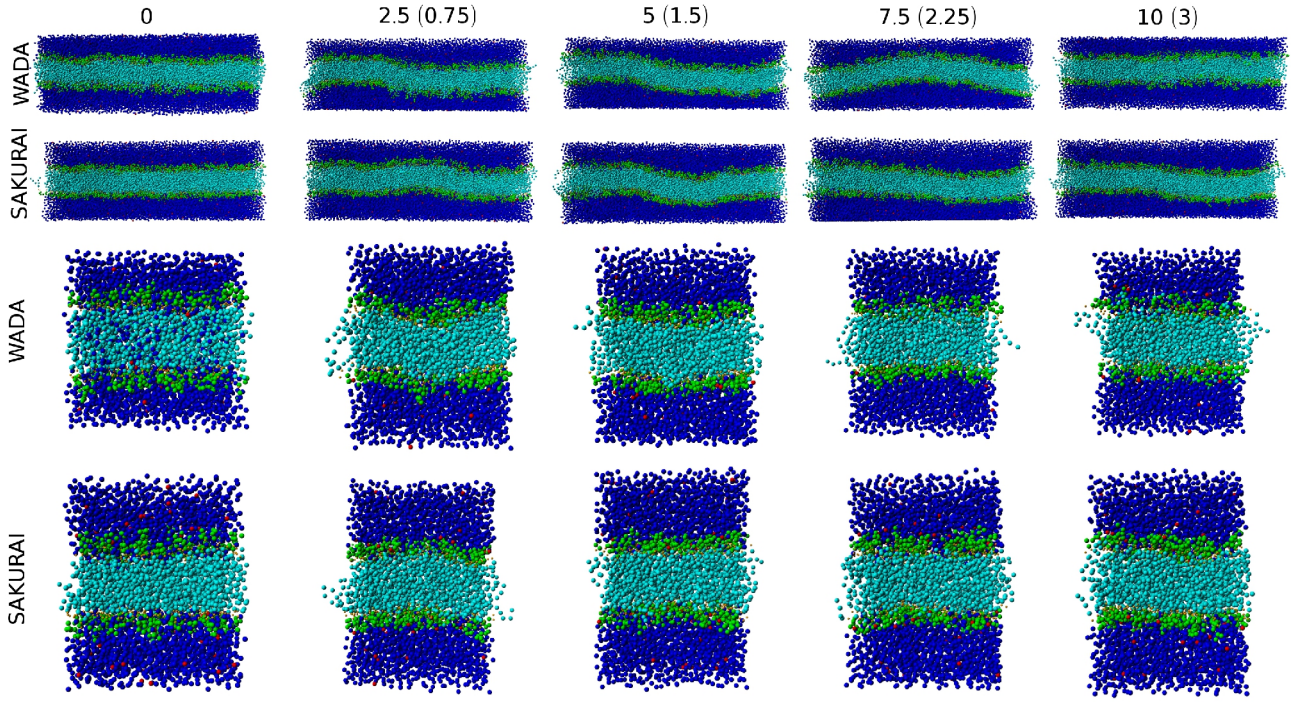


Figure 2.8: *The frames of all four simulations at the given time in μs (data in brackets for the shorter simulation of large membrane of Sakurai's composition). Undulation is obvious mainly for the bigger membranes. Water beads are shown in blue, ions are red, lipid headgroups are green, glycerol beads orange and tails cyan.*

calculated) and so the peaks belonging to the individual regions of the system are wider in the big membranes than the small ones. This undulation occurs since the very beginning of the simulations with the same extend over the whole simulation as was examined by a comparison of the overall density profiles for every nanosecond of the simulation (data not shown).

The widening of the overall electron density profiles aggrees with the membrane thickness determination for what these curves are used. As can be seen in table 2.11, big membranes are thicker than small ones and it was not that straightforward to determine the exact maxima on the overall electron density curve. Interestingly, although when determined from small membrane patches the system of Wada's composition is thinner then the Sakurai's one, while when the bigger patches are used for the membrane thickness determination the situation is opposite. This may point out, and is supported by a visual inspection of the membranes, that the Wada's membrane is more undulated than the Sakurai's one. Searching for an explanation, the various membrane composition is the most straightforward reason. The main difference is in the proportion of MGDG, that at a low level of hydration tends to adopt an inverted hexagonal phase, (more in Wada's membrane) and of charged lipids SQDG and PG (more in Sakurai's membrane).

The APL of Wada's membrane is a bit bigger than the one of the Sakurai's membrane (≈ 4% for both, big and small membrane), which can be explained by different APLs of individual

		APL [nm^2]	thick [nm]	trans [K]	phase (310 K)
Wada	*small*	0.643	3.7	225 – 230	LQ
	big	0.633	4.6		LQ
Sakurai	*small*	0.618	3.9	245 – 250	LQ
	big	0.610	4.4		LQ

Table 2.11: *APL, membrane thickness and phase transition temperature for the large (2400 lipids) and small (200 lipids) thylakoid membranes of Wada's and Sakurai's composition. Both values calculated for the large Wada's membrane are the same, if calculated in the range from 0.5 to 3 μs and after 5 μs, showing the membrane is already in the first interval in equilibrium.*

components of both membranes (see table 2.9). The difference of APL between big and small systems can be then explained by the membrane undulation.
The only major difference in the behavior of the membrane of Sakurai's and Wada's composition shown in table 2.11 is the phase transition temperature, that is for the membrane of Wada's composition of about 20 K lower that for the membrane of Sakurai's composition. Considering the phase transition behavior one must be aware of that the Martini model does not reflect this property well, but as the values for both membranes are affected by the same systematic error, their relative comparison makes sense.
A visual inspection of both big membranes did not show any tendency of a domain formation as expected according to the Israelachvili's theory described in chapter 2.1.1. The last frames from the two simulations of the big membranes are shown in figure 2.9.
Radial distribution functions (RDFs) were used to quantify potential preferences in lipid grouping and to examine if there are some short-time-range attractive forces, that drive certain lipids together. The RDFs for lipid headgroups of big membranes are shown in figure 2.10. Small membranes provide a similar view. In the Wada's membrane both neutral lipids prefer being surrounded by the charged ones, namely MGDG prefers the company of PG and DGDG is most frequently surrounded by SQDG. Charged PG prefers to be surrounded by neutral lipids, no matter its kind. Interestingly, the radial distribution function of SQDG around PG is higher than the one of PG. This can be explained by the division of the head-group of SQDG to a charged (bead S6) and a polar (other 3 beads) part and by the interaction of PG with the polar part of SQDG headgroup (for the confirmation see figure 2.11). The distribution of charged lipids around charged lipids (especially PG) shows another interesting thing, and that is that where the other lipids show the secondary maximum of the radial distribution function the curves for PG and SQDG exhibit a minimum while the tertiary maxima are higher than the maxima of the neutral lipids. As the APL of all four lipids is similar, this may mean, that mutual interactions of charged lipids are arranged by their interaction with an neutral lipid. The small peak on the position of the secondary maximum of SQDG may be caused by an attractive interaction of the polar part of the sugar headgroups. SQDG is the only lipid whose radial distribution function with the other molecules of the same kind of lipid is not the lowest but is the second highest (after it s interactions with DGDG). Looking at the RDF curves one must see the difference in the primary maximum of the distribution of DGDG around DGDG and the other curves. This double maximum may mean that there are two different modes of mutual DGDG interactions that have the maximum shifted with respect of each other. This may be related

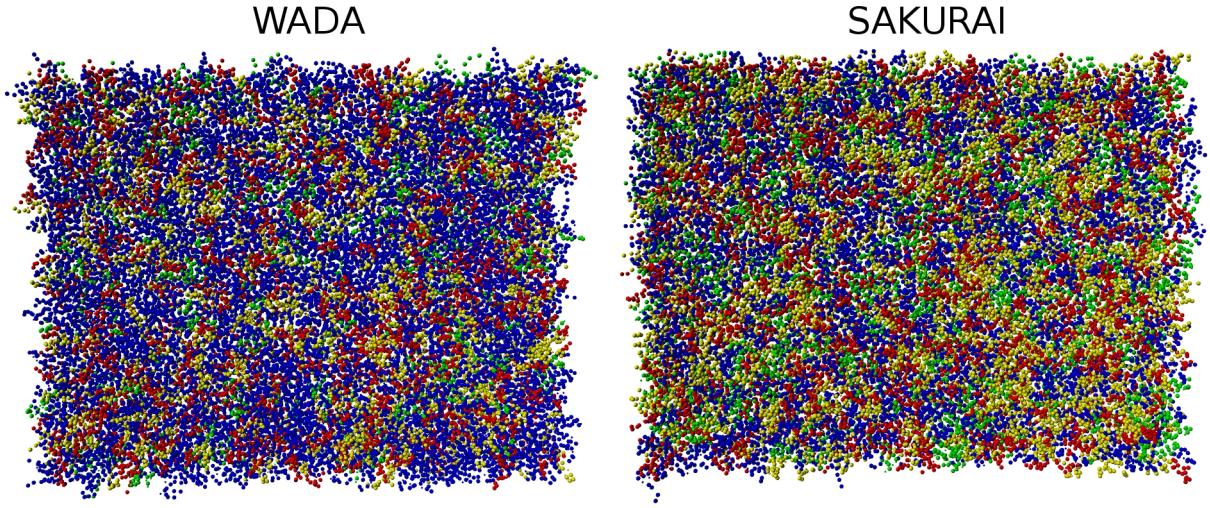


Figure 2.9: *Last frames of simulations of large membranes proves that there is no tendency in both systems to form domains. As the different lipids are shown in different colors (MGDG blue, DGDG red, SQDG yellow and PG green) these patches also demonstrate the difference in the composition of both membranes.*

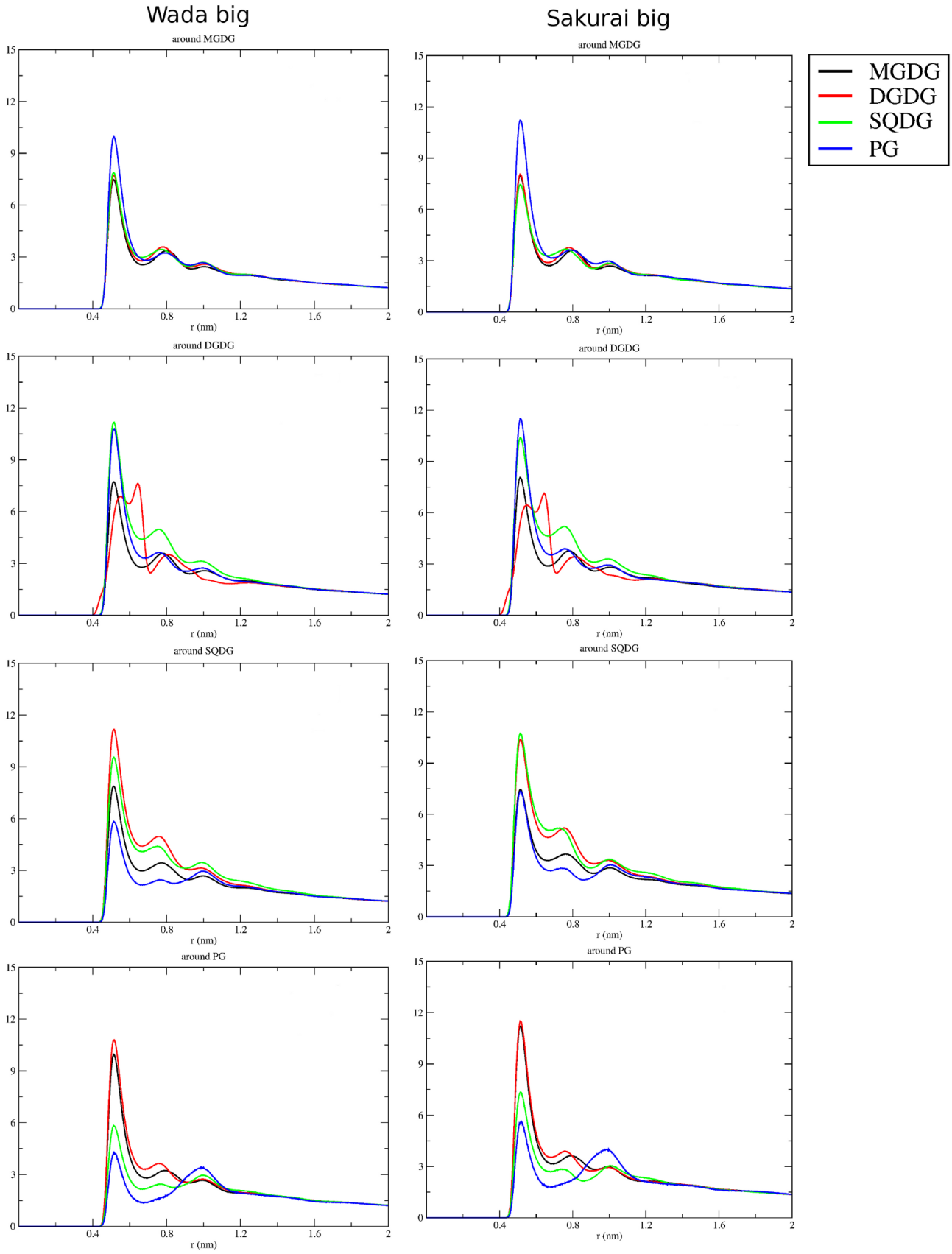


Figure 2.10: *RDFs of the lipid headgroups around the headgroup of another lipid for the large patches.*

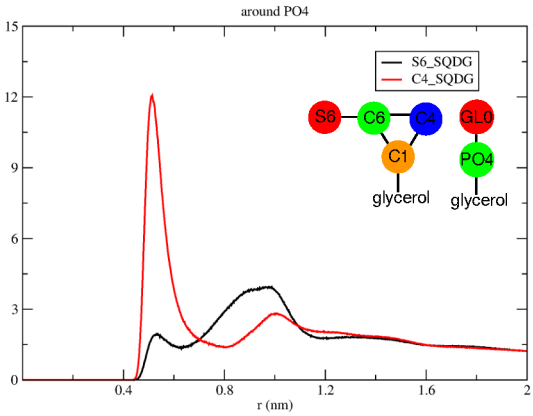


Figure 2.11: *RDFs of two distinct parts of SQDG represented by the beads S6 (ionic part) and C4 (polar part) around the PO4 bead (ionic) of PG show that SQDG is oriented with its polar part towards PG.*

to the 2 sugar residues making the headgroup of the DGDG and the two ways of their possible interaction with respect to the reference lipid.

RDFs for the Sakurai's membranes, that are in comparison with the above characterized Wada's membrane enriched with charged lipids on expense of mainly MGDG, show a bit different preference of the individual lipids. Here, PG is the most abundant lipid in the vicinity of MGDG and DGDG and the less abundant lipid around PG and SQDG. DGDG does not prefer its own company, maybe because its headgroups are that big that they need some smaller molecule to fill the space around them. This is supported by that DGDG is in both membrane compositions surrounded by lipids of small APL (PG and SQDG, that have very similar APL). MGDG with a bigger APL is in much less abundance around DGDG. PG is surrounded by molecules with bigger APL – MGDG and DGDG – than those with the smaller headgroup, what supports the previous hypothesis. Here we must consider the effect of the charge, discussed above. Another possible point of view on this issue is that the charged lipids try to be as far from one another as possible. The most effective way how to do that is an arrangement of the membrane, where the charged lipids alternate with the neutral ones. We may hypothesize, that if the membrane would contain even less charged lipids, they would be further with respect to each other than they are in the models presented here.

To conclude this session, thylakoid membranes the compositions simulated do not form membrane domains, as expected. Although the mutual preferences of various lipids depend on the membrane composition, there are some general trends common for both membranes (the size of the membrane does not influences this preferences). These are: PG prefers being surrounded by an uncharged lipid, with DGDG a bit favored than MGDG. The charged lipids have a depression on their RDF curves where the others have the secondary maximum, while their tertiary maximum (especially for PG) is higher than the tertiary maximum of uncharged lipids which indicated their accumulation in the 3rd shell around the lipid. The most likely candidate for the role of a linker between two molecules of PG is MGDG that is highly preferred to be surrounded by PG. Or in case of the less charged membrane of Wada's composition the second charged lipid, SQDG, may play this role, when there is not enough free PG. In the membrane of Sakurai's composition, where the proportion MGDG:PG is 2.3 instead of 7.4 like in the Wada's membrane, the space around MGDG is fully occupied by PG. The second charged lipid, SQDG, that has a smaller affinity to MGDG than PG does not manage to get close to MGDG and is repelled from the neutral lipid by an electrostatic interaction of PG. This makes SQDG to be the less abundant lipid in the vicinity of MGDG in the Sakurai's membrane. Also the second neutral lipid – DGDG – prefers to be surrounded by charged lipids, in case of Sakurai's membrane by PG, in case of the less charged Wada's membrane by both, SQDG and PG, with similar probability. This may be explained by the lack of free PG in membrane (this leads to a hypothesis that PG is more attracted to MGDG than DGDG). Both membranes show, that SQDG prefers either the company of DGDG and interestingly (especially for the Sakurai's membrane) its own proximity. MGDG – PG interaction is the most preferable interaction in the

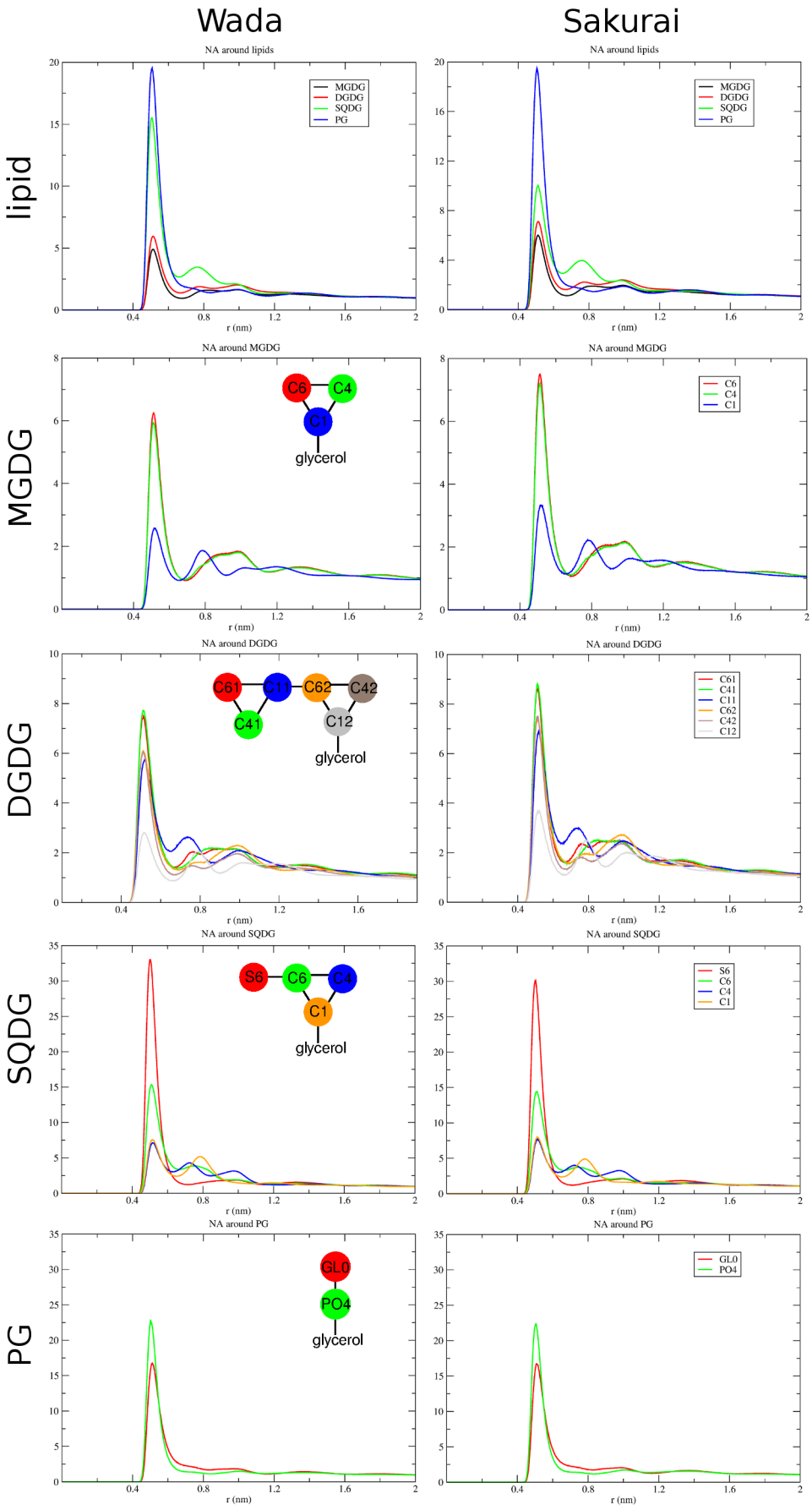


Figure 2.12: *RDFs of counterions around lipids and their individual beads for the large membranes.*

membrane. DGDG also tends to interact with PG, but as most of PG is busy in an interaction with MGDG it interacts with the second charged lipid – SQDG. In the vicinity of SQDG, DGDG or SQDG can be most likely found. The mutual competition between MGDG and DGDG to capture as much PG as possible, and the effort of PG and SQDG to drift apart, do not allow the formation of domains. An electrostatic repulsion between the charged lipids is probably the main force for the arrangement of lipids within the thylakoid membrane.
An effort to construct the most likely arrangement of lipids in the membrane failed, as it was impossible to arrange the lipids in order to reproduce mutual preferences of all lipid headgroups (according to the most favorable lipids on the positions on maxima of RDFs).
RDFs were also used to examine the distribution of counterions around individual lipids and their headgroup beads for big membranes of both compositions, see figure 2.12. Sodium in both membranes shows the same preferences to individual lipids and their beads, only neutral lipids in the membrane of Sakurai's composition has values of RDF a bit higher and interestingly there are less ions around SQDG in Sakurai's membrane. As expected, ions are much more attracted to charged lipids than neutral ones, with higher preference for PG than SQDG. RDF of sodium around PG differs from the others in the presence of the primary minimum where the other lipids have the secondary maximum. SQDG is the lipid with the highest secondary maximum of all the lipids. The RDFs for MGDG and DGDG are very similar to each other, with the secondary maximum lower than the tertiary one and with a very low secondary minimum.
Both curves for the RDF of sodium around the individual headgroup beads of PG show the same trend with an absence of the peak at the position where most of the other lipids have the secondary maximum. This can be explained by a repulsion of the other ions by the sodium that is present in the first shell around the lipid. It is not surprising that the PO4 bead attracts ions more then the GL0 bead.
The S6 bead of SQDG is interestingly the bead with the highest primary maximum of RDF of sodium. The curve exhibits the same shape as the curves for PG and its individual beads. The higher attractivity of this bead to sodium than the PO4 bead could be explained by the presence of other polar beads in its vicinity. The relatively high primary maximum of RDF of sodium beads around the C6 bead may be a result of the proximity of this bead to the previously mentioned S6 bead, contrary to which, its RDF shows the secondary maximum, where expected. RDFs for sodium beads around the C4 and C1 beads have a very similar behavior when concerning their primary maxima, that differ from the maxima of higher order in the way that the secondary maximum of the

	Wada – small				**Wada – big**			
	MGDG	*DGDG*	*SQDG*	*PG*	*MGDG*	*DGDG*	*SQDG*	*PG*
C1A – C2A	0.131	0.102	0.542	0.555	0.124	0.127	0.513	0.517
C2A – C3A	0.177	0.195	0.401	0.393	0.157	0.168	0.368	0.392
C3A – C4A	0.115	0.129	0.209	0.069	0.121	0.130	0.204	0.098
C4A – C5A				0.044				0.028
C1B – C2B	0.535	0.535	0.558	0.568	0.500	0.501	0.524	0.529
C2B – C3B	0.376	0.367	0.424	0.416	0.365	0.360	0.375	0.408
C3B – C4B	0.204	0.205	0.240	0.248	0.202	0.192	0.206	0.238

	Sakurai – small				**Sakurai – big**			
	MGDG	*DGDG*	*SQDG*	*PG*	*MGDG*	*DGDG*	*SQDG*	*PG*
C1A – C2A	0.133	0.163	0.583	0.569	0.121	0.127	0.551	0.557
C2A – C3A	0.170	0.175	0.445	0.491	0.172	0.188	0.416	0.431
C3A – C4A	0.163	0.163	0.254	0.111	0.139	0.151	0.241	0.124
C4A – C5A				0.037				0.040
C1B – C2B	0.553	0.555	0.595	0.599	0.534	0.541	0.564	0.560
C2B – C3B	0.424	0.453	0.438	0.464	0.409	0.403	0.426	0.450
C3B – C4B	0.250	0.250	0.256	0.307	0.233	0.225	0.215	0.271

Tables 2.12: *"Order parameters" for lipid tails in the examined membranes.*

C4 bead is with respect to the one of the C1 bead shifted a bit towards the origin of the coordinates. There is also a difference in the heights of the maxima of higher order, where the C1 bead has the highest secondary maximum of all the beads of SQDG, while the C4 bead has the highest tertiary maximum. As will be shown below, the height of the primary maxima of the S6 and C6 beads is similar to those of PG beads, while the other beads of SQDG – C4 and C1 beads – behave in this respect more similarly as DGDG and MGDG. This is in agreement with the above proposed hypothesis, that the headgroup of SQDG can be divided into two parts, one of which behaves as a charged lipid while the other one shows a behavior closer to a sugar.

For MGDG there are ions more likely to be found in the proximity of the beads C4 and C6, whose RDFs show a very similar course, with secondary and tertiary maxima flowing into one. The last C1 bead has the primary maximum of about half of the size than the other two headgroup beads and also its secondary and tertiary maxima are markedly separated.

A very similar behavior is observed for DGDG, where the C4 and C6 beads have a similar course for both galactose units, with both C1 beads having a lower number of sodium ions in their proximity. The height of the RDF curve of the C1 bead of the upper headgroup is at the level of RDFs of the lower headgroup, what may be connected with the presence of the polar beads of the second headgroup, that attract the ion. The C1 bead of the lower sugar, that links the headgroup to glycerol, is the one less attractive to ions. The curves exhibit the similar shape as the curves of MGDG.

Concerning the slope of the tail bonds to the z-axis (how is an order parameter in the CG simulations defined), shown in tables 2.12, the Sakurai's membrane is a bit more ordered than the Wada's one, which is likely to be caused by the higher number of less saturated lipids (SQDG and PG) in the former membrane as well as by the smaller extend of undulation. The undulation may effect the order parameter as it is computed in the way it tilts the membrane normal that is no more vertical to the z-axis of the simulation box and to which as an reference the slope of a bond is computed. It is under debate whether the overall undulation averages out. Generally, smaller membranes show more ordered lipid tails. As they are almost planar and so with very little undulation, this may be the explanation. In all the membranes, the sn-1 tails of MGDG and DGDG show much lower order parameters than the sn-1 tails of SQDG and PG. The main reason for this is obviously their high degree of unsaturation. Interestingly, identical behavior is not seen for PG, but the behavior of its saturated tail can be divided into two parts: the first two values show the behavior of the saturated tails, while the last two exhibit a random orientation of the given bond with respect to the z-axis. As described in the next chapter, PG exhibits the same behavior in the atomistic simulations as well.

The lateral diffusion coefficient (see figure 2.13 and table 2.13) is another value showing a different behavior for the membranes of different compositions. Absolute numbers are similar, with the lateral diffusion a bit faster in the less-charged membrane of Wada's composition. This is in agreement with the general knowledge, that the charged lipids in MD simulations slow down lateral diffusion [Zhao et al., 2008]. Generally, there is no difference between the speed of the lateral movement of lipids in the big and small patches (as expected). The main difference between lateral diffusion of the membranes of the two different compositions is in grouping lipids with similar lateral diffusion coefficients. In the less-charged Wada's membrane, MGDG moves faster than the other three lipids. On the other hand, in the membrane of Sakurai's composition, PG joins MGDG in

	Wada		**Sakurai**		
	small	*big*	*small_3*	*small_10*	*big*
LPMG	6.2	6.1	5.1	6.0	4.9
LPDG	4.9	5.2	4.1	5.6	4.2
DPSQ	5.2	5.4	4.0	5.3	4.3
IPPG	5.0	5.1	5.2	6.1	4.9

Table 2.13: *Lateral diffusion coefficients (10^{-5} cm^2/s) for membranes simulated. Only the small membrane of Sakurai's composition shows different values when computed between 1.9 – 3.0 μs and 8.9 – 10 μs.*

the group, that moves faster than DGDG and SQDG. Interestingly, the outer leaflet of the thylakoid membranes (of higher plants) is relatively enriched by PG and MGDG while the inner one contain relatively more DGDG and SQDG. It was reported [Dietrich et al., 2001; Filippov et al., 2004] that the lateral diffusion in membrane rafts is slower that outside of the domain. We may hypothesize, whether this different lateral diffusion coefficients may be joined with this effect, but there is some stronger force (probably of an electrostatic nature) acting against the domain formation that overweights in general.

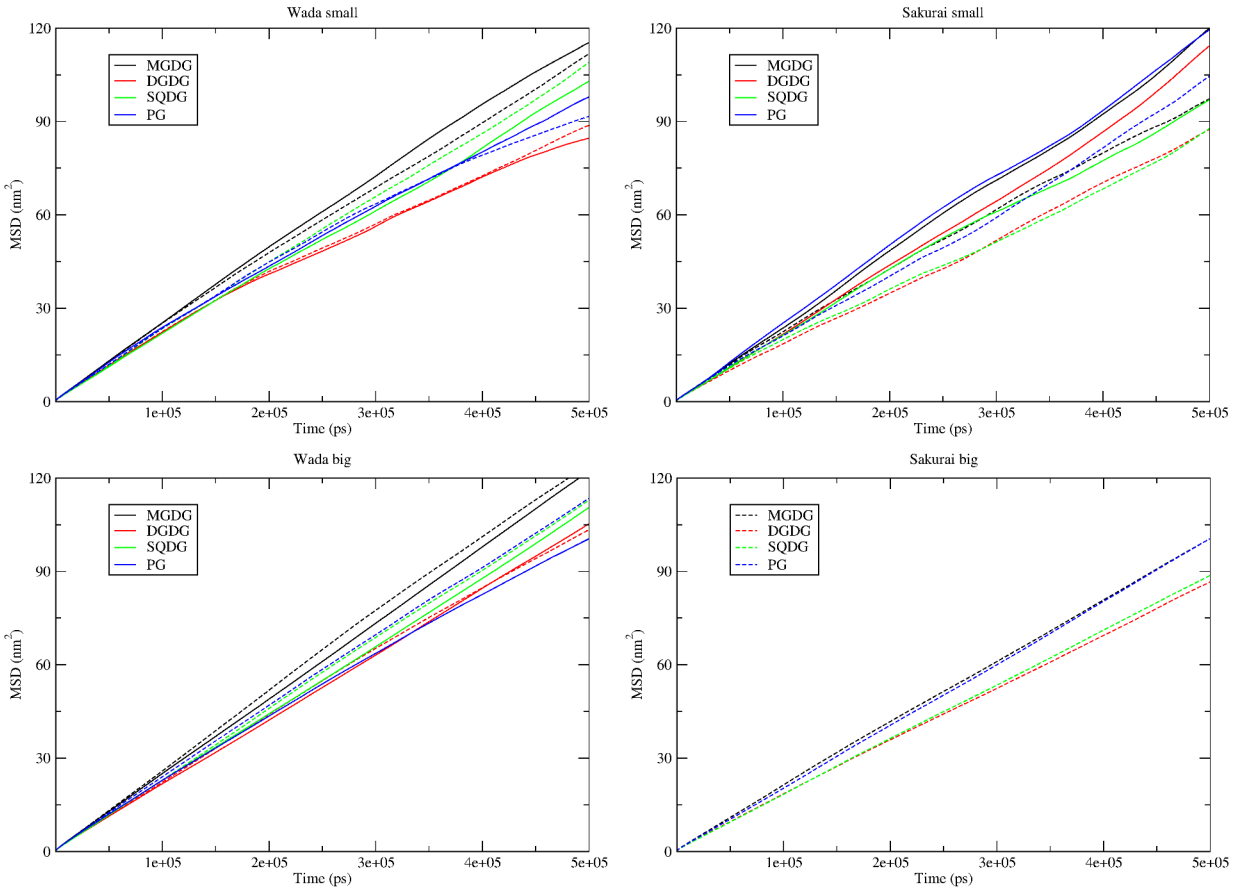


Figure 2.13: *MSD for all simulations between 1.9 – 3.0 μs (dashed line) and 8.9 – 10 μs (full line) that served to determine the lateral diffusion coeffiecient. The large membrane of Sakurai's composition was simulated for 3 μs only.*

2.4.4 Thylakoid membrane – atomistic simulations and their comparison with CG simulations

Atomistic simulations of thylakoid membranes of both, the Wada's and Sakurai's composition, form a bilayer in LQ phase at 310 K (for some frames of the simulations see figure 2.14). Their APL and membrane thickness (i.e. the properties based on which the CG membranes were parametrized) are in excellent agreement with CG simulations of the membranes of the same size (referred to as small membranes) as can be seen in table 2.14. The membrane thickness is identical for the given membrane composition, while APL differs by 2% resp. 1.2 % for the Sakurai's resp. Wada's membrane. This can be considered as another proof of the correctness of the developed CG parameters for thylakoid glycolipids. The new information, provided by this comparison of atomistic and CG thylakoid membrane simulations, is that not only properties of pure membranes, but also properties of the mixtures are in the mutual agreement. This shows the correct capturing of intermolecular interactions among various lipids. Similarly to the CG simulations the Sakurai's membrane has smaller APL (of about 5 %) and the bigger membrane thickness (of about 5 %) when compared with the Wada's one.

		APL [nm^2]	thickness [nm]
Wada	*atomistic*	0.635	3.7
	CG	0.643	3.7
Sakurai	*atomistic*	0.605	3.9
	CG	0.618	3.9

Table 2.14: *APL and membrane thickness of the atomistic and small CG membranes show excellent agreement.*

The electron density profiles for both membranes are shown in figure 2.15. The widening of curves of Wada's membrane is connected with a movement of the bilayer in the water bath, as can be seen from the picture of individual simulation frames (figure 2.14) and from the shift of the electron density profile of lipids for different time frames, shown in figure 2.16. The electron density profiles show the decrease of water towards the center of the membrane, especially in the hydrophobic tail region of the membranes. Lipids are oriented with their tails towards each other and headgroups on the interface with water. The higher maxima on the electron density profile of

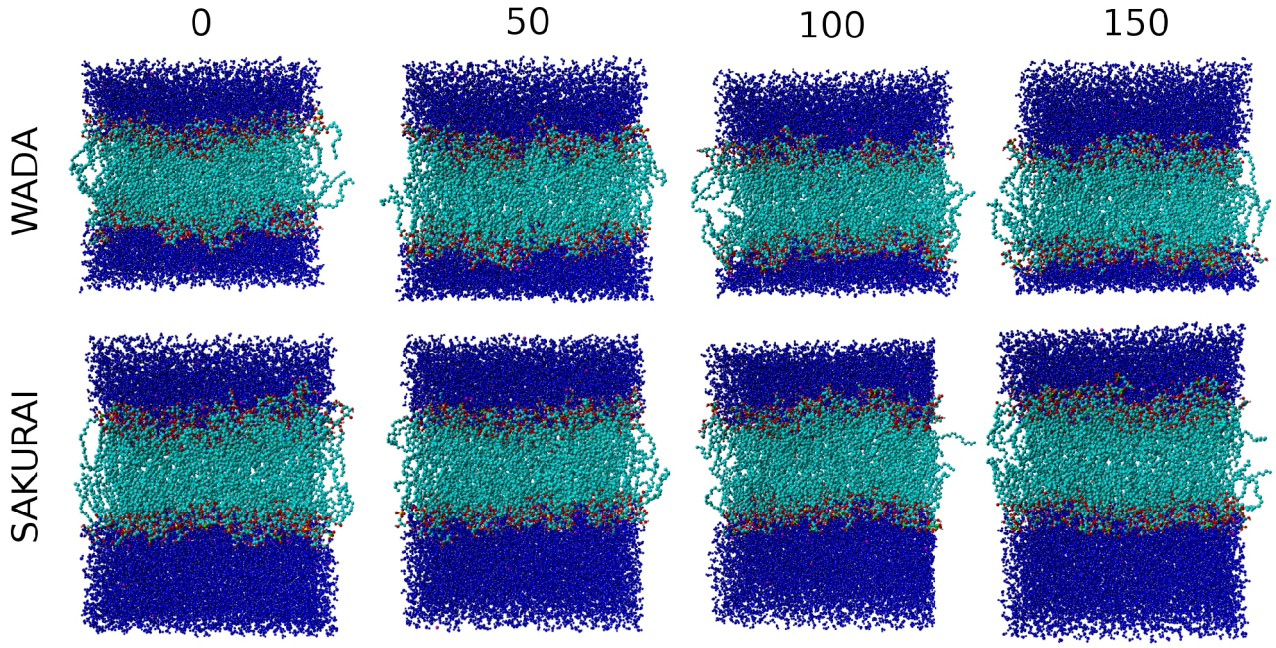


Figure 2.14: *Frames of atomistic simulations in given times (in ns). Notice the shift of Wada's membrane in the last frame. Water molecules are shown in blue, ions in magenta, in lipids carbon is highlighted cyan, oxygen red, hydrogen gray, sulfur green and phosphorus yellow.*

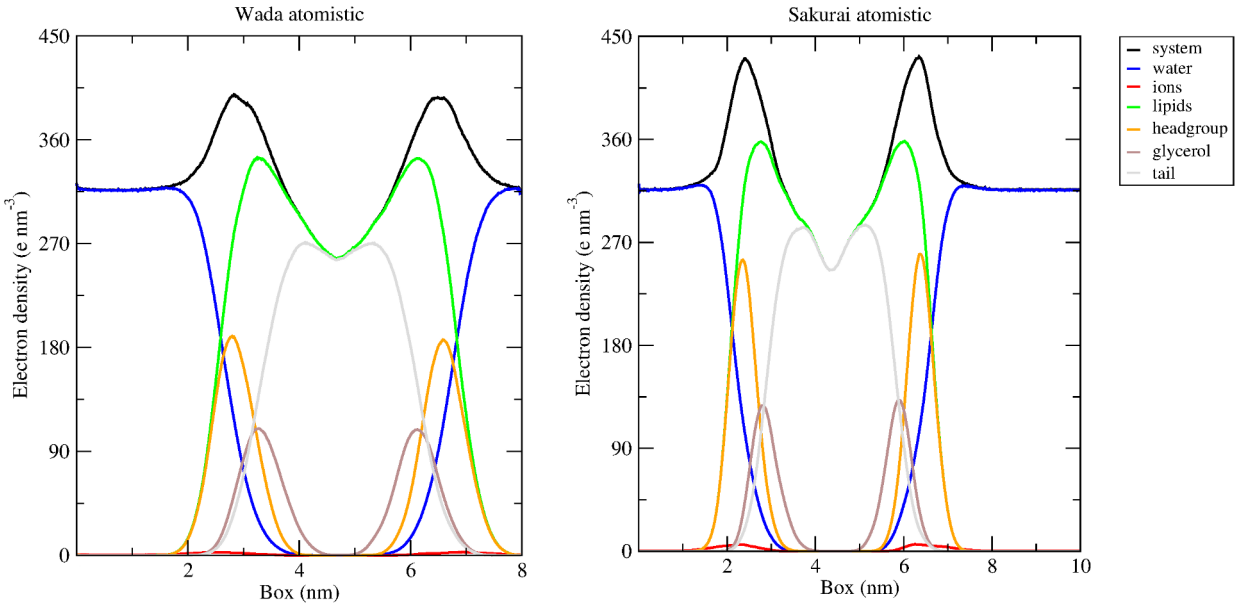


Figure 2.15: *Electron density profiles of the atomistic simulations after 60 ns.*

the whole system for the Sakurai´s membrane than for the Wada´s one are caused by the higher level of hydration of the former system. The only major difference between the atomistic and the CG electron density profiles (shown in figure 2.7) is the distribution of ions with respect to the membranes. The maxima of the curves for ions are at the same position as the maxima for lipids in the CG simulations, while in the atomistic simulations the maxima of ions are a bit “above” the maxima for lipids. This means that in the CG simulation the surface of the membrane is neutralized by ions, while in the atomistic simulation there are two regions of charges of opposite polarity. Although the later system is highly dynamic, it may in some extend behave as a capacitor. This behavior of the atomistic simulations is in this respect in agreement with the theory [Israelachvili, 1973] and we must keep in mind that in this respect the CG simulations do not reflect the reality well. However, it should be discussed whether the correct reproduction of this has any influence on the long-time-scale effects generally studied in CG simulations. My humble guess is no...
MSD curves, that are used for determination of lateral diffusion coefficients are shown in figure 2.17. As MSD for all lipids is so small, especially for the membrane of Wada's composition, that in the given time a lipid does not move over the distance of even its own size, I conclude, that it is impossible to determine the lateral diffusion coefficients of lipids in the atomistic simulations of the thylakoid membranes. The other thing the MSD curves show is that there is faster lateral diffusion

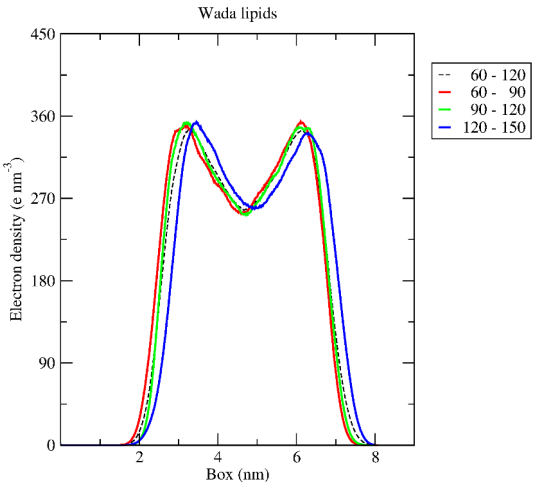


Figure 2.16: *Electron density profiles of lipid bilayer in Wada's membrane show a shift in the simulation cell.*

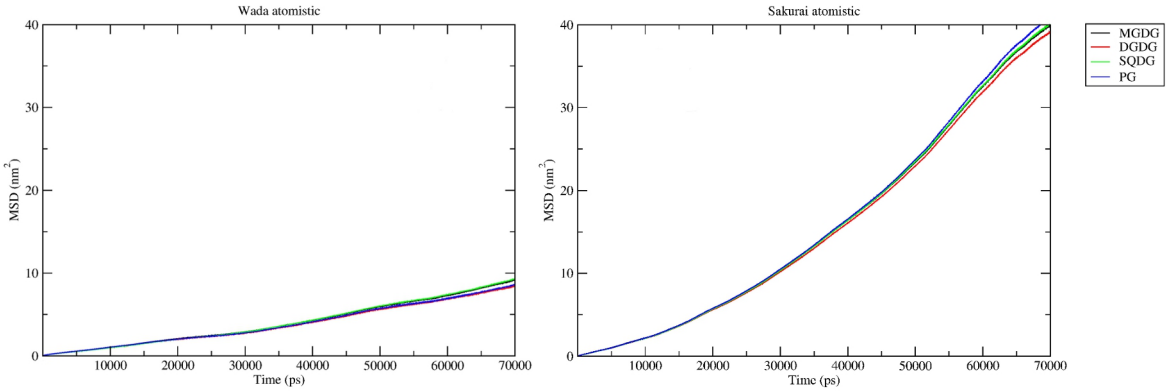


Figure 2.17: *MSD curves for individual lipids in the atomistic membranes determined after 65 ns of simulation.*

in the membrane of Sakurai's composition than of the Wada's one, that is in contrast with the CG simulations of the given membranes, where the lateral diffusion coefficients were determined from much longer simulation and far later from the start of the simulation. The longer simulations that might help to solve this issue were not performed, due to their computational time demands. Therefore the radial distribution functions of lipids with respect to another lipid were not calculated as it is not expected that any tendency on grouping could be observed in this relatively short time and for lipids of very slow lateral diffusion. On the other hand the last frames of the simulations show a different lipid arrangement than the initial membrane set-ups, see figure 2.18, what points onto some lateral diffusion and maybe also some gathering tendency is present in these membranes.
RDFs were used to quantify a possible preference of sodium ions interactions with oxygen atoms of various lipids. Oxygens were chosen as they bear a negative partial charge and so they are more likely to be involved in an interaction with the positively charged sodium ions than other atoms bearing the positive partial charges. To be able to compare these values with the values obtained by the CG simulation, RDFs of oxygens making one CG bead were summed. The resultant graphs are shown in figures 2.19 and 2.20. Apart from several exceptions discussed further down, both membrane compositions show the same distribution of sodium ions around individual oxygen atoms.
For MGDG the O6 oxygen is the most attractive atom for sodium ions. Its RDF is typical by a very high primary maximum and a very low secondary maximum, maxima of higher order are not obvious. The RDF of sodium around the O4 oxygen has a lower primary maximum than the previous one, but it has a higher secondary and tertiary maxima, that are well defined. As these two atoms are in the CG model mapped into one bead, designed as C6, it is not surprising that this bead is the one most attractive bead to sodium ions. In the CG simulation, there is the C6 bead not the only bead preferred by ions. A similar number of sodium ions is found around the bead C4, made of the O2 and O3 oxygens. Although oxygens O2 and O3 in the atomistic simulations show much lower primary maxima of RDF, oxygens O4 and O6, and their secondary and tertiary maxima are of the same height as for the first two atoms. The atom O3 in both membranes exhibits a marked peak inbetween where RDFs of the others atoms have the secondary and tertiary maximum. This may be caused by an interference of the sodiums grouped around the neighboring O4 (and maybe also O6) atom. The higher preference of sodium ions to the oxygens making bead C6 than to the oxygens forming bead C4 in atomistic simulations may be explained by a preferred conformation of the MGDG headgroup in the atomistic simulations, which is not captured in the CG simulations. Atoms O1 and O5 do not show any primary maxima at all, what means that sodiums are not primarily attracted to these oxygens (they are not alcohol oxygens like the previous ones). The maxima of the higher order are probably made of sodium ions primarily attracted by some other atoms. The absence of the primary maxima on these curves of oxygens mapped into the C1 bead is the major

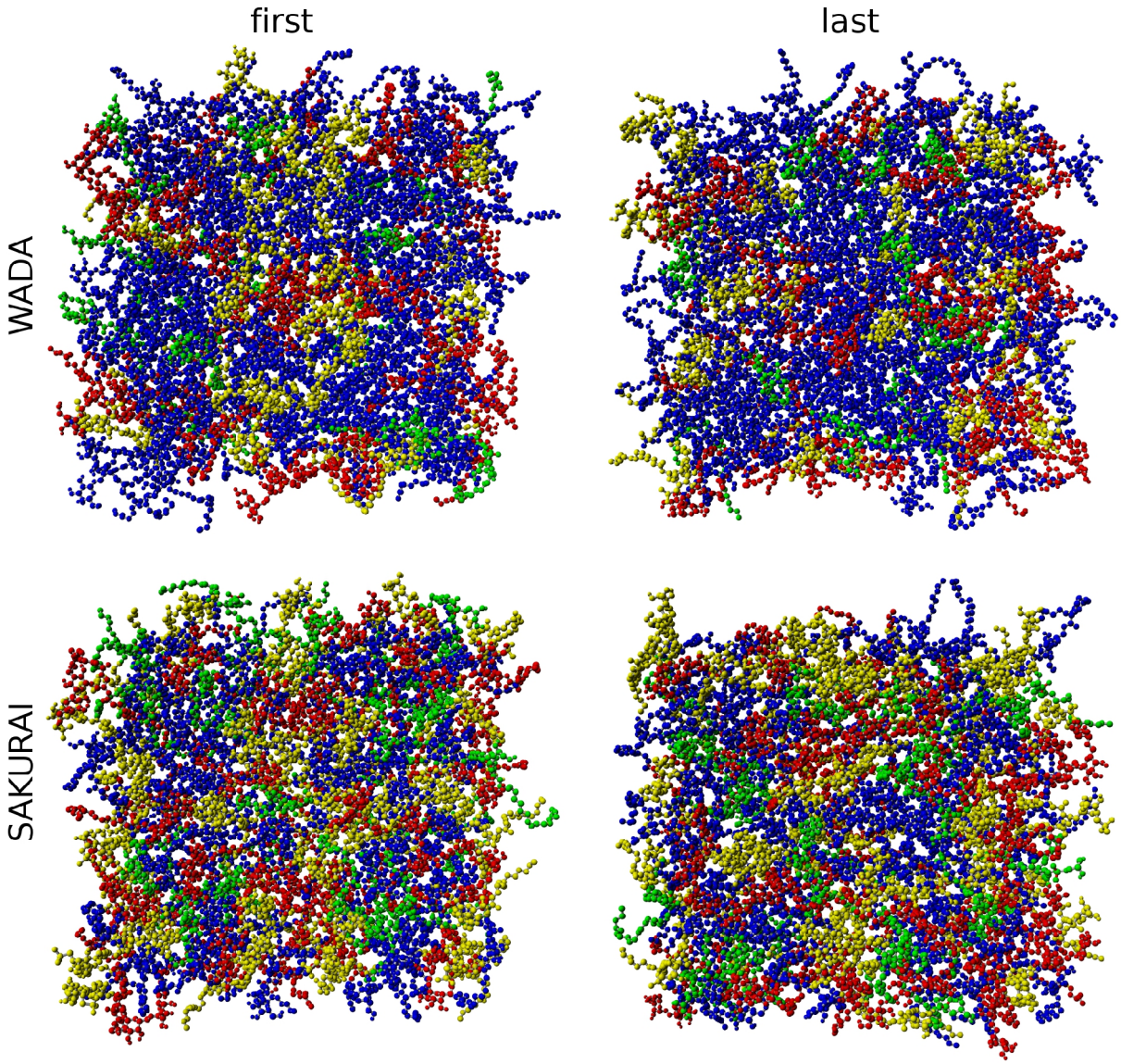


Figure 2.18: *First (after equilibration runs) and last (at 150 ns) frames of the atomistic simulations. MGDG is shown in blue, DGDG in red, SQDG in yellow and PG in green.*

difference between the the atomistic and CG simulations, where this bead has some sodium ions in its primary shell.

A preferred conformation of the sugar headgroups seem to influence the RDFs of sodium ions around the oxygen atoms in the headgroup of DGDG as well. Although the distributions do not exhibit the same preferences for both membranes, there are some trends these membranes have in common. Similarly to MGDG, DGDG also shows that the non-alcohol oxygens (O51, O52, O11, O12) are not that attractive for sodium ions as the alcohol ones. This is demonstrated by the absence of the primary maxima on the RDF functions of sodium ions around these oxygens. Although maxima of higher order are present, they may be the result of clustering of the ions around other oxygens. As oxygens O26, O12 and O13 are the most attractive oxygens for the sodium ions, we may hypothesize, that in the preferred conformation of DGDG headgroup these atoms are exposed to the solution. A relatively little number of sodiums around the O42 oxygen can be explained by its vicinity to the O62 oxygen, that is not directly bound to the saccharide ring, but is linked to it via the carbon C62 and may have a bigger radius in which it may capture ions in solution. As in the shade of the O62 oxygen, there are not so many sodium ions available, that can interact with the O42 oxygen. A similar effect may be involved in the relatively low primary maximum of the O14

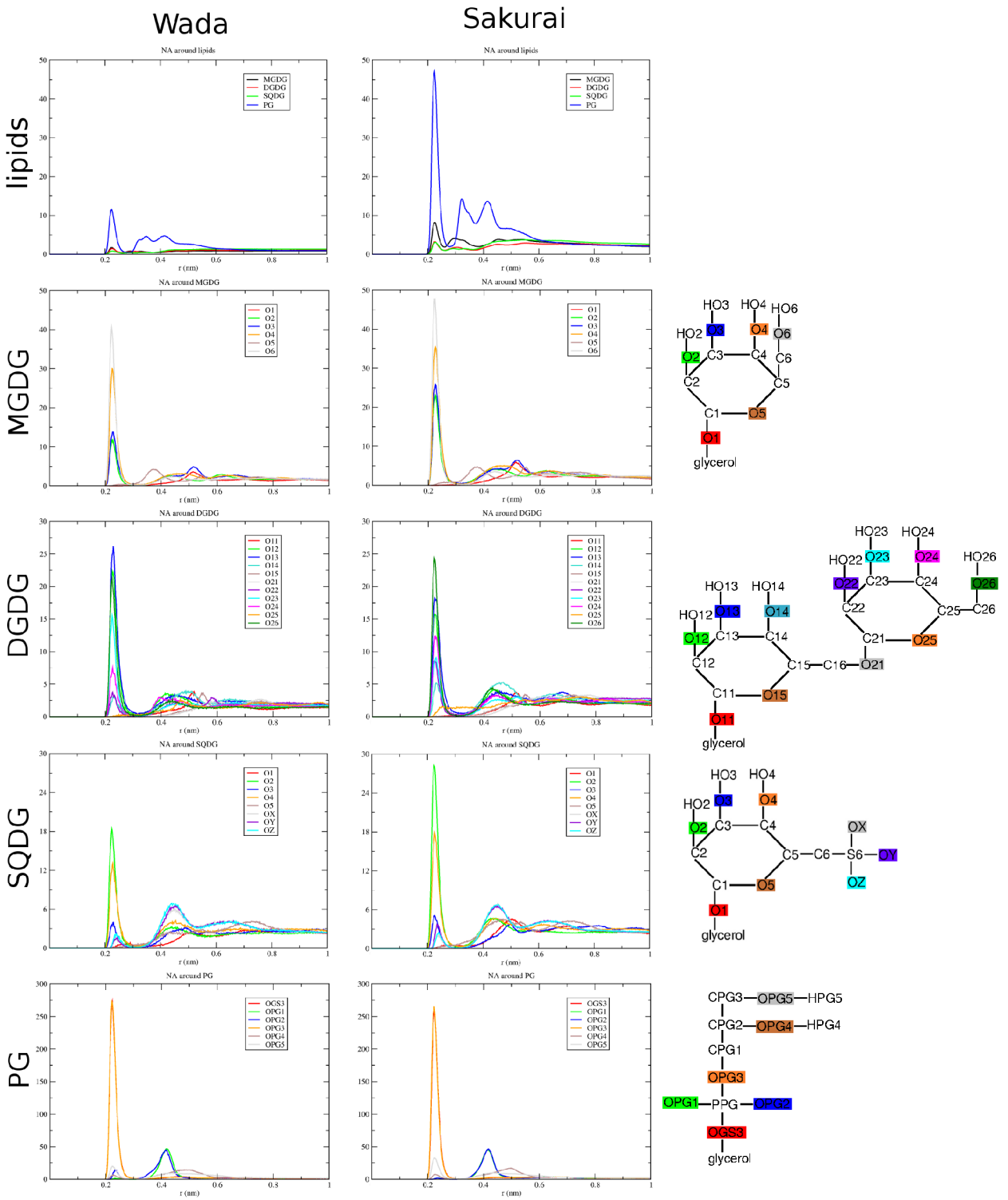


Figure 2.19: *RDFs for atomistic simulations.*

oxygen. The shift of the relatively high secondary maximum supports this hypothesis. Differences between the individual membrane compositions appear when the RDFs are considered with respect to the atoms summed into the beads. While in the more charged membrane of Sakuari's composition oxygens mapped into the beads C42 (O12 and O13) and C41 (O24 and O26) show primary maxima of similar height, in the membrane of Wada's composition the primary maximum of bead C42 is considerably higher than the next maximum of bead C41. This could be explained by the fact that atoms making bead C41 are involved in interactions with several other atoms from the headgroup region of the membrane or by a different conformational preferences of the headgroup of DGDG in

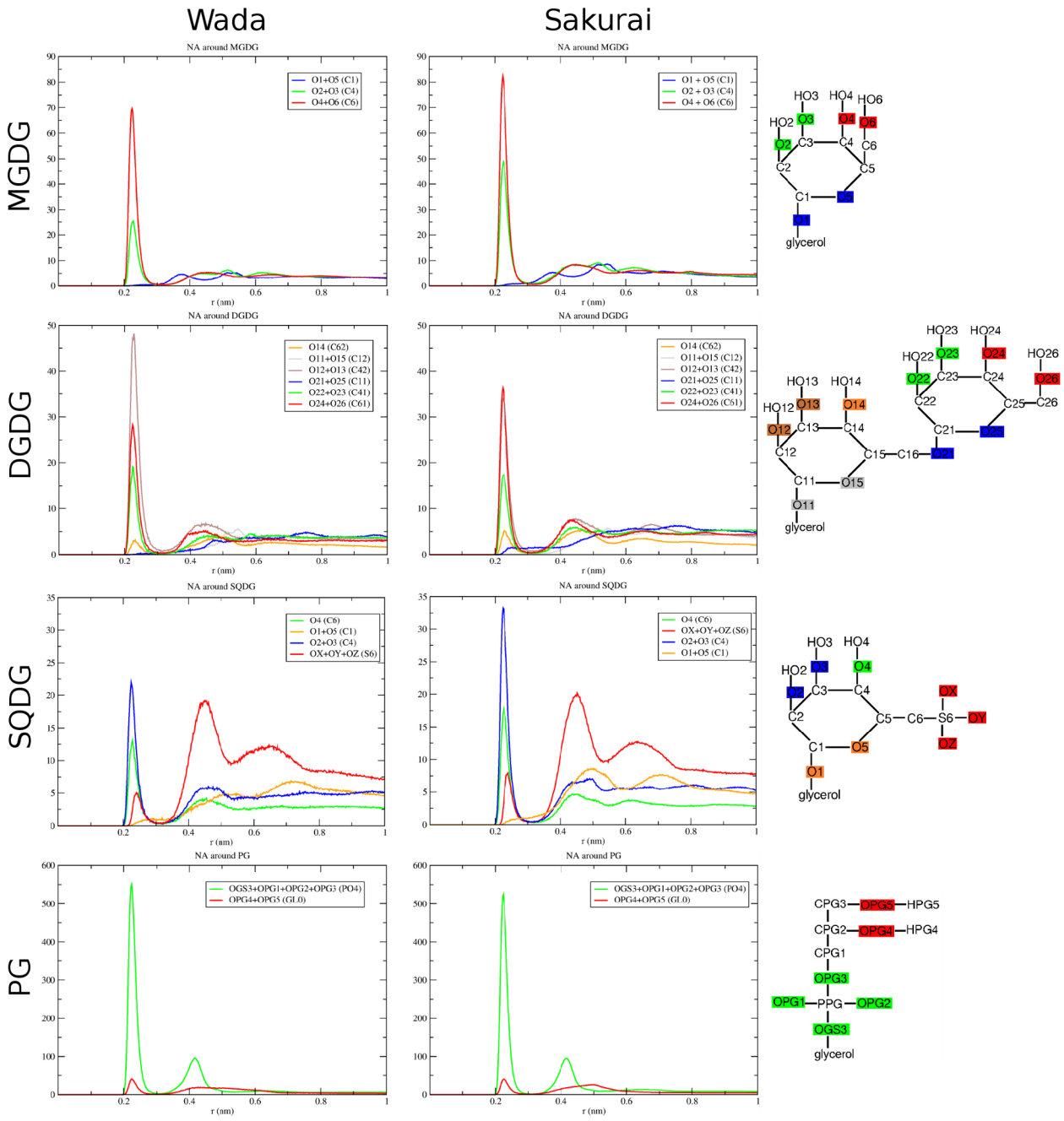


Figure 2.20: *Sum of RDFs of sodium ions around oxygen atoms of lipids to represent beads.*

the membranes of different compositions. It is not surprising that beads C11 (O52 and O12) and C12 (O51 and O11) show RDF without primary maxima, as curves from which they are made do not have the primary maximum as well. In both membrane compositions the bead C62 (made by oxygen O14) is much lower then bead C61 (O22 and O23) as it is made of one oxygen only, that by its own is not very attractive for sodium ions.

RDFs of sodium ions around the headgroup oxygens of SQDG are interesting in the way that sodiums are more preferably found in the first shell around the saccharide oxygens than oxygens of the SO_3^- group. The maxima of higher order are higher for the SO_3^- group than for the alcohol oxygens. Like for the other lipids, the atoms O1 and O5 do not exhibit any primary maxima of their RDF curves. The maxima of the higher order are obvious, but is under debate, if this is not just a result of a proximity of the other atoms. Summing up RDFs obtained for the individual oxygens into the beads results in curves with higher primary maxima for the saccharide part of the molecule

while maxima of the other orders are higher for the SO_3^- group. This is in contrast with the RDFs obtained from the CG simulations where ions are over the whole course of the distances attracted more to the S6 bead than to the saccharide beads. As for the other lipids with the sugar headgroup, the bead made by non-alcohol oxygens is not well described in the CG representation.
Sodium ions are most likely found around the OPG3 and OGS3 atoms of PG. The height of the RDFs of sodium ions around these atoms is much higher than the other curves. The oxygen OPG5 has the higher primary maximum of RDF curve than the other alcohol oxygen of PG, designed as OPG5. As PG is in the CG representation mapped into two beads only and mainly as the two oxygens with that high RDF are grouped within one bead it is not surprising that after summation into the CG representation the PO4 bead (phospho group) shows much higher RDF than the GL0 bead (the rest of the headgroup). The only difference between the curves obtained by the atomistic and CG simulation is the (relative) height of the curves.
The deuterium order parameters for the tails of the lipids (separately for each lipid to be consistent with the CG simulations) are shown in figure 2.21. All graphs show the different behavior of the saturated (at the second carbon of glycerol, both tails for SQDG) and unsaturated tail (first carbon of glycerol, not for SQDG) of the lipids. The saturated tails show the order parameter that is close to those referred to in phospolipidal membrane simulations. Saturated tails of MGDG, DGDG and PG show a more obvious “odd-even effect” [Vermeer et al., 2007] than those of SQDG. This may be caused by the presence of the second unsaturated tail. The unsaturated tails, especially those of galactolipids MGDG and DGDG, show order parameters absolutely different from the ones of the saturated tails. The order parameter for DGDG even starts with negative values (and for MGDG with values close to zero), that is a behavior that was not reported before (as I am aware). The other remarkable feature of the curves are the depressions in the order parameters on the positions of the carbons 10 and 13 (PG exhibits the same behavior) and so they are connected with the presence of double bonds, as expected. One would expect the same deep depression at the position of the carbon 7 (the second carbon of the first double bond), but the (much shallower) minimum is at the position of the carbon 8. Interestingly, the similar minimum at this position can be seen at the curve for PG, that has only two double bonds in its saturated tail. I have no explanation for this type of behavior. The unsaturated tails of PG behaves on its start as the saturated tails and approaching to where its double bonds are its adopts the behavior of the unsaturated tails MGDG and DGDG. We may discuss in which expend the behavior of the unsaturated tails of MGDG and DGDG is influenced by the behavior of the last unsaturated tail of PG as it seems that the depression at the carbon 8 that is in common for all the three lipids is a result of the “odd-even effect” of PG. This would mean that the ordering tendency of the tails of MGDG and DGDG has bigger effect than has the double bond on the 6th position and also that PG participates in ordering of the other lipids. The membranes of both the compositions show the same trends in the order parameter behavior.

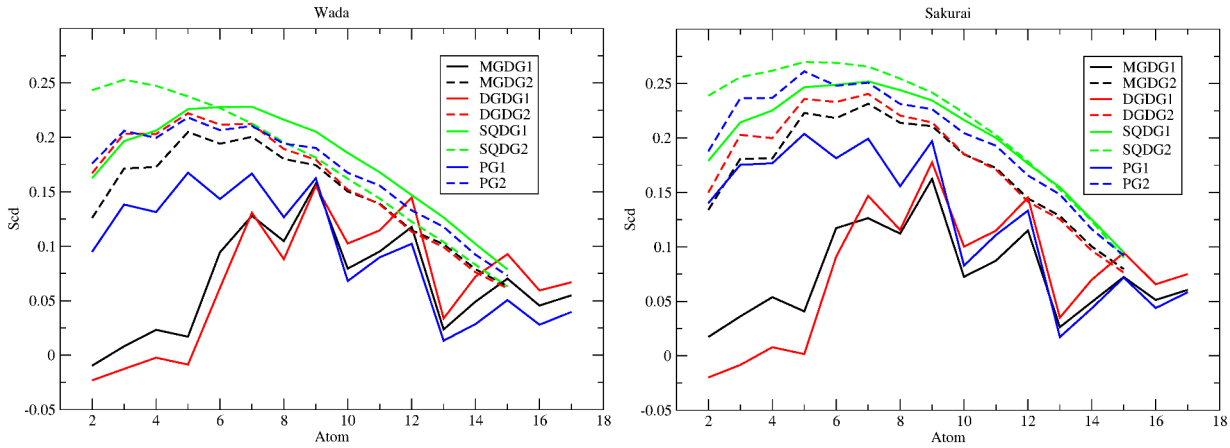


Figure 2.21: *Deuterium order parameters for lipids in atomistic simulations. Fatty acid bound to sn-1 carbon is shown in full line, while the one on sn-2 carbon in dashed line.*

2.5 Conclusions

Martini force field parameters that enable running a simulation with a timestep of up to 40 fs were developed for the thylakoid membrane glycolipids MGDG, DGDG and SQDG. These parameters, based on the atomistic simulations of the pure DS glycolipids in Gromos force field, reproduce the APL and membrane thickness and for SQDG also the phase transition temperature in excellent agreement with the properties obtained from atomistic simulations for both, the pure DS glycolipids as well as for the model of the thylakoid membrane from the cyanobacterium Synechocystis. Contrary to this, my CG model of the glycolipids was not able to capture the highly disordered behavior of the tail bound on the first carbon of glycerol of MGDG and DGDG (that probably has to do with the interaction of the tail with the headgroup region), what may be solved by a change of the bonded parameters for the first tail bead of these lipids. The other disagreement between the atomistic and CG simulations is in the behavior of counterions at the membrane-water interface. While in the atomistic model the ions prefer to stay above the membrane, in the CG model they prefer to enter the headgroup region of the membrane, what has an effect onto the shape of RDFs of sodium counterions around the individual atoms resp. beads. The other effect influencing the RDFs of sodium ions around lipids is some preferred conformation in which their headgroups are. As expected the lateral diffusion is slower for the atomistic simulations than for the CG ones.

Models of the thylakoid membrane from Synechocystis of two different compositions reported in the literature were compared with each other. Generally, it can be said, that these two models show only very little differences in membrane properties such as APL and membrane thickness and also in the properties connected with electrostatics. In fact, the only difference found is a different lateral diffusion coefficient and the phase transition temperature, that as discussed in the previous text is born with a systematic error.

None of the membranes examined forms membrane domains, however, there are certain preferences in the lipid arrangement. The main driving force hereby is a mutual repulsion of charged lipids resulting in a preferred alternation of charged and neutral lipids. SQDG, as a lipid of saccharide nature with a charged group attached, exhibits a bivalent behavior in the way that its charged part (S6 and C6 beads) behaves like a charged PG lipid, while the behavior of its second part (beads C4 and C1) is closer to MGDG and DGDG. This explains why PG behaves like a “more-charged” lipid than SQDG, as the former one does not have the sugar part that from one side shields the charge. Results also show that the charged lipids prefer interactions with MGDG rather than with DGDG.

3.1 Oxygenic photosynthesis

Oxygenic photosynthesis* (further only photosynthesis) is a process that approx. 2.2 billions years ago enabled organisms to colonize the land. It is because its side-product, oxygen, saturated the atmosphere of the Earth and made it breathable and the ozone (O_3) layer become thick enough to prevent the surface of the Earth from the ultraviolet light [e.g. Palinska, 2008; Gounaris et al., 1986]. However, the main purpose of photosynthesis is energy production [e.g. Gounaris, 1986 et al.; Ort and Kramer, 2009], that is useful not only for phototrophic organisms itself but also by eating their bodies for herbivorous and omnivorous animals.

Photosynthesis is a light-driven process that synthesizes energetically rich saccharides from carbon dioxide and water via light-driven reactions (for review see e.g. [Ort and Kramer, 2009]). Not all organisms, but only green plants, algae and cyanobacteria, are able of oxygenic photosynthesis. In eukaryotic organisms, there is a specialized organelle called chloroplast, in which photosynthesis takes place.

Photosynthetic processes can be, with respect to the necessity of light, divided into two part: a light and a dark phase. In the later one, atmospheric oxygen is fixated into energetically rich saccharides using NADPH and ATP yielded in the light phase of the photosynthesis. As my work does not concern this part of the photosynthesis, it will not be described here in a detail.

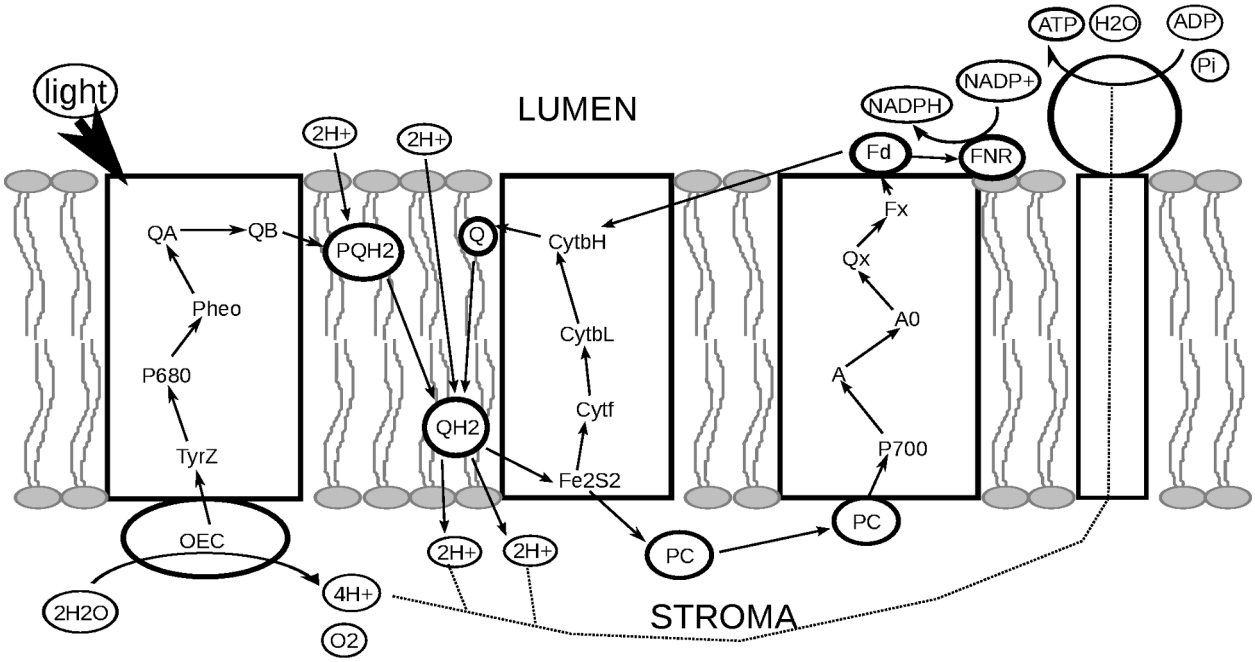


Figure 3.1: *Scheme of photosynthesis. Adapted according to [Singhal et al., 1999]*

The main aim of the light-dependent part of photosynthesis is to capture sunlight and convert it into energetically rich molecules, ATP and NADPH, with water as the input. To do that, four protein-cofactor complexes are necessary. The first of them is photosystem II, a water:plastoquinone oxidoreductase, that is described in more detail in the next chapter. Electrons then continue via plastoquinone to the cytochrome *b6f*, that may either in a cyclic electron flow, resulting in NADPH synthesis, act like a ferredoxin:plastocyanin oxidoreductase or in an acyclic electron flow work like a platoquinol:plastocyanin oxidoreductase. Electrons of the acyclic flow are then excited by the next plastocyanin:ferredoxin oxidoreductase, called photosystem I, where they are excited and are finally used for ATP synthesis by the last engine, called F0-F1 ATPase (for review see e.g. [Gounaris et al., 1986; Ort and Kramer, 2009]. For a scheme of the photosynthetic electron flow see figure 3.1.

**__anoxygenic photosynthesis__: Apart from the above described oxygenic photosynthesis using water as a source of electrons, an other compounds such as H_2, H_2S and various organic acids can be used. Then we deal with the anoxygenic photosynthesis.*

All four complexes are embedded in the thylakoid membrane on which an electrostatic gradient, necessary for the ATP synthesis, is created. As the subject of my study, this membrane is described in more detail in chapter 2.1.1.

3.1.1 Photosystem II

Photosystem II (PSII), the first supercomplex of the photosynthetic apparatus, is a water:plastoqoinone oxidoreductase, where two water molecules are split into 4 electrons, 4 protons and an oxygen molecule, that is released into the environment (reviewed e.g. by [Govindjee et al., 2010]). An electron obtained from water oxidation via the Kok cycle [Kok et al., 1970], is transferred to the tyrosine Z (Y_z, Tyr161 in *Thermosynechococcus vulcanus* [Umena et al., 2011]) of the D1 protein and further to the pair of chlorophyll *a* molecules, designed as P680, and to the molecule of pheophytin. The electron then finally gets via plastoquinone Q_A to the last part of its way associated with PSII, to a mobile plastoquinone Q_B by which it is transferred to the next photosynthetic complex, the cytochrome *b6f* (for review see e.g. [Govindjee et al., 2010; Britt, 2001]). For graphical illustration of the electron pathway in the PSII see figure 3.1 above.
The overall chemical equation of a reaction associated with the PSII can than be written as:

$$2H_2O + 2PQ + 4h\nu \text{ ---------- } O2 + 2PQH_2$$

where PQ stands for an oxidized form of plastoquinone and PQH_2 for its reduced form, a photon is described by *hv* [Govindjee et al., 2010].
Till now, there were crystal structures of the PSII at atomistic resolution (≤ 3.5 Å) solved only for the cyanobacterium *Thermosynechococcus* sp. [e.g. Guskov et al., 2009; Umena et al., 2011; Ferreira et al., 2004], see figure 3.2. PSII is composed of more than 40 subunits, that are attached to the complex either permanently or temporary (for review see [Shi et al., 2012]), and of many cofactor molecules such as chlorophyll, pheophytin, β-carotene… Structures of some subunits, that are missing in the above mentioned crystals, have been solved independently from various organisms [e.g. Mabbitt et al., 2009; Balsera et al., 2005].
The smallest unit capable of an electron transfer *in vitro* is composed of D1 (PsbA) and D2 (PsbD) core subunits, an inner antenna CP43 (PsbC) and CP47 (PsbB) subunits and interestingly of a low-mass PsbI subunit [Namba and Satoh, 1987; Webber et al., 1989a]. As the subject of my study, the PsbI subunit is described in more detail in the following chapter. PsbA and PsbD form a heterodimer binding cofactors necessary for the electron transfer. PsbA is the subunit most prone to a photodamage. To avoid the necessity of rebuilding the whole PSII after the damage of one subunit, PsbA undergoes a process called D1 turnover, in which this unit is replaced (for review see [Barber and Andersson, 1992]). PsbA and PsbD are surrounded by the inner antenna formed by PsbB and PsbC subunits that bind additional light-conducting pigments, mainly chlorophyll and carotenoid molecules, that are used to conduct photons to the oxygen evolving complex. The other important part of the PSII is the cytochrome *b559* (Cyt *b559*) composed of two subunits, PsbE and

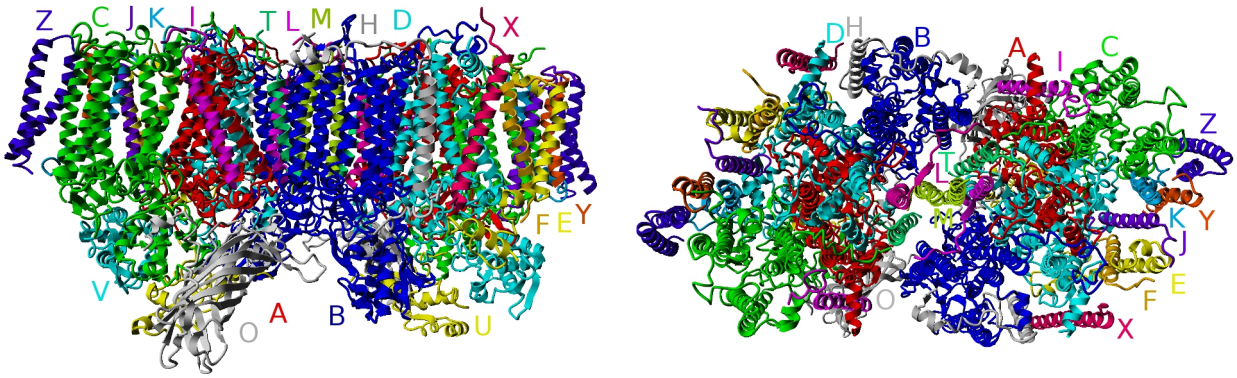


Figure 3.2: *Arrangement of PSII as seen from the edge (left) and from the stromal site of the membrane (right). Some loops of proteins are missing in the crystal structure 3ARC used.*

PsbF. This cytochrome binds a heme molecule and is not involved in primary charge transfer [Burda et al., 2003]. The primary charge separation occurs at the oxygen-evolving complex, that is composed of four manganase ions and a calcium ion linked by μ-oxo bridges [Guskov et al., 2009]. This subsystem undergoes a series of redox reactions resulting in the release of an oxygen molecule and four protons into the lumen and four electrons that travel through the photosynthetic complexes. This five-step process is referred as the Kok cycle [Kok et al., 1970]. An oxygen-evolving complex is located on the lumenal site of the thylakoid membrane, in close vicinity of the D1 and CP47 proteins [e.g. Umena et al., 2011]. From the other site it is capped by soluble extrinsic subunits, that protect and stabilize this cluster (for review see e.g. [Bricker et al., 2012; Roose et al., 2007]). Participation of different subunits in this cap was one of the first differences revealed in the architecture of the PSII among different kingdoms of photosynthetic organisms. It is generally known, that the biggest extrinsic subunit, PsbO, is the only common extrinsic subunit for all organisms. In higher plants and green algae it is joined by PsbP and PsbQ subunits and by the PsbR chain (reviewed by [Bricker et al., 2012; Roose et al., 2007]), that is expected to be partly embedded in the membrane [Webber et al., 1989b]. In cyanobacteria, there are the subchains PsbU and PsbV (cytochrome *c550*) (reviewed by [Bricker et al., 2012; Roose et al., 2007]) and a homologue of PsbP protein [Kashino et al., 2002] from higher plants and green algae designated as CyanoP. Some cyanobacteria (for exceptions see e.g. [Thornton et al., 2004]) have a homologue of PsbQ [Kashino at al., 2002] of higher plants and algae referred to as CyanoQ. The last two subunits are not present in the crystal structures of the PSII available so far [e.g Umena et al., 2011; Loll et al., 2005; Broser et al., 2010]. The protein cap of the oxygen evolving complex of red algae is somewhere inbetween the one of higher plants and green algae and cyanobacteria on the other site. Apart from the omnipresent PsbO, it is composed of the PsbP, PsbU, PsbV subunits and of a homologue of the cyanobacterial CyanoQ, PsbQ' subunit [Roose et al., 2007]. As there is some work done on the PsbP protein presented in this thesis there is another chapter dedicated to this subunit.

More than half of the PSII components are chains with a molecular weight smaller than 15 kDa. Many of them are not crucial for the photosynthesis, although they adjust its function [Shi et al., 2012]. They may participate in the PSII assembly (e.g. PsbK [Iwai et al., 2010], PsbH [Iwai et al., 2006], PsbI [Dobakova et al., 2007]) and act as a pigment binders (such as PsbZ or Psb30 [Barber and Iwata, 2005]). Different subunits may play different roles across kingdoms. Some subunits, mainly those apart form the reaction core may only be attached to the PSII temporary (for a recent review see [Shi et al., 2012, Nixon et al., 2010]).

An other interesting issue connected with the investigation of PSII is its assembly. In the year 2001, it was shown [Zak et al., 2001], that D1, D2, cyt b559 and PsbO, but not the inner antennae CP43 and CP47, are present in the plasma membrane of the cyanobacterium *Synechocystis PCC6803*. Supported by the presence of CtpA endoprotease, that is necessary for the PSII assembly, exclusively in the plasma membrane, Zak et al. [Zak et al., 2001] postulated, that the early steps of PSII assembly occur in the plasma membrane. This theory was adjusted (for review see e.g. [Nickelsen et al., 2011; Komenda et al., 2012]) after the finding that the thylakoid membrane may converge onto the plasma membrane in cyanobacteria [van de Meene et al., 2006]. This joins are considered to be the place of the early steps of PSII assembly.

PSII assembly is studied by construction of various knock-out mutants [e.g. Komenda et al., 2004; Schwenkert et al., 2006] and its general nature is known (for review see [Nixon et al., 2010] on which this paragraph is based).

Cytochrome *b559* (i.e. subunits PsbE and PsbF) is considered to be the nucleation factor of the PSII assembly [Komenda et al., 2004]. After its settlement into the membrane, the D2 subunit is joined [Komenda et al., 2004; Komenda et al., 2008], followed by a dimer formed of D1 and PsbI units [Dobakova et al., 2007]. It is assumed [Dobakova et al., 2007], that PsbI binds D1 unit during the synthesis of the later protein and stabilizes it. During this, a 16 amino acid C-terminal extension is

cleaved by the CtpA endoprotease making from the precursor D1 (pD1) an intermediate D1 (iD1) form still having 8 amino acid precursor on its C-terminus (number of amino for *Synechocystis PCC6803*) [Komenda et al., 2007; Inagaki et al., 2001]. After the next step, when CP47 [Komenda et al., 2004] together with PsbH and some other small subunits [Komenda et al., 2005] are attached, this intermediate extension of the D1 is cleaved as well. Already assembled PsbI subunits are necessary for the CP43/PsbK dimer [Boehm et al., 2011] linkage to the forming PSII [Dobakova, 2007].
As there are different extrinsic subunits present in the different kingdoms of a photosynthetic organisms, assembly pathes differ as well. The common feature is, that extrinsic subunits are not able to assemble to the core of PSII before maturing of the PsbA protein [Diner et al., 1988]. The simplest situation is in green algae, where chains PsbO, PsbQ and PsbP can bind independently *in vivo* [Suzuki et al., 2003]. No information about the assembly or binding mode of green-algal PsbR was found. In cyanobacteria, subunits PsbO and PsbV, that are needed for binding of PsbU [Shen and Inoue, 1993], can bind independently to intrinsic proteins [Burnap and Sherman, 1991; Shen and Inoue, 1993], however presence of PsbO and PsbU stabilizes binding of PsbV [Shen and Inoue, 1993]. As far as I am aware, no study focusing on the assembly of CyanoP and CyanoQ subunits was published. There are two independently binding subunits in red algae PsbO and PsbQ' [Enami et al., 1998]. The other two proteins join then the system paralelly [Enami et al., 1998]. PsbO of higher plants was shown to bind independently to the core of PSII [Miyao and Murata, 1983a]. PsbP, that is necessary for PsbQ binding [Miyao and Murata, 1983b], binds to PSII only in the presence of PsbO and PsbR proteins [Miyao and Murata, 1983b; Suorsa et al., 2006].
PsbJ protein is necessary for attachment of PsbR [Suorsa et al., 2006]. The exact moment of this protein linkage is not known, they can even join the forming PSII simultaneously. There is much less information available about the stages, when the small subunits attach the nascent PSII complex, although it was shown that a presence of ones depends on the others [Nixon et al., 2010]. One of this lines starts with PsbK protein, that is needed for the attachment of PsbZ and after that also Psb30 [Iwai et al., 2010]. PsbM joins PSII via $PsbT_c$ [Iwai et al., 2004]. The final step in PSII assembly is the dimerization of the system. Several subunits join the being assembled PSII only temporary to enable or to restrict binding of other chains [e.g. Roose and Pakrasi, 2004; Yao et al., 2007].
As pigments – mainly β-carotene and (bacterio)chlorophylls – are necessary for the proper work of PSII, their assembly was examined. The stable PSII can only be formed in the presence of a carotenoid with at least one β-ionylidene ring [Bautista et al., 2005]. Presence of chlorophyll *a*, that can, interestingly, be replaced by chlorophyll *b* [Xu et al., 2001] or di-vinyl chlorophyll *a* [Tomo et al., 2009] is necessary for the correct D1 protein translation and via that for PSII core complex formation [He and Vermaas, 1998].
In all [e.g. Guskov et al., 2009; Umena et al., 2011; Ferreira et al., 2004; Loll et al., 2005] but one [Broser et al., 2010] crystal structure available, PSII is shown to be a dimer. Although there is an ongoing debate nowadays, if the dimerization is [Umena et al, 2011] or is not [Takahashi et al., 2009] an artifact of protein crystallization. In some studies (e.g. [Dobakova et al., 2007]) the ratio of monomeric and dimeric PSII complexes is measured and so this issues stays still unsolved.

3.1.2 PsbI protein

The PsbI protein was discovered in 1988 by Ikuechi and Inoue [Ikuechi and Inoue, 1988a] in chloroplasts of spinach (*Spinacia oleracea*) and wheat (*Triticum* spp.) as a 4.8 kDa protein, that is associated with the intrinsic proteins of PSII. Soon after its discovery, PsbI protein from spinach, tobacco (*Nicotiana tabacum*) and liverwort (*Marchantia polymorpha*) was sequened and its tertiary structure was predicted [Ikuechi and Inoue, 1988b]. The prediction is in agreement with the present knowledge about the structure of PsbI, that was determined by X-ray crystallography [e.g. Umena et al, 2011; Loll et al., 2005]. This chloroplast-encoded protein was stated to be conserved among

the higher plants. PsbI was found to be a member of the minimal PSII complex [Webber et al., 1989a], i.e. the minimal unit capable to photosynthetize *in vitro*. Apart from PsbI, other members of this minimal complex are the core proteins PsbA and PsbD and the internal antenna subunits PsbB and PsbC [Nanba and Satoh, 1987; Webber et al., 1989].

In the pre-crystal era of the PSII examination, Tomo et al. [Tomo et al., 1993] reported that PsbI cross-links via its lysine 3 with the PsbD protein (two possible interaction sites were reported) and/or with the PsbE protein (*note: with respect to the up-to-day sequence of spinach, there is no lysine at the position 3 but there is threonine there, whose presence was in the very early sequences of PsbI from spinach unsure. This threonine is surrounded by leucines on both sites, the closes N-terminal lysine is at 5th position. We may speculate, that this is the cross-linking partner.*)

The position of PsbI within PSII was clarified when crystal structures of cyanobacterial PSII were published (the first crystal structure with the complete PsbI protein was released in 2004 by Ferreira et al. [Ferreira et al., 2004]). In all structures, there is PsbI located at the periphery of the complex, in the vicinity of PsbA and close to the PsbC subunits (see figure 3.3). Apart from several lipids (sometimes only fragments of fatty acids) there are chlorophyll and β-carotene molecules found in close proximity of the PsbI subunit. Crystal structures also revealed that PsbI is a single-helical transmembrane protein with a short (14 amino acids out of 38) C-terminal loop on the stromal side of the thylakoid membrane.

The physiological role of PsbI was examined by the construction of knock-out mutants for higher plant tobacco (*Nicotiana tabacum*) [Schwenkert et al., 2006], eukaryotic green alga *Chlamydomonas reinhardtii* [Kunstner et al., 1995] and for cyanobacteria *Synechocystis PCC6803* [Dobakova et al., 2007; Ikuechi et al., 1995] and *Thermosynechococcus elongatus* [Kawakami et al., 2011]. Although all mutants were able to grow photoautotrophically, the role of PsbI was not found to be unique.

In *C. reinhardtii* [Kunstner et al., 1995] removal of PsbI leads to a decrease of oxygen evolution activity and to an increase of photosensitivity. The amount of the D1 subunit present in the mutant was decreased by only 10 – 20 % of the wild type level.

The PsbI subunit in tobacco [Schwenkert et al., 2006] seems to have several functions. Its absence reduces the level of PsbA, PsbB, PsbC subunits and of the extrinsic chain PsbO to about 50% compared to the wild type. More importantly, it increases the level of PSII monomers and trimeric LHCII on expense of dimeric PSII and PSII-LHCII supercomplexes. This is interpreted as an important role in higher level architecture of this complex, although PsbI is not essential for the proper PSII monomeric formation [Schwenkert et al., 2006].

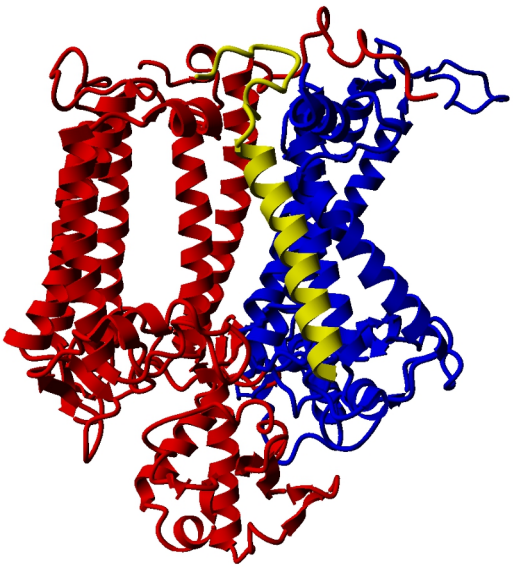


Figure 3.3: *PsbI protein (yellow) and its neighbors in PSII, units PsbA (blue) and PsbC (red). Crystal structure 3ARC is used here.*

An earlier study of PsbI in Synechocystis [Ikuechi et al., 1995] revealed that the absence of this subunit decreases PSII steady-state oxygen evolution activity of about 30% and slightly increases photosensitivity. The same was confirmed by the later study [Dobakova et al., 2007], that is focused mainly on the role of PsbI in PSII assembly. From this point of view, the study showed that PsbI binds to D1, whose turnover it accelerates, and is necessary for binding of the CP43 subunit to PSII. Authors further hypothesize, that PsbI may stabilize a D1 nascent subunit before being incorporated into a *de novo* formed PSII. Unassembled PsbI protein is shown to be removed from the

thylakoid membrane by FtsH protease (coded by the gene *slr0228* in Synechocystis). Absence of the PsbI subunit also alters the Q_A midpoint potential under continuous forward electron flow and reduces phosphorylation of PsbA and PsbD.
Similarly to Synechocystis, the photosynthetic activity of *Thermosynechococcus elongatus* [Kawakami et al., 2011] decreases to 70 – 80% of the wild type level and the amount of PSII dimers is decreased in favor of monomers. Stability of the dimer of *ΔpsbI* PSII was proved by X-ray crystallography with the resolution of 3.45 Å. The resultant crystal structure is unfortunately not available in the ProteinDataBank.

3.1.3 PsbP protein and its homologues

PsbP was first detected in spinach (*Spinacea oleracea*) thylakoid membranes in early 1980ies [Akerlund and Jansson, 1981; Akerlund et al., 1982] and so its behavior and function are well described (for a recent review see e.g. [Roose et al., 2007; Ifuku et al., 2008; Bricker et al., 2012]). Found in higher plants and green algae but not in cyanobacteria and red algae, PsbP protein is known to regulate Ca^{2+} and Cl^- requirements of oxygen evolution of PSII and to protect (with the PsbQ protein) the manganese cluster [Miyao and Murata, 1984; Ghanotakis et al., 1984]. It was also reported to bind manganese ions [Bondarava et al., 2005]. Furthermore, PsbP was shown [Yi et al., 2007] to be necessary for the accumulation of PSII units, mainly its PsbB and PsbD chains. Its (and subsequently also PsbQ protein) assembly to resp. its removal from the PSII is joined with structural changes in the PSII supercomplex, most likely in PsbA and PsbD subunits [Roose et al., 2010; Bricker et al., 2012], resulting in a shift of the light-harvesting antenna CP29 [Boekema, et al., 2000].
Its assembly to PSII was already discussed in chapter 3.1.1. Here I will remind, that in green algae all extrinsic proteins bind to PSII independently of each other [Suzuki et al., 2003], while in higher plants, PsbP, that is necessary for PsbQ binding, binds to PSII in the presence of PsbO and PsbR [Miyao and Murata, 1983a,b; Suorsa et al., 2006]. A large number of unassembled PsbO, PsbP and PsbQ proteins was reported in the thylakoid lumen of higher plants (*Pisum sativum*) and green algae (*Chlamydomonas reinhardtii*) [Hashimoto et al., 1997; de Vitry et al., 1989]. In higher plants was shown [Murakami et al., 2002] that the number of unassembled lumenal PsbP decreases when the amount of free PsbO is lowered. The same happens to PsbQ when the amount of free PsbP is reduced [Ifuku et al., 2005a]. These findings led Ifuku et al. [Ifuku et al., 2008] to predict the presence of some proteases in the lumen, that degradates unassembled oxygen evolving proteins. To preserve this, bound and free proteins are in a rapid binding equilibrium [Hashimoto et al., 1997].

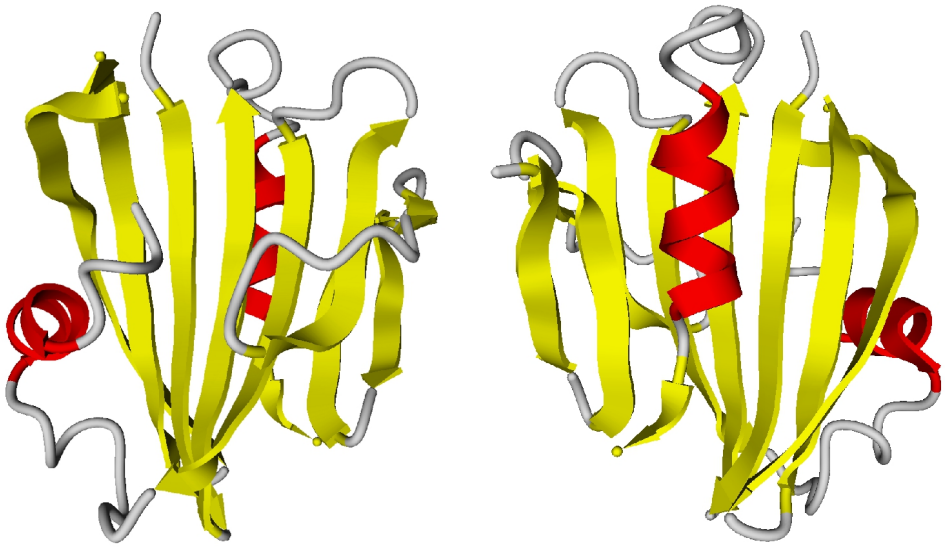


Figure 3.4: *Structure of PsbP protein from tobacco (pdb code 1V2B). N-terminus and two loops are missing in the structure.*

PsbP protein is expected to interact with others proteins of the oxygen evolving complex via electrostatic interactions of its negatively charged amino acids with PsbO and PsbQ proteins [Bricker and Frankel, 2003; Meades et al., 2005].
The 3D structure of a PsbP protein was unknown till 2004, when the crystal structure of PsbP from tobacco (*Nicotiana tabacum*) resolved at 1.6 Å was published, pdb code 1V2B, see figure 3.4 [Ifuku et al., 2004]. The protein is composed mainly of β-strands, forming a 6-strand antiparallel β-sheet, that is its central structural feature. Another anti-parallel β-sheet (2 β-strands) is found at the N-terminus of the structure. Apart from this, there are two α-helices present in the crystal structure, one at its C-terminus and the second one on the opposite site of the central β-sheet. The crystal structure is incomplete with the regions between the amino acids 1 to 15, 89 to 106 and 136 to 140 (incl.) not resolved because of disorder [Ifuku et al., 2004]. Authors have identified two domains in the crystal structure: the first one containing amino acids 1 to 53 and the second one the rest of the protein (the central β-sheet). This is supported by the finding [Kuwabara and Suzuki, 1995] that the bond between amino acids 58 and 59 is susceptible to proteolysis (experiments on spinach). The structure of PsbP protein is similar to the one of a Mog1p protein (pdb code 1EQ6), a nuclear transport Ran GTPase of yeast, as shown by a DALI search performed by Ifuku et al. [Ifuku et al., 2004].
From the functional point of view, there are two region of major importance. The first one is the C-terminal helix (or more exactly the last 10 C-terminal amino acids), that is necessary for effective protein folding [Roffey and Theg, 1996]. The other region of importance is the N-terminus of PsbP protein, as mutants lacking the first 15 resp. 19 N-terminal residues were not able to activate oxygen evolution [Ifuku and Sato, 2001; Ifuku et al., 2005b].
Apart from the above described “real” PsbP protein, its homologues were found in higher plants, green algae and cyanobacteria (for review see e.g. [Ifuku et al., 2008; Bricker et al., 2012]). A genomic analysis of *Arabidopisis thaliana* [de las Rivas et al., 2004; Sato, 2010] revealed genes for two “real” PsbP proteins (one of which is not transcribed) and 8 genes for PsbP homologues. Among the 8 homologues, the two with the highest similarity with PsbP are referred to as the PsbP-like (PPL) proteins, while the others as PsbP domains [Ishihara et al., 2007]. Interestingly, PPL proteins were shown to have a different function in *Arabidopsis thaliana* [Ishihara et al., 2007]. While the first one (designed as PPL1) has a significant sequence similarity with cyanoP and is necessary for effective repair of the photodamaged PSII, the later one (PPL2) is needed for

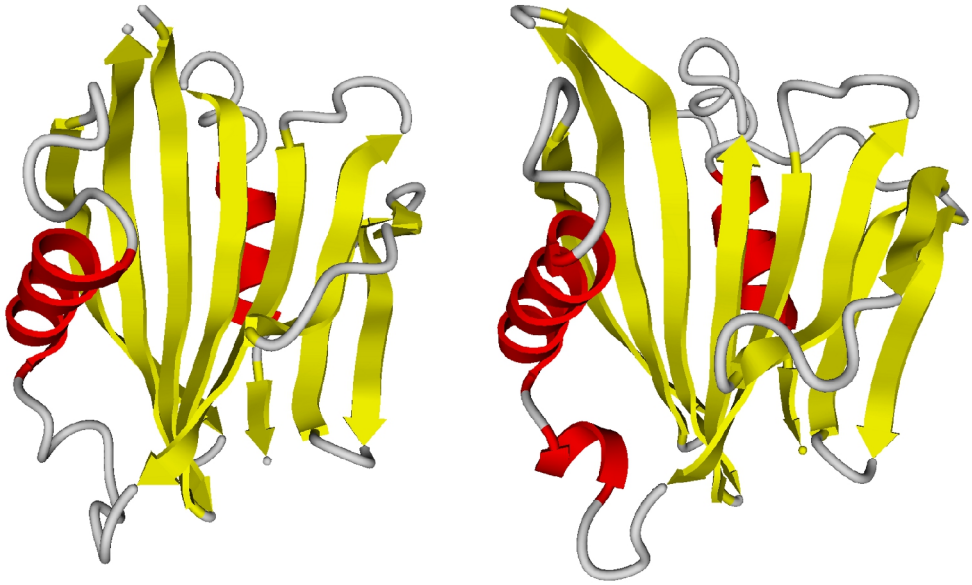


Figure 3.5: *Structures of cyanobacterial homologues of PsbP. Structure from* T. elongatus *(pdb code 2XB3; missing parts) was obtained by crystalography, while the one from* Synechocystis PCC6803 *(pdb code 2LNJ) by NMR. To compare with the PsbP from tobacco see figure 3.4.*

accumulation of the chloroplastic NDH complex, that is involved in a cyclic electron transfer around PSI under stress conditions.
In cyanobacteria, a PsbP homolog, that is referred to as CyanoP, is expected to be a lipoprotein, with a lipidation site on its N-terminus [Thornton et al., 2004]. Although not essential for proper work of the PSII under a wide range of conditions, some conditions can be found, under which cyanoP influences the activity of PSII [Thornton et al., 2004] and may participate in stabilization of charge separation in PSII [Sveshnikov et al., 2007].
As cyanoP is missing in all available crystal structures of PSII, its 3D-position within the system stays unclear. Its 3D structure (pdb code 2XB3) was resolved at 2.8 Å [Michoux et al., 2010] and is almost identical with the one of PsbP of tobacco, see figure 3.5, although these two proteins share a sequence homology of about only 20%. Similarly to the tobacco structure, also the structure from *Thermosynechococcus elongatus* is incomplete with the first 23 amino acids and the residues 133 to 137 missing. To solve this issue, an NMR structure of a cyanoP from Synechocystis was published recently (pdb code 2LNJ) [Jackson et al., 2012], see figure 3.5. It is not surprising that this structure is very similar to the one of *T. elongatus*, with an overall RMSD of the structure with lowest energy to the crystal structure of 1.55 Å. The most important feature of this structure is, that it is complete and so it provides a view onto the N-terminus of the protein, that is a very variable structure. The other differences can be found in the position of the loop, which is not surprising, as loops are generally the most flexible parts of proteins.

3.2 PsbI protein – aims

PsbI is a minor protein of PSII that attaches to the complex in early stages of its assembly and works as a link of PsbA and PsbC core proteins, as described in cyanobacteria. Apart from this, chlorophyll and carotene molecules are found in the vicinity of this protein and so PsbI may participate in their binding.

Research done on PsbI so far is based on experimental work, mainly on effects of various knock-out mutants on PSII complex. There was almost no attention paid to the sole behavior of PsbI not attached to PSII, mainly because the free protein is enzymatically degraded.

My task was to characterize isolated PsbI protein by the means of bioinformatics and computational biology using a model of PsbI from cyanobacterium *Synechocystis PCC6803* with the aim to describe its behavior in the thylakoid membrane and examine possible mutual interactions of PsbI proteins.

3.3 Methods

As there is no crystal structure of the PsbI protein from cyanobacterium Synechocystis available, a homology model was computed. The sequence AAC43720.1 (GenBank code) was downloaded from the PubMed server and used as target sequence. The PsbI chain from the crystal structure 3ARC [Umena et al., 2011] (PSII from *Thermosynechococcus vulcanus*), which is the crystal structure of PSII with the highest resolution (1.9 Å) available, was used as template. With respect to the very high homology of 71% of these two sequences and their identical length, the alignment (figure 3.6) was done manually by aligning the sequences. Ten homology models were calculated by Modeller 9v7 [Sali and Blundell, 1993], the best of which was chosen according to the distribution of amino acids in the Ramachandran plot and the Modeller self-evaluation function. The best model was converted into the CG representation by the *martinize* script, that is available on the official Martini web-page.

Protein sequences for sequence analysis were downloaded in November 2011 form the PubMed internet interface (keyword "PsbI"). If necessary, transit peptides were removed and incomplete sequences were left out. Only unique sequences were used for further work. All sequences were aligned in ClustalX [Larkin et al., 2007] and their sequence logo was created using the Weblogo internet interface [Crooks et al., 2004].

CG systems with one copy of PsbI were set up by the *insane* script, followed by the same minimization protocol as used for membranes without protein (i.e. 1000 steps of the steepest descent energy minimization, and a series of 1000 steps of NVT MD simulation followed by 1000 steps of NpT MD simulations with timesteps of 1ns, 2ns, 5ns and 10ns). Simulations were performed in Gromacs [Hess et al., 2008], with the standard Martini force field for proteins [Monticelli et al., 2008], solution and PG [Marrink et al., 2007] and with the above reported parameters for MGDG, DGDG and SQDG. An elastic network was not necessary to be used for the secondary structure stabilization of the proteins. The simulations ran with a timestep of 20 fs (the maximum possible timestep for proteins) for 1.2 μs. Temperature was kept at 310 K by separately connecting two groups (water + ions and protein + membrane) to the rescale-velocities thermostat once in 0.5 ps. Pressure was coupled semiisotropically by connecting the system once per picosecond to the Berendsen barostat at 1 bar with compressibility of $3e^{-5}$ bar^{-1}. Center of mass motion was removed separately for the upper and lower leaflet of the membrane and for the rest of the system. The standard Martini maintainance (long range interactions cut at 1.2 nm apart from the beads, the shift algorithm used to preserve grouping of the charges at the cut-off distance) of long-range forces (Coulombic and van der Waals) was used. Analysis of the systems were performed using Gromacs tools, if not otherwise stated. Periodic boundary conditions were used.

Mutual interactions of more than one PsbI unit were examined only in the thylakoid membrane of Wada's composition. As can be seen from table 3.1, two different proportions of lipids to one PsbI unit were used. In the first – sparse – case the final frame of the simulation of one PsbI in the membrane was used as input for building bigger systems. The frame was multiplicated and 2 resp. 4 frames were put next to each other resp. onto a grid of 2 × 2 frames. Than the new system underwent the same minimization procedure as if it were built by the *insane* script. Both simulations ran for 5 μs.

To speed up the aggregation process and to enable simulations of more PsbI units, a building block composed of 1 PsbI protein, 29 LPMG, 10 LPDG, 7 DPSQ, 4 lPPG, 938 water beads (i.e.

Figure 3.6: *Alignment used for homology modeling of PsbI. Thermos stands for T. vulcanus, Synecho for Synechocystis PCC6803.*

purpose	protein	MGDG	DGDG	SQDG	PG	NA+	CL-	W	tails
single PsbI	1	174					1	2081	LP
	1		174				1	1784	LP
	1			174		173		2224	LP
	1				174	173		2195	LP
	1	112	32	30	14	43		2667	TM
	1	76	42	59	33	91		2551	TM
big PsbI	2	224	64	60	28	85		5334	TM
	4	448	128	120	56	175		10668	TM
small PsbI	2	68	20	14	8	10		1876	TM
	4	136	40	28	16	20		3752	TM
	9	261	90	63	36	90		8442	TM
	16	464	160	112	64	160		15008	TM
atomistic	1	115	29	32	16	47		7287	TM

Table 3.1: *List of simulations performed with PsbI. TM stands for thylakoid membrane (i.e. LP for MGDG and DGDG, DS for SQDG and lP for PG).*

18.8 water beads per lipid) and 10 sodium beads was created by the *insane* script and minimized by the standard procedure. Then this block was multiplied and systems consisting of 2, 4, 9 and 16 of these blocks were made. Having been built, these systems underwent the “after-insane” minimization again. Systems containing 2 and 4 proteins ran for 2.4 μs, the bigger ones with 9 resp. 16 PsbIs ran for 7.5 μs.

An atomistic model of the Wada's thylakoid membrane after 150ns of relaxation was used as the input for atomistic simulations. The protein was inserted into the membrane by the *g_membed* [Wolf et al., 2010] routine of Gromacs, that removed 5 molecules of LPMG, 3 molecules of LPDG and 10 water molecules. After that the system was minimized by 5000 steps of steepest descent minimization followed by a series of molecular dynamics simulations with the following settings: timestep of 0.5 fs and τ_p = 5ps, for the next two simulations the timestep was risen to 1 fs and system was connected to the Berendsen pressure bath once in 2 ps resp. 1.5 ps. All three simulations ran for 0.5 ns and position restraints (F_r = 1000 N) were applied to the protein in the first two simulations. In the last minimizing simulation running for 0.15 ns, the system was connected to a pressure bath once a picosecond and the timestep of the simulation was set to 1.5 fs. The production simulation ran with a timestep of 2 fs for 250 ns and the pressure of the system was kept at 1 bar semiisotropically by the Berendsen barostat to which the system with a compressibility $4.6e^{-5}$ bar^{-1} was connected once per picosecond. All simulations were performed in Gromacs with the Gromos 45A4 force field for the sugar parts of the lipids and the Gromos 53A6 force field for the rest of the system. The electrostatic interactions were cut off at a distance of 1.4 nm, beyond that value the reaction field approximation with a relative permittivity of 62 bar^{-1} was used. Periodic boundary conditions with a distance of 0.9 nm for neighbor searching were used, the neighbor list was updated every 5 steps. Temperature of the system was kept at 310 K by connecting the membrane grouped with protein and solution separately to a temperature bath (the rescale-velocities thermostat) every 0.5 ps.

3.4 PsbI subunit – results and discussion

3.4.1 Sequence analysis

All sequences used for the sequence analysis and the more comprehensive sequence logo are shown in figure 3.7. The bigger the letter in the sequence logo, the higher probability of an occurence of the given amino acid at the given position.

PsbI protein from various organisms has from 34 to 42 amino acids with clearly defined hydrophobic and hydrophylic/charged regions. It is obvious that the former region ranging from methionine 1 to leucine 24 (both amino acids are conserved among all the sequences; numbers are used in agreement with the alignment) belongs to the transmembrane helix of the protein, while the other ones form the loop. It is also evident, that the sequence involved in the helix is more conserved than the one forming the loop. Furthermore the helix is in all organisms composed of 24 amino acids, which means that this length is of some meaning (its length is probably the same as the thickness of the membrane).

Looking closer at the helix region, we realized that the (almost) conserved amino acids are in intervals of about 4 residues, and thus are above each other in the helix in 3-D space. Crystal structures reveal that conserved amino acids are on the side of the helix by which PsbI faces the first N-terminal transmembrane helix of the PsbA subunit and the chlorophyll molecule that is sandwiched between PsbA and PsbI in the given region, see figure 3.8. The ring of a β-carotene molecule is found close to the C-terminus of the PsbI helix in proximity of a region of 4 conserved amino acids (phenylalanine 21 and 23, glycine 22 and leucine 24). We may hypothesize, that at least some of these residues are involved in binding or right positioning of these pigments into PSII. Other conserved residues of this cluster together with other conserved amino acids of the helix may play role in correctly orienting PsbI when assembled with PsbA. Residues considered to be conserved are methionine 1, leucine 4 and lysine 5, valine 8, valine 12 (that is in 3 cases replaced by isoleucine, an other highly hydrophobic amino acid [Kyte and Doolittle, 1982]), phenylalanine 14 (5× replaced by leucine, once by serine) and phenylalanine 15 (once mutated to leucine), leucine 18 (6× replaced by isoleucine and 3× by valine, both, similar to leucine, very hydrophobic residues) and phenylalanine 19 (once replaced by phenylalanine resp. by a proline) and the conserved C-terminal region of this helix, where phenylalanine 21, glycine 22, phenylalanine 23 (2× replaced by leucine) and leucine 24 are found. I will not discuss the conserved residues one after another, but just mention, that the presence of charged residues close to the end of the transmembrane helices is usual and it is involved in stabilizing the protein in the membrane [Killian et al., 1996]. The loop region, that in different organisms has from 10 to 18 amino acids, is characteristic by a high number of charged residues ranging from 2 to 7, none of which is histidine (positively charged). Furthermore, aspartic acid 27 and arginine 30 are conserved within all sequences and there is a strong tendency to have two positively charged residues close to the position 35, one of which is arginine and the other one lysine (in some groups they are at positions 34 and 35, while for others they are aligned at positions 35 and 36). The loop usually ends with one or two negatively charged residues in about half of all cases (51 out of 91 sequences) separated by an apolar amino acid. Two prolines at positions 28 and 32 are almost conserved within all sequences (the first one is twice replaced by alanine and once by threonine, while the second one is once mutated into arginine). Proline

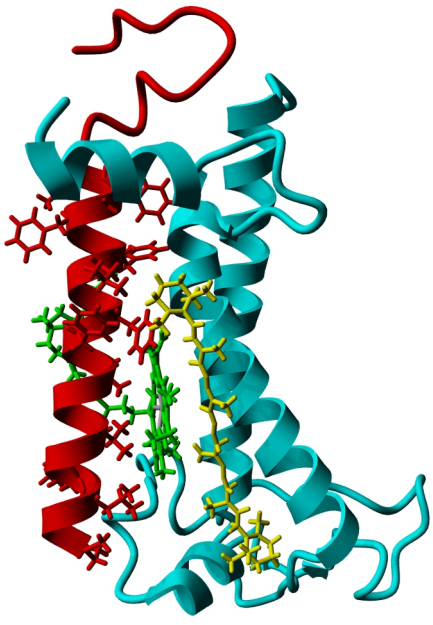


Figure 3.7: *Conserved amino acids in PsbI protein (red) and their position with respect to PsbA protein (cyan, cut after residue 140) and neighboring molecule of chlorophyll (green) and β-carotene (yellow).*

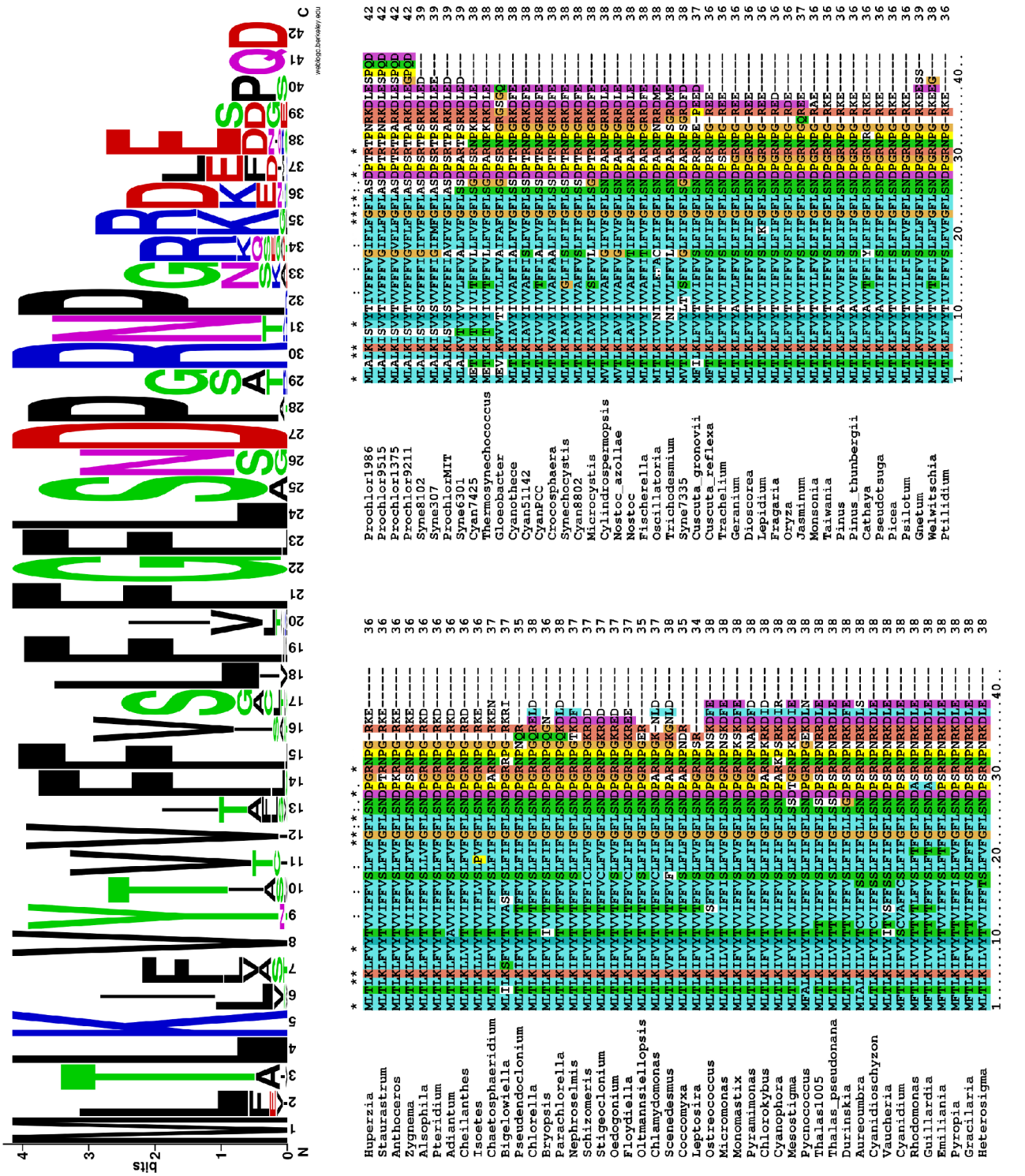


Figure 3.8: *Sequences of PsbI protein and sequence logo (made by weblogo server [Crooks et al., 2004]) used for sequence analysis. The bigger the character in the sequence logo the more frequently this amino acid occurs in the alignment at a given position.*

is interesting as its amine nitrogen is part of the ring and so it introduces bends into the backbone of the protein. The species with a long loop region of PsbI (< 40 amino acids; some prochlorococci) have another proline at the position 40. Apart from these the other amino acids making the loop are (apart from glycine) of polar nature.

```
                ****** ** ** :*:*********.** *******:*
PsbI_avarage    MLTLKIFVYTVVIFFVSLFIFGFLSNDPGRNPGRKDLE
Synechocystis   MLTLKIAVYIVVGLFISLFIFGFLSSDPTRNPGRKDFE
                1.......10........20........30........
```

Figure 3.9: *Comparison of the "average" sequence of PsbI protein with the one from Synechocystis, that is dealt with in the thesis.*

A comparison of PsbI from Synechocystis and the "average" PsbI protein made by the most abundant amino acids at the given position is shown in figure 3.9. Interestingly, the "average" PsbI is just a hypothetical sequences, that is not present in any organism. It is closest to e.g. bracken (*Pteridium aquilinum*), Taiwan Plum Yew (*Cephalotaxus wilsoniana*), or Bladder Campion (*Silene vulgaris*), with those it has a homology of 94%.
The "average" PsbI sequence has a homology of 79% (30 identical amino acids out of 38) with PsbI from Synechocystis that is the subject of my study. Five of the differing amino acids are in the helix region (position 7, 10, 13, 14 and 16) and 3 are in the loop area (positions 26, 29 and 37). All changes are at the positions where we observe a larger variety in amino acids, with the exception of leucine 14. Phenylalanine is usually conserved at this position, with the exception of 5 species having leucine here and one with serine. Synechocystis is also one of the organisms that have hydrophobic amino acid between the two negatively charged residues at the C-terminus.

3.4.2 Homology modeling of PsbI

With respect to the high homology (71%, i.e. 27 out of 38 amino acids identical) of PsbI sequences form Synechocystis and *T. vulcanus*, that was used as the template structure for the homology modeling, it is not surprising, that all the 10 models are almost identical as can be seen from their fit onto the template structure in figure 3.10A and 3.10B. The differences between them are mainly in the side-chains orientation. Model 3, that is in the detail shown in figure 3.10C, 3.10D and 3.10E, was chosen as the best model according to a visual comparison with the template structure, the distribution of amino acids in the Ramachandran plot and the self-evaluation Modeller objective function.

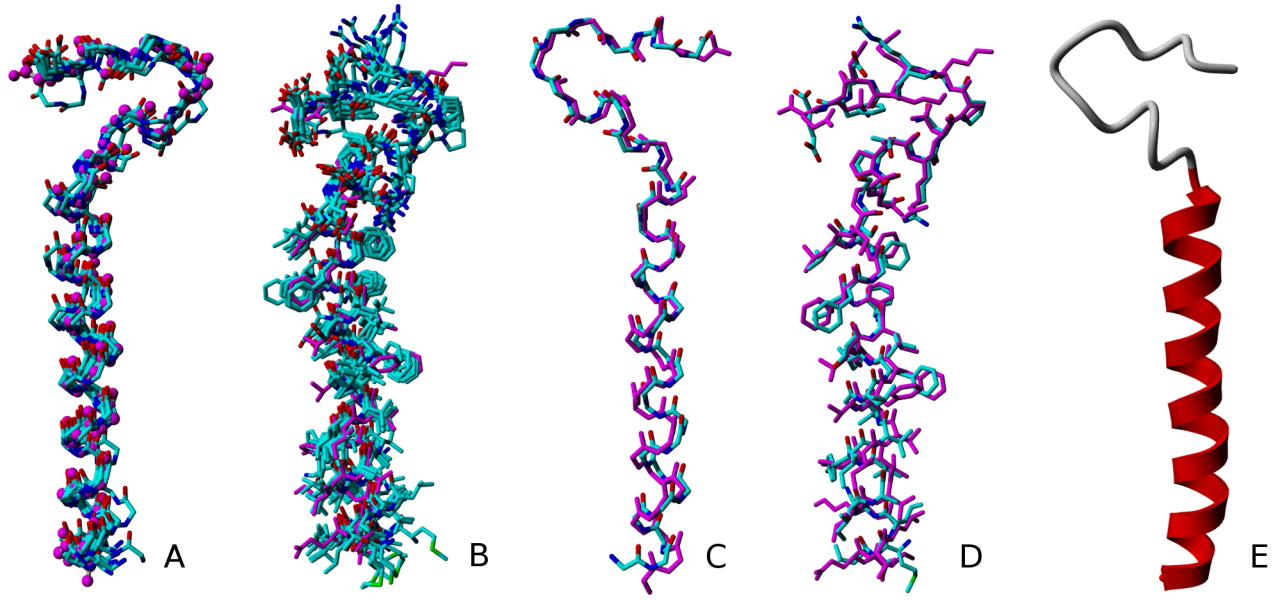


Figure 3.10: *Comparison of all models obtained by homology modeling with the template structure (3ARC_I) shown in magenta. Not only backbones (A) but also side-chains (B) show very good agreement among each other. The best model 3 is compared with the template structure for backbone (C) and side-chains (D). The final frame (E) shows the resultant model in ribbon representation.*

3.4.3 Coarse-grained simulations of one PsbI protein in membranes

All simulations show a lipid bilayer in LQ phase with solution entering into the headgroup region of the membrane. The loop of the protein lays down onto the membrane rapidly after the start of the simulation (within the first 30 ns), what indicates, that the erected loop in the crystal structure is a result of the protein's interactions with other subunits of PSII, most likely with its closest neighbor, the PsbA subunit. A visual inspection of the simulations does not show any ion bound to PsbI, neither do lipid headgroups interact with the protein as shown by a listing of the lipids (defined by the last headgroup bead) in the distance up to 1 nm from the last center of mass of the helix. Although some of the lipids approach the protein, they will after some while go away from its proximity as the result of diffusion (data not shown). As the loop of the protein is very flexible (see later) it is not expected that it could bind any lipid.

Estimations of the APL and membrane thickness are almost the same as for systems without PsbI (see table 3.2), which indicates that insertion of the protein does not change membrane properties, with the exception of LPSQ, where protein insertion causes a phase transition from gel to LQ phase. To characterize the protein behavior, root mean square deviation (RMSD) was calculated for the backbone beads of the whole protein and separately for the backbone beads of the helix (amino acids 1 to 24) and of the loop (residues 25 to 38) over the course of the whole simulations. As the tilt of the loop occurs within the first 30 ns and the length of the whole simulation is 1.2 μs, I do not think, that it could influence the average RMSD values dramatically. RMSD curves in figure 3.11 and averaged values of RMSD in table 3.3 show that in all cases the helix does not fluctuate from the initial structure as much as the loop. RMSDs of the loop and also of the whole protein oscillate between two values which may indicate two configurations. As can be seen from the graphs, the protein embedded in different membranes tends to stay in the conformation with lower energy over a different time.

	LPMG		**LPDG**		**LPSQ**		**LPPG**		**Wada**		**Sakurai**	
	protein	*membr.*	*protein*	*membr.*	*protein*	*membr.*	*protein*	*membr.*	*protein*	*membr.*	*protein*	*membr.*
APL [nm^2]	0.683	0.684	0.696	0.719	0.673	0.572	0.715	0.701	0.652	0.643	0.630	0.618
thick.[nm]	3.5	3.5	3.7	3.7	3.7	4.4	3.9	4.0	3.8	3.7	3.9	3.9

Table 3.2: *Comparison of APL and membrane thickness for systems with and without PsbI estimated for the small membrane patches.*

Averages of the RMSD of the protein inserted into the membrane of a different composition differ up to 40% of the values of the lowest one with respect to the highest one and there is no general rule for the characteristics that make the system to have the bigger resp. smaller RMSD. The average RMSD and RMSD of the loop is significantly higher for the LPMG membrane than for other systems. The high abundance of MGDG in the thylakoid membrane of the Wada's composition may be the reason, why this membrane has higher values of RMSD for the system and the loop than the thylakoid membrane of Sakurai's composition. RMSD of the protein in the LPDG membrane is interesting by its lowest absolute values but highest percentage of standard deviation with respect to averaged values. A similar situation can be seen for the thylakoid membrane of Wada's composition. PsbI in the LPDG membrane also shows the significantly highest RMSD for the helix (with the smallest percentage of standard deviation). Here, the biggest size of lipid headgroups and the effort

	protein			**helix**			**loop**		
	average	*std.dev*	*%*	*average*	*std.dev*	*%*	*average*	*std.dev*	*%*
LPMG	0.602	0.086	14.29	0.125	0.024	19.20	0.516	0.097	18.80
LPDG	0.383	0.125	32.64	0.182	0.030	16.48	0.361	0.071	19.67
LPSQ	0.482	0.083	17.22	0.140	0.029	20.71	0.470	0.073	15.53
LPPG	0.524	0.115	21.95	0.116	0.026	22.41	0.393	0.095	24.17
thylakoid_W	0.489	0.139	28.43	0.123	0.029	23.58	0.356	0.139	39.04
thylakoid_S	0.411	0.090	21.90	0.129	0.033	25.58	0.329	0.073	22.19

Table 3.3: *Averages of RMSD (in nm) of PsbI protein embedded in different membranes.*

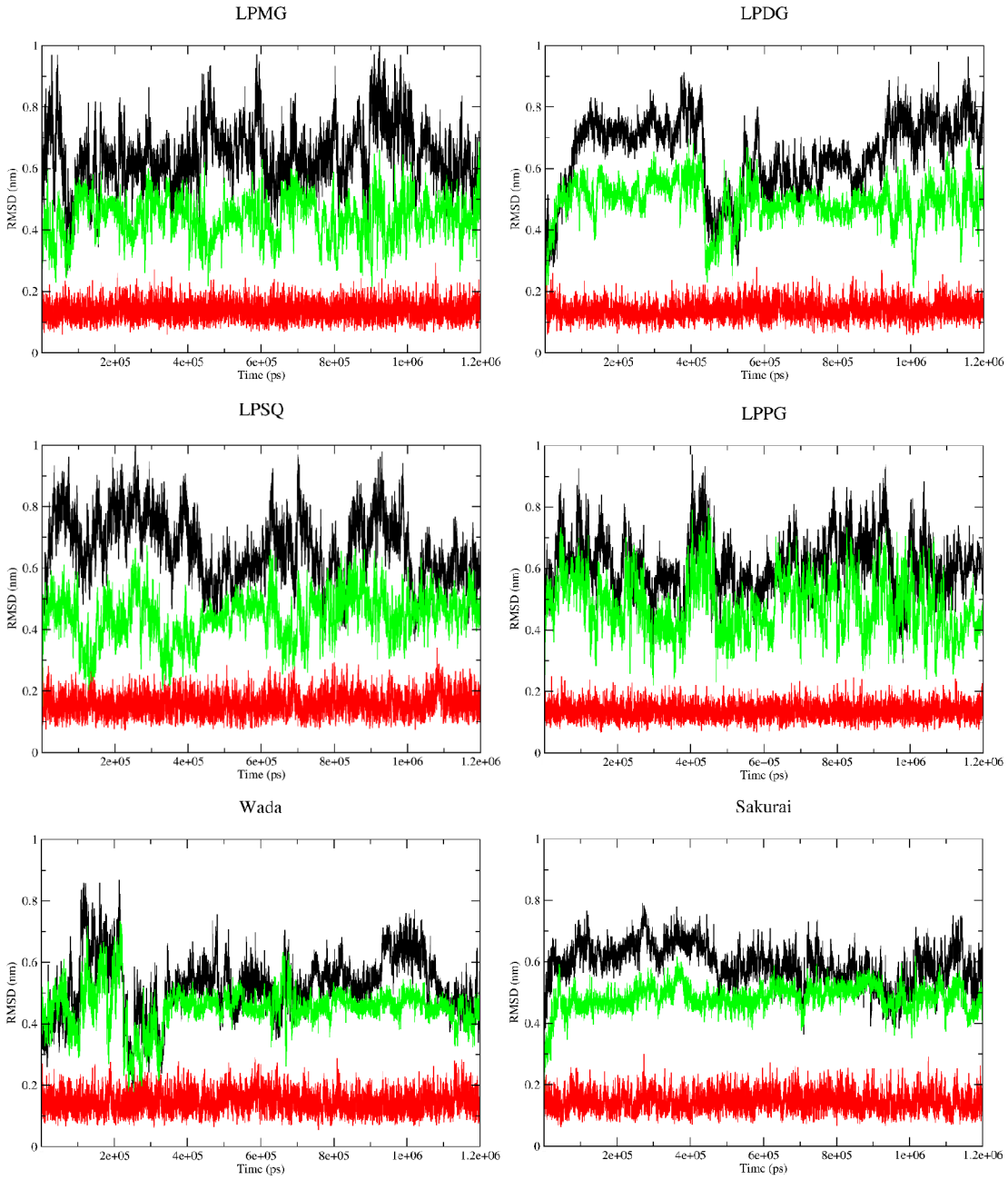


Figure 3.11: *RMSD for backbone beads for the protein (black) and separately for the helix (red) and the loop (green) region.*

of the transmembrane part of the protein to go around it might be an explanation for this. The loop has the highest values of RMSD for LPMG and SQDG membranes and interestingly the lowest one for both thylakoid membranes (although the differences here are not very high).
With respect to the shape of the RMSD curves as well as the fraction of the standard deviation to the absolute value of the given quantity, we may conclude, that the protein switches between (at least) two equilibrium positions, but it is not stable in any of them for a long time. With respect to the length of the simulation, running the simulation even longer would not bring any considerable improvements. This behavior does not depend of the membrane composition (for the examined membranes).
The higher flexibility of the loop in comparison to the helix is confirmed by the root mean square fluctuation (RMSF) curves shown in figure 3.12. Starting with residue 25 (the first one of the loop),

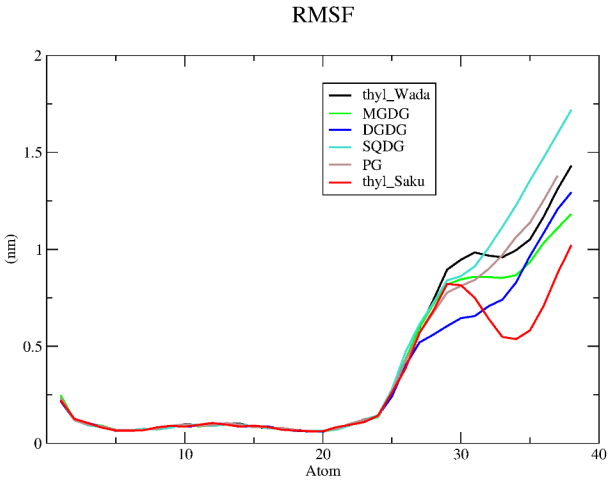
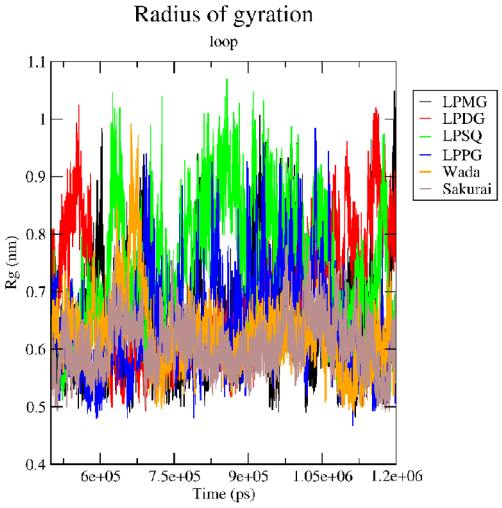


Figure 3.12: *RMSF for backbone beads of simulations with 1 PsbI protein.*

the RMSF starts growing rapidly. For the proteins embedded in the charged membranes (SQDG and PG) this growth is almost linear with a small inflection at the position of 29th amino acid. RMSF of the loop in the DGDG membrane behaves similarly the previous ones, although it has two inflections at the position of the amino acids 27 and 30. The others loops – concerning their RMSF – are similar to those already reported till the amino acid 29 resp. 31, where they grow rapidly. In the later case, there is an inflection point at amino acid 29 for PsbI in the membrane of Wada's composition. The region of the common behavior of all loops is followed by amino acids having a constant RMSF and from residue 34 it increases for PsbI in the MGDG membrane. PsbI in the Wada's membrane shows a slight decrease in RMSF between residues 31 and 33, and from residue 34 it increases again. In case of the loop of PsbI in Sakurai's membrane, this decrease is much more obvious and occurs for residues 29 and 34. RMSFs in the helix region is higher at its ends than in the central part.

Radius of gyration was used to characterize the loop behavior in even more detail, see table 3.4 and figure 3.13. Like RMSD of the loop-only region, the radius of gyration is smallest for both thylakoid mixtures, especially the Sakurai's one. The highest value of the radius of gyration is obtained from the simulation of PsbI in SQDG membrane, the second highest average is in the DGDG membrane. Curves of the radius of gyration show that the loop oscillates between two conformations, one with a higher radius of gyration the other one with smaller one. Similar behavior was seen in the RMSD curves for the loop. The averaged values of the radius of gyration can be interpreted as the time the loop spends in the conformation with the lower radius of gyration as all the curves have the lower (and also the higher) value of a similar level. This can be interpreted that the thylakoid mixture stabilizes the loop region.



	radius of gyration		
	average	*std.dev*	*%*
LPMG	0.651	0.094	14.4
LPDG	0.693	0.113	16.3
LPSQ	0.751	0.106	14.1
LPPG	0.641	0.083	12.9
thylakoid_W	0.622	0.062	10.0
thylakoid_S	0.605	0.045	7.4

Figure 3.13 and table 3.4: *Radius of gyration for backbone beads of the loop of PsbI after 500 ns of simulation.*

	tilt		
	average	*std.dev*	*%*
LPMG	154.8	8.9	5.7
LPDG	154.2	8.3	5.4
LPSQ	156.0	7.8	5.0
LPPG	155.8	8.7	5.6
thylakoid_W	157.9	8.3	5.3
thylakoid_S	162.0	7.0	4.3

Table 3.5: *Tilt of the helix of PsbI.*

A tilt of the helix (defined by the backbone beads of amino acids 2 to 5, resp. 21 to 24) to the membrane normal shows similar values for all systems simulated, see table 3.5 (graphs not shown). It may be just an coincidence, that the average value of the tilt is smallest for the thylakoid membranes and biggest for the neutral lipids.
The RDF was used to see, if there are any preferences of certain lipid headgroups around the center of mass of backbone beads of the protein, see figure 3.14. The primary maxima of the unmixed membranes reaches higher values for charged lipids, LPSQ and LPPG, than for neutral LPMG and DGDG. All curves but the one for LPPG have some tiny maxima at about 0.8 nm from the COM of the helix. LPPG has a wide primary minimum there. The primary minima are deeper for uncharged lipids LPMG and LPDG, while the shallowest one is for LPSQ. The secondary maximum occurs for all lipids in the same distance from the helix and is the highest for LPSQ and lowest for LPPG. A secondary minimum and tertiary maximum are present in the curves, too.
Considering mixed membranes, we must be aware that the membrane of Wada's composition is smaller than Sakurai's one (188 vs. 210 lipids). RDF curves for all lipids around the helix show similar shapes for all except SQDG and also similar heights for the given lipid in both membranes. RDF for SQDG in Wada's membrane is approximately twice as high as in Sakurai's membrane till the secondary maximum. Then the difference between them decreases getting to a similar level at about 3 nm, i.e. in the region where arrangement of lipids around the helix is random. RDF for SQDG is typical by the major reduction of its secondary maximum, with an obvious tertiary maximum that is half the size of the primary maximum for Wada's membrane, resp. at a level of its 60 % for Sakurai's membrane. For the later membrane the RDF increases over its further course, while for the former one it stays almost constant after its tertiary maximum. The distribution of the PG headgroups around the COM of the helix is similar to the one of SQDG of Sakurai's membrane. The secondary maximum is turned into a wide primary minimum. The absence of the secondary maximum, resp. its major reduction in the case of SQDG can be explained (similarly as for the membranes without helix) by the mutual repulsion of the charged lipids. Here we may hypothize about an effect of charged amino acids of the loop region of PsbI onto the behavior of lipids, but as the protein from Synechocystis has the same number (three) of positively and negatively charged amino acids, we can assume that their overall effect will averaged out. The secondary maximum in the RDF curve of DGDG is smaller than the tertiary one, or in other words, the tertiary maximum is adjusted as it is of the same height as the primary one, while the secondary maximum is about 75%

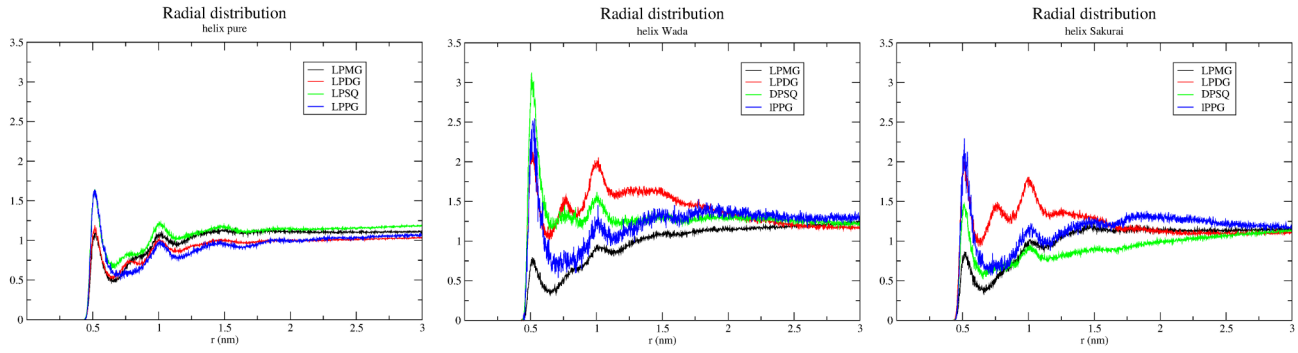


Figure 3.14: *RDFs of headgroups of lipid around backbone beads of the helix of PsbI.*

of the height of the primary maximum. MGDG then shows a curve with obvious but very small (<1) primary maximum, a negligible secondary maximum that is in the position where the curve grows to the tertiary maximum, that is concerning its height at the level of the primary maximum and after that it grows even further.
Generally spoken, although there are some regions with prefered lipid appearance, the absolute values of the RDF are relatively low. As mentioned earlier, there are no lipids associated with PsbI over a significant time, so we must bear in mind, that the RDF curves show approaching of lipids along PsbI. As the RDFs for the two membranes with different compositions show similarity of the results, they can be considered as representative ones. It can be concluded from the graphs in figure 3.14 that the helix of the protein repells MGDG in the membranes of both compositions and that in the less charged membrane of Wada's composition, it attracts SQDG.
Mutual distributions of lipid headgroups with respect to each other were also examined (data not shown). In the membrane of Wada's composition, preferences are identical as those of the membrane without PsbI, while in the membrane of Sakurai's composition, there are some changes. The distribution of lipids around MGDG is identical in both systems with and without protein. The most abundant lipid around DGDG is PG and not SQDG (both the lipids are charged) like in the membrane without protein. The preference of PG and SQDG is also switched in the vicinity of PG, but for this lipid both charged lipids are less favorable ones. While in the membrane without PsbI DGDG and SQDG were of the same abundance, in the membrane with protein DGDG is more preferable in the vicinity of SQDG than SQDG. Before making any definite conclusion about this issue, we must realize that the simulation of the pure membrane ran almost 10times longer than the one with protein (as the purpose of the later simulations was to describe the dynamics of the protein, the simulations are long enough for this purpose). As the membrane of Wada's composition has faster diffusion than Sakurai's one, and the mutual distribution of the lipids is the same no matter the presence of the protein, the most likely explanation for the observed differentiation of the lipids in the membrane of Sakurai's composition with the protein embedded is that it did not reached an equilibrium yet.
The lateral diffusion coefficient is another property connected with the differentiation of the membrane. Estimations of the lateral diffusion coefficients are shown in table 3.6, MSD curves based on these estimations are shown in figure 3.15. When comparing presented MSD curves with those obtained from simulations without protein, we must be aware that the former were calculated in a shorter time (0.7 μs) than the later ones (1.1 μs) and that the total length of the simulations with protein is shorter (1.2 μs) than those without the protein (10.0 μs). An attempt to determine the lateral diffusion coefficient for the backbone beads of the helix of the protein failed as the helix does not move more than its diameter (data not shown).
The overall lateral diffusion of lipids in membranes with PsbI inserted is in agreement with data obtained for simulations without protein. This means, that in unmixed membranes LPMG moves fastest, followed by LPPG with a bit slower lateral diffusion. Lateral diffusion coefficients are slower for LPSQ and mainly for LPDG. The thylakoid membrane of Sakurai's composition exhibit similar arrangement of the lateral diffusion, i.e. MGDG and PG moves faster than DGDG and SQDG. Contrary to that, in the membrane of Wada's composition MGDG moves faster than the other three lipids, as observed for simulations without protein. Differentiation of MSDs in mixed membranes presented above is not that obvious as in thylakoid membranes without the protein, what may be caused by the different time over which the simulations were performed and that in the simulations with the protein the equilibration of the lateral diffusion of lipids is not fully established yet.

	diffusion coefficient [e-05 cm/s]			
	MGDG	*DGDG*	*SQDG*	*PG*
pure	6.1	2.8	3.8	5.6
thylakoid_W	4.3	3.7	4.1	3.5
thylakoid_S	3.9	3.1	3.4	3.8

Table 3.6: *Estimates of lateral diffusion coefficients of lipids in membranes with PsbI.*

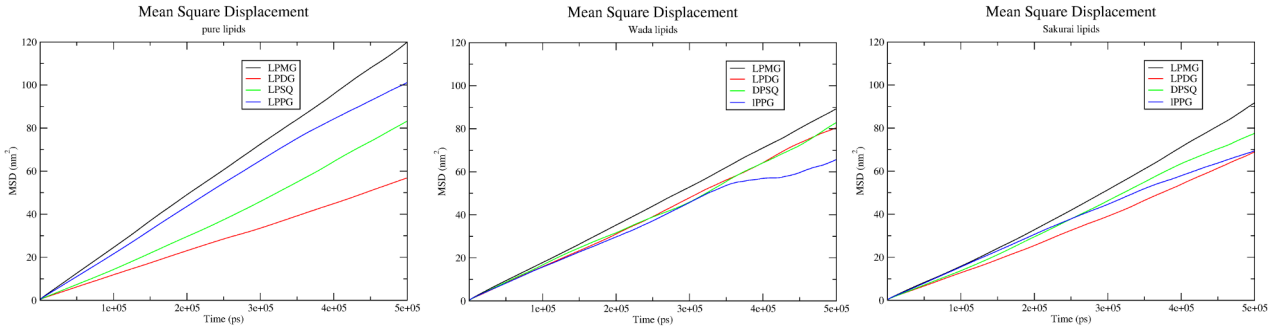


Figure 3.15: *MSDs for lipids in membranes with PsbI.*

There was an effort to characterize the movement of the loops by describing the mutual stereochemical properties of the backbone beads of amino acids 24, 29, 33 and 38. These were chosen because of their position in the initial model of PsbI. Their averaged values incl. standard deviations are shown in tables 3.7, graphs are not shown. All values are interesting by their relatively high fluctuation mainly for interactions gathering more than two beads. It is impossible to find any mutual dependence of the given bonds, angles and improper dihedral angle with respect to each other. For bonds, we can say, that the closer the two beads are, the more conserved the given bond distance among simulations of protein in various membranes is and for more distant beads the average distance has the biggest split of the values. None such rule can be found for angles and improper dihedrals.

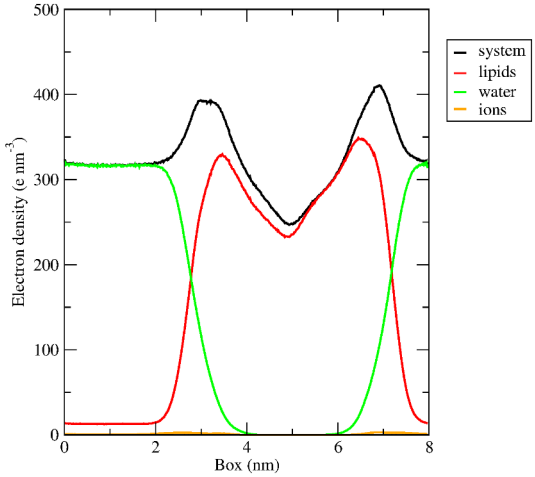
	MGDG			**DGDG**			**SQDG**		
	average	*std.dev*	*%*	*average*	*std.dev*	*%*	*average*	*std.dev*	*%*
24|29	0.904	0.231	25.6	0.724	0.169	23.3	0.972	0.216	22.2
24|33	1.082	0.363	33.5	1.161	0.354	30.5	1.236	0.517	41.8
24|38	1.404	0.562	40.0	1.627	0.571	35.1	1.962	0.650	33.1
29|33	0.843	0.154	18.3	0.796	0.167	21.0	0.894	0.128	14.3
29|38	1.511	0.357	23.6	1.552	0.257	16.6	1.756	0.367	20.9
33|38	1.006	0.217	21.6	1.044	0.149	14.3	1.129	0.209	18.5
24|29|33	79.800	33.800	42.4	100.500	27.100	27.0	89.400	44.800	50.1
24|29|38	87.100	36.700	42.1	83.400	42.000	50.4	93.100	47.000	50.5
24|33|38	114.900	33.500	29.2	95.500	29.800	31.2	112.800	35.800	31.7
29|33|38	68.900	41.600	60.4	117.000	21.500	18.4	124.300	29.000	23.3
improper	-7.900	86.100	-1089.9	37.700	68.500	181.7	-15.300	113.800	-743.8

	PG			**thylakoid_W**			**thylakoid_S**		
	average	*std.dev*	*%*	*average*	*std.dev*	*%*	*average*	*std.dev*	*%*
24|29	0.877	0.214	24.4	0.854	0.196	23.0	0.867	0.201	23.2
24|33	1.037	0.360	34.7	0.774	0.271	35.0	0.866	0.171	19.7
24|38	1.504	0.519	34.5	1.311	0.346	26.4	1.230	0.275	22.4
29|33	0.795	0.158	19.9	0.847	0.131	15.5	0.829	0.163	19.7
29|38	1.434	0.392	27.3	1.669	0.263	15.8	1.575	0.293	18.6
33|38	1.061	0.201	18.9	1.123	0.153	13.6	1.112	0.144	12.9
24|29|33	82.100	36.900	44.9	55.300	22.900	41.4	59.900	16.300	27.2
24|29|38	82.500	44.300	53.7	49.900	21.600	43.3	48.900	20.500	41.9
24|33|38	94.700	37.700	39.8	86.500	26.700	30.9	76.700	20.700	27.0
29|33|38	104.500	34.600	33.1	119.200	25.800	21.6	110.800	24.700	22.3
improper	1.500	106.300	7086.7	53.900	52.200	96.8	23.800	57.100	239.9

Tables 3.7: *Mutual stereochemical properties of backbone beads of amino acids 24, 29, 33 and 38 were used to characterize the loop movement.*

3.4.4 Atomistic simulation of one PsbI protein in the thylakoid membrane

Insertion of PsbI does not influence the APL and membrane thickness of the thylakoid membrane simulated at atomistic resolution, as can be seen in table 3.8. The electron density profile shown in figure 3.16, is very similar to the one of the membrane without the protein. This means that water and lipid molecules are separated and ions prefer to group right above the membrane. Like in the CG representation, the atomistic system shows the membrane in LQ phase. The protein starts laying down onto the membrane already during the relaxation runs and the loop does not return to its initial position over the whole simulation. Visual inspection further shows, that the bend on the loop region, that is present in the crystal structure and so in the initial homology model, is not persistant in the loop. Contrary to the CG simulations, the first (N-terminal) turn of the helix unfolds in the atomistic simulation.



	APL [nm^2]	thickness [nm]
with protein	0.639	3.8
membrane	0.635	3.7

Table 3.8: *Insertion of PsbI protein into Wada's membrane does not considerably change its APL and thickness. Atomistic analysis after 100 ns from the start of the simulation.*

Figure 3.16: *Electron density profile for the atomistic simulation of the PsbI protein in the thylakoid membrane. Protein is excluded from the electron density calculations.*

The same characteristics as in the CG simulations were used to characterize the behavior of PsbI. RMSD of C_α atoms as a function of time, figure 3.17 left, shows that the protein does not reach equilibrium in 250 ns, what is a highly unusual behavior for isolated protein of the given size (38 amino acids). Similarly to CG simulations RMSD of the helix is smaller than those of the protein and the loop that are almost identical. It can be concluded, that the protein does not adopt any stable conformation. A similar trend can be seen in the radius of gyration of C_α atoms of the loop as a function of time, figure 3.17 right, as here the system does not settle in any stable conformation,too.

RMSF for C_α atoms was calculated after 100 ns of the simulation to quantify the contribution of the individual C_α atoms into the overall motion of the protein, figure 3.18. Similarly to the CG

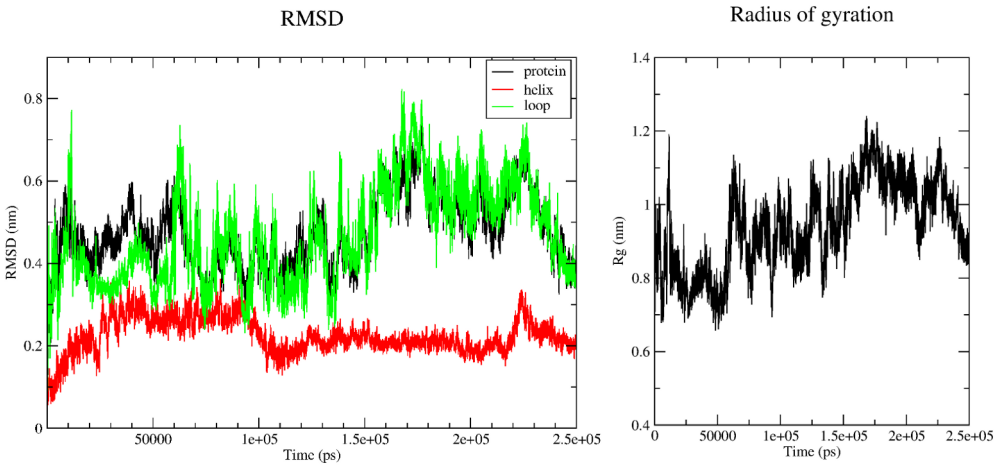


Figure 3.17: *RMSD for C_α atoms of the whole protein and separately for its loop and helix (left) and radius of gyration for the C_α atoms of the loop (right).*

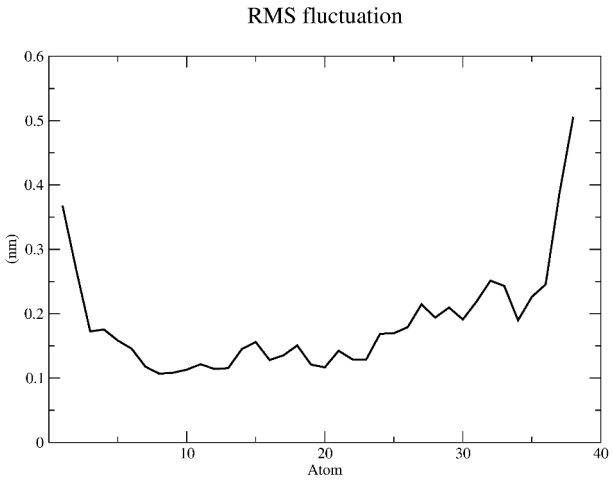


Figure 3.18: *RMSF of the C_α atoms of the atomistic simulation after 100 ns of the simulation.*

simulation, the loop region has a higher RMSF than the helix, although the growth of the curve starting at residue 23 (in contrast to the amino acid 24 in the CG simulation) is not that steep. The much higher RMSF at the N-terminus of the protein, that cannot be seen in the graphs for the CG simulations (figure 3.12) confirms, that the first turn of the helix is unfolded. The other similarity with the CG simulations is the small hump in the loop region with a minimum at the position of residue 34. The same position of the minimum was observed in the CG system of the protein inserted into the thylakoid membrane of Sakurai's composition, while PsbI in Wada's thylakoid membrane and in pure LPMG has this minimum at position 33. The C-terminus of the loop fluctuates more than the rest of the system. The range of the fluctuation of helix is of similar values for the atomistic and CG simulations, while the loop fluctuates much less in the atomistic than in the CG representation.

To describe the behavior of the loop in more detail, hydrogen bonds between its atoms were analyzed. Figure 3.19 shows that the number of hydrogen bonds varies with time between none to nine. The smaller the number of hydrogen bonds, the higher probability it occurs. The hydrogen bond existence map (see figure 3.20) shows, there is no hydrogen bond stable over the whole course of the simulation (and thus persitant), although there are some present over a considerably long time or appearing repeatedly. This supports the conclusion, that there is no stable structure in the loop of PsbI. Further, visual inspection did not reveal any ion interacting with the protein and so this issue was not addressed anymore.

With respect to the relatively short time of the simulation, there was only little attention payed to search for potential hydrogen bonds between the protein and lipids. Obviously, analysis (see figure 3.19) reveals their existence, what is not surprising as saccharides, prone to form hydrogen

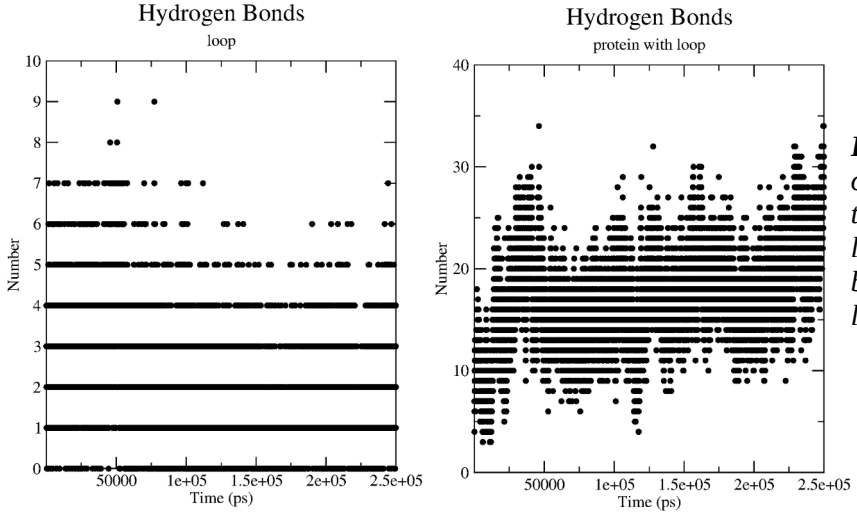


Figure 3.19: *Graphs of the number of hydrogen bonds as a function of time between the residues in the loop of the PsbI protein (left) and between protein and membrane lipids (right).*

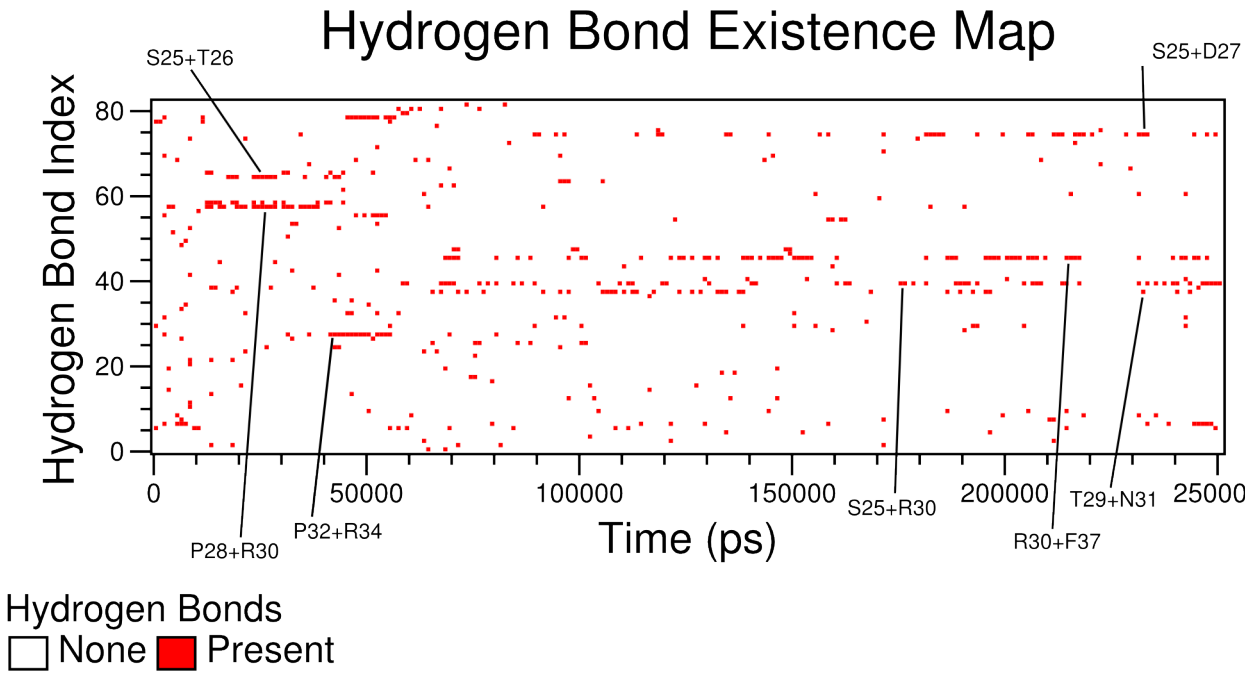


Figure 3.20: *Existence map for hydrogen bonds between amino acids of the loop. Hydrogen bonds appearing for a longer time are identified.*

bonds, make the majority of the lipids of the thylakoid membrane. However, the simulation is not long enough to let the lateral diffusion of the lipids prove and so we cannot judge if the existing hydrogen bonds between the protein and lipids are a coincidence of their mutual proximity or are present temporary. In order to address this issue, several frames of the simulation were used to hydrogen bond analysis in Yasara, that uses a more strict criterion for a hydrogen bond search than Gromacs. Results shown in table 3.9 do not suggest longer existence of the hydrogen bonds, although the conclusion is based on the insignificantly small number of six frames.
Like in the CG simulations, distances, angles and improper dihedrals between C_α atoms of amino acids 24, 29, 33 and 38 were determined as a function of time, see figure 3.21, Their average values are shown in tables 3.10. Before comparing the behavior of the atomistic and the CG simulations one must be aware that the former simulation was performed for a much shorter time and so the

frame [ns]	hydrogen bonds						
100	M1+M174	K5+M167	K5+M168	L24+M62	D25+M81	L2+S219	T3+S219
	Y9+S211	K35+S126	D36+S126				
125	M1+M179	K5+M168	L24+M62	S25+M85	S26+M85	D27+M85	R34+M96
	D27+D105	R34+D110	K5+S211	Y9+S211	K35+S126		
150	S25+M85	S26+M85	D27+M85	N31+M86	R34+M95	K35+M86	D27+D105
	L2+S220	Y9+S211	R34+S118	K35+S126			
175	M1+M179	L2+M169	T3+M179	K5+M168	D27+M85	R30+M63	K35+M86
	E38+M96	Y9+S211	K35+S126				
200	M1+M169	T3+M179	L24+M64	S26+M64	S26+M85	D27+M83	D27+M85
	T29+M83	R30+M63	K35+M86	Y9+P223			
225	K5+M160	K5+M168	S25+M85	D27+M85	R34+M59	T29+D105	L2+S220
	R34+S118	K35+S126	K5+P233	Y9+P233	E38+P134		
250	M1+M179	L2+M169	T3+M175	T3+M179	K5+M168	I6+M168	S25+M85
	D27+M83	T28+M86	R30+M63	K35+M86	K5+D192	D27+D109	T29+D109
	K35+D109	E38+D109	E38+D110	K5+S211	K35+S126	E38+S126	

Table 3.9: *Hydrogen bonds between the protein and membrane lipids at the given time frames. The first binding partner is an amino acid, the second one is a lipid. MGDG is shortened as M and its bonds are shown in blue, DGDG (D) is shown in red, SQDG (S) in brown and PG (P) in green.*

system did not have any opportunity to map the conformational space as well as the later systems. Knowing this, it is not surprising that standard deviations of average values for atomistic simulation are smaller than the ones for CG simulations. The other general property of average values obtained from the atomistic simulation is that all the values are higher than those obtained by the CG simulation (but the average for angle 29, 33, 38 where the average of the CG simulation of the PsbI in MGDG membrane is smaller). Interesting, the order of the bond length of atomistic simulation is identical with the CG simulation of the PsbI in the DGDG membrane, order of angle sizes differs from all CG simulations. It is not surprising that closest residues have a smaller mutual distance than the more further. The same can be concluded for angles. The improper dihedral angle fluctuate randomly over the whole course of the atomistic simulation, what is the same behavior as was observed for all CG simulations.

The behavior of the helix was further characterized by its tilt with respect to the z-axis of the system (time dependance of the tilt of the helix on time not shown). Similarly to CG simulations the helix is tilted of a bit more than 20°, but to the other site of the axis (22° instead of 158° on average). The atomistic curve differs from the CG one, that it is more structured and not of the same value over the time as in CG simulations.

To conclude and compare the behavior of the protein at the atomistic and CG level, we may state that the CG simulations reflect the overall behavior of PsbI in good agreement with the atomistic simulations, mainly with respect that the protein does not disturb the behavior of the membrane and in its inability to adopt a stable conformation. On the other hand, the RMSF of the loop in the atomistic simulation is relatively smaller (in comparison with the helix) than in the CG simulations. Interestingly, the hump present in the RMSF of the loop is present in both types of simulations of PsbI in the thylakoid membrane. The first N-terminal bend of the helix is not fixed in the atomistic simulation, what is the matter of an extend in which the helix is defined in the CG simulation. In none of the simulations ions and lipids associate with the protein.

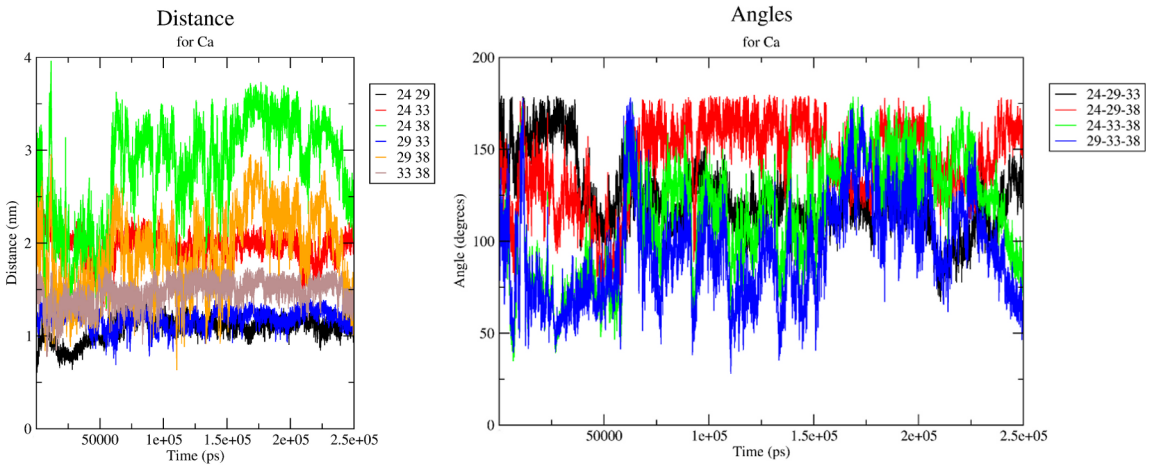


Figure 3.21: *Time dependance of mutual distances and sizes of angles of C_α atoms of amino acids 24, 29, 33 and 38.*

atoms	24|29	24|33	24|38	29|33	29|38	33|38
dist [nm]	1.14±0.074	1.96±0.122	3.04±0.347	1.19±0.093	2.03±0.397	1.50±0.114

angle	24|29|33	24|29|38	24|33|38	29|33|38	improper
value [°]	116.7±13.1	151.1±15.1	126.3±22.7	99.9±25.8	-65.3±147.5

Tables 3.10: *Averages of distances, angles and improper dihedral angle between C_α atoms of residues 24, 29, 33 and 38.*

3.4.5 Large thylakoid membranes with 2 and 4 PsbI proteins

Systems of 2 resp. 4 PsbI units embedded into the thylakoid membrane of Wada's composition were simulated to describe the mutual interactions of more PsbI proteins. Both system contained 188 lipids per protein and are hydrated by 14.2 water beads per lipid (the exact composition of these systems can be found in table 3.1).

No major attention was payed to the properties of the membrane and to the protein-membrane interactions as in the previous chapter previous I already described that the protein does not interact neither with the membrane nor changes its properties. A visual inspection of both systems shows lipids forming a bilayer in LQ phase.

The system with the two PsbIs is the only one of all systems with more PsbI proteins examined, where the proteins do not aggregate. Although the proteins get close to each other for several times, they never manage to form a stable dimer and depart after a short time. As can be seen from figure 3.22, the COM of the backbone beads of the helices (amino acids 1 – 24) never get closer than 2.8 nm and only twice – at approx. 3.0 and 3.35 μs – closer than 3 nm. At about 0.6 and 1.05

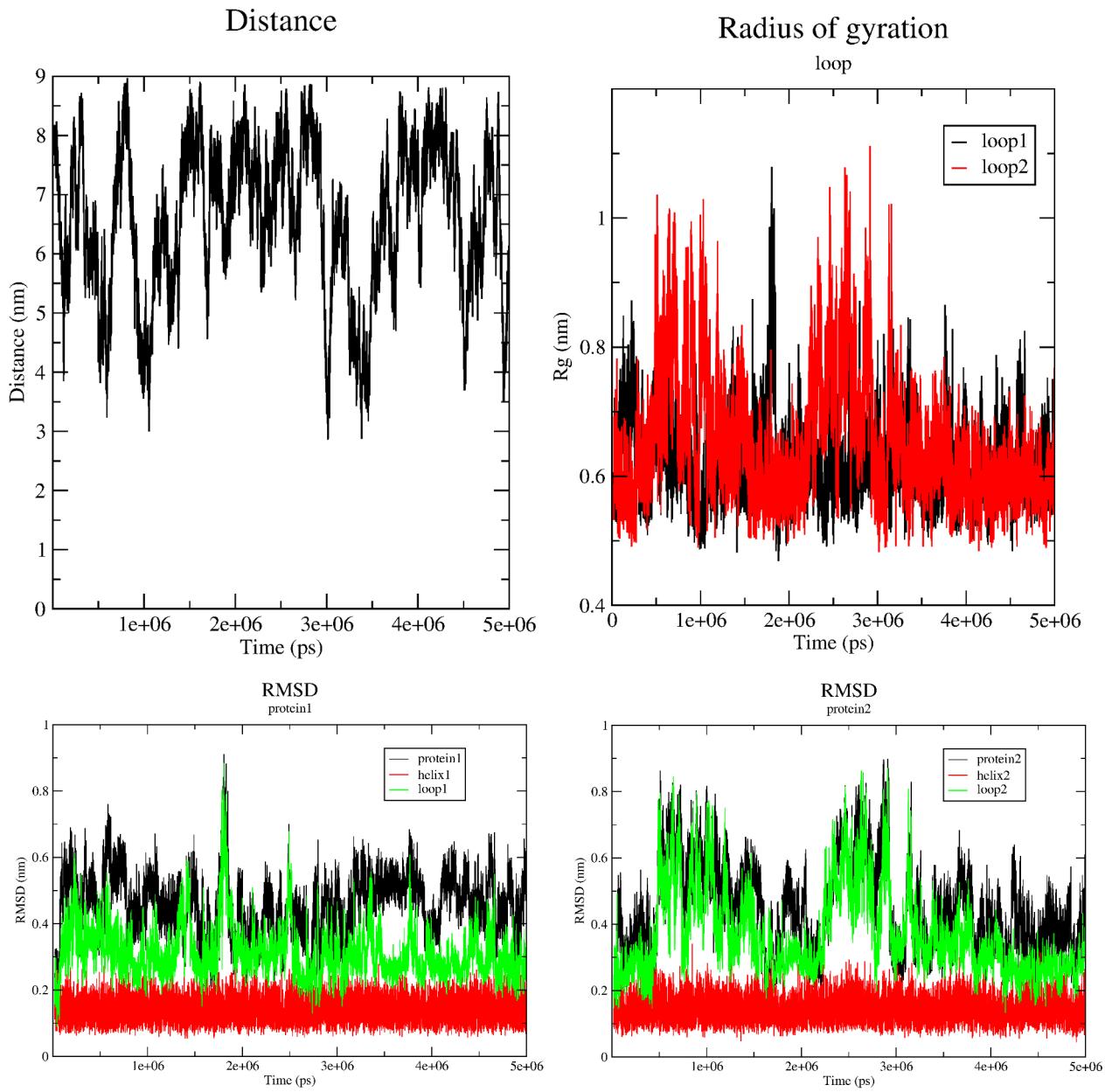
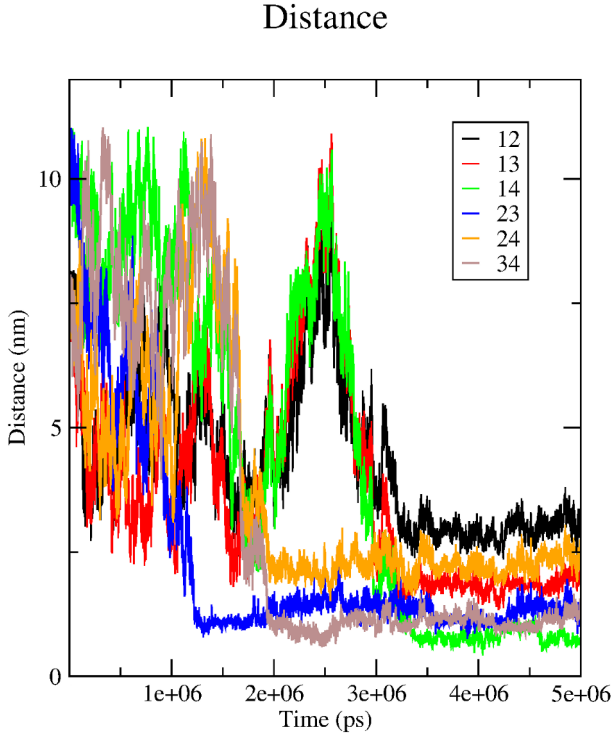


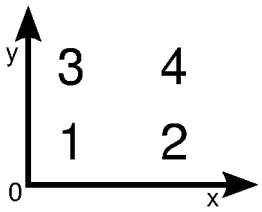
Figure 3.22: *Properties of a large simulation with 2 PsbI proteins.*

	RMSD protein [nm]		RMSD helix [nm]		RMSD loop [nm]		gyration loop [nm]		tilt helix [°]	
	average	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*
protein1	0.456	0.090	0.138	0.033	0.313	0.075	0.620	0.061	158.4	9.4
protein2	0.452	0.135	0.140	0.034	0.369	0.126	0.645	0.093	157.0	9.4

Table 3.11: *Averages and their standard deviations (over the whole simulation) of protein properties determined for a system with 2 PsbI units.*

μs of the simulation, COMs of the helices are closer than 3.5 nm to each other. This confirms that the two PsbI units do not dimerize and behave more or less like free proteins.
This is supported by the analysis of the properties of the proteins. Like for single protein in the different membranes, RMSD of the backbone beads of the helix is smaller than the ones of the loop and the whole protein. For both proteins, RMSD of their loops fluctuate between two different values that may represent different conformations. While the protein1 is almost exclusively found in the conformation with lower RMSD the second one spends a considerable time in the conformation with higher RMSD, see figure 3.22. A similar behavior is seen for the radius of gyration (calculated for the backbone beads of the loop only), figure 3.22. The helices were further described by their tilt with respect to the z-axis (data not shown), which is also in an agreement with the behavior of the helices in the systems with one protein. For the averaged values and their standard deviations of all the properties determined, see table 3.11.
Contrary to the behavior of the system composed of 2 PsbI proteins, in the system made of 4 PsbI units aggregation occurs as can be seen from the mutual distances of the COM of the backbone beads of their helices in figure 3.23. Interestingly, the resultant aggregate is of a rod-like shape and not of a circular shape, which indicates, that there are regions that are more susceptible to mutual interactions. The aggregation of the whole system starts by dimerization of proteins 2 and 3 (for the initial setting of the proteins see figure 3.23) at the timepoint 1.2 μs. Protein4 starts approaching the existing dimer at 1.7 μs and after proper orientating it joins the dimer at 1.9 μs by binding to protein3. Monomer1 approaches the trimer at about 3.0 μs and after orienting it binds to the system at 3.2 μs. The behavior of protein1 is interesting as it had approached the dimer together with unit4 at 1.6 μs (i.e. a bit before unit 4) but it did not manage to bind to the system.
The behavior of the proteins was further characterized by computing RMSD for the backbone beads of the whole protein, its helix and loop, see figure 3.24. As expected from the behavior of the





Figures 3.23: *Mutual distances in dependence of time and initial position of proteins in a large simulation of 4 PsbI units.*

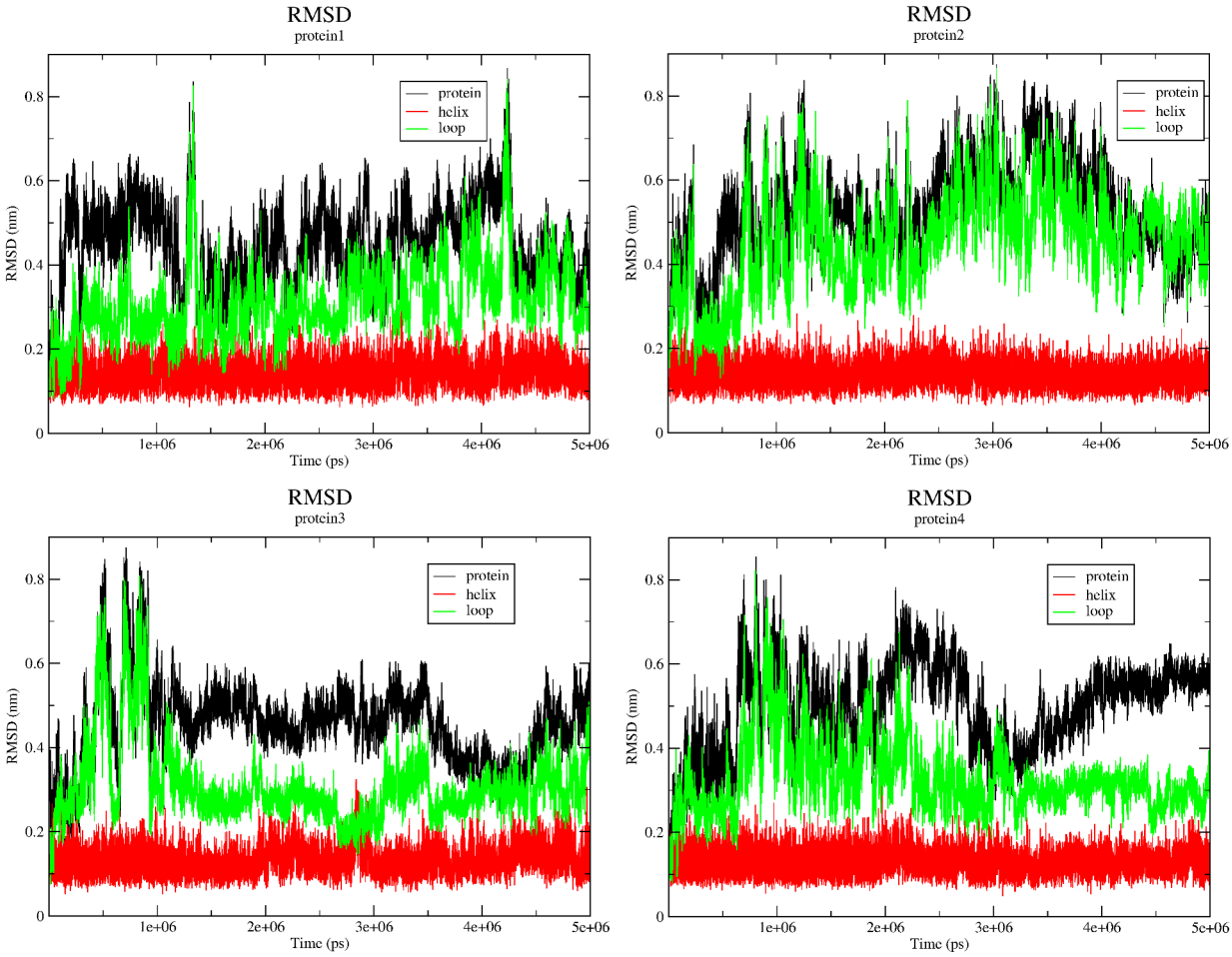


Figure 3.24: *RMSD of backbone beads of protein and separately for the loop and helix.*

previous systems, the RMSD of the helix is much smaller than the others one and is not influenced by the mutual approach of the proteins when multimerizing. The only exception is helix3, that between 2.8 and 2.95 μs increases its RMSD. At this time protein1 approaches the already existing trimer. Around this time (a bit broader on both sides) the RMSD of loop3 is at the level of the RMSD of the helix and occasionally it is even lower. The RMSD of loop3 is characterized by being closer to the RMSD of the helix than to the whole system in the period after binding protein2 and before protein1 associates with the system, that leads to the increase of the RMSD of loop3. Similarly the RMSD of whole protein3 decreases between 3.5 and 4.5 μs. Both RMSDs of loop3 and protein3 exhibit significant peaks before the protein's interaction with protein2. Similarly, the RMSD of loop4 decreases after binding protein4 to the already existing dimer 2 + 3, but binding of protein1 increases it and decreases the RMSD of the whole protein4. This can be interpreted as an interaction between proteins 3 and 4 that stabilizes their loops. Protein2 shows the most significant difference between the height of the helix and loop with protein. Till about 0.7 μs there is a very little difference between the three curves, but afterwards the curves of loop and protein, that are very similar, increase. No interactions with other protein influences these curves. This means that

	RMSD protein [nm]		**RMSD helix [nm]**		**RMSD loop [nm]**		**gyration loop [nm]**		**tilt helix [°]**	
	average	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*	*average*	*std.dev.*
protein1	0.449	0.099	0.139	0.030	0.311	0.087	0.619	0.060	156.1	9.6
protein2	0.525	0.111	0.137	0.027	0.451	0.111	0.679	0.091	155.8	10.3
protein3	0.453	0.094	0.131	0.032	0.308	0.092	0.631	0.069	157.5	9.5
protein4	0.495	0.105	0.129	0.027	0.325	0.081	0.608	0.067	156.3	9.1

Table 3.12: *Averages and their standard deviations (over the whole simulation) of protein properties determined for the system with 4 PsbI units.*

the loop freely fluctuates within the space what is a situation similar to the loops of unbounded proteins. The curves for protein1 and loop1 are remarkable by the sharp increases at 1.3 and 4.2 μs, that can be assigned to the approach of this unit to the dimer 2+3 resp. to some rearrangement within the tetramer with respect to the unit 3 whose behavior of the RMSD alters at a similar time as well.
Like in the other simulations, the loops were described by their radius of gyration, showing the same behavior as the RMSD of the loops (data not shown).
Additionally to the RMSD, the helices behavior was described by their tilt to the z-axis (data not shown). As expected from the analyze of the RMSD, neither here are any significant differences from the helix behavior in the systems with one protein. The average values of all characteristics are shown in table 3.12.

3.4.6 Small thylakoid membranes with several PsbI proteins

These simulations were performed to see if an increasing number of PsbI units speeds up the aggregation of the proteins as it may seem from the simulations of PsbI proteins embedded in the larger thylakoid membranes (see previous chapter). To speed up the aggregation, the proportion lipid-to-protein was decreased in comparison with the previous large simulations. For comparison, simulations with 2 and 4 protein units were set up in this lipid-to-protein ratio.

Visual inspection of all systems reveals, that membranes are bilayers in LQ phase, with water and ion beads entering into the headgroup region of the membranes. As already shown, the presence of PsbI in the membrane does not change its APL and membrane thickness (and visual inspection did not show any disturbances of membranes) and thus these properties of membranes were not studied in detail.

The other common feature of all systems containing 2, 4, 9 or 16 PsbI units is that the proteins aggregate. This process is initiated by a dimerization of two neighboring units in the time order of 100 ns, and is usually finished within 600 ns. In case of the larger systems made by 9 resp. 16 proteins some of the units stay unpaired. The dimers are stabilized by interactions of the helices and are stable till the end of the simulation. Dimers and unpaired proteins may approach each other by a random motion and then in dependence of their mutual orientation multimerize into clusters of higher order. This multimerization occurs in two ways: either the proteins interact by their helices resulting in rod-like structures or they interact by their loops, which provides more rectangular multimers. The later structures are less stable than the former ones and they may split or reorganize after the approach of some other proteins or multimers. Not all approaches of proteins or their complexes lead to a multimerization or to a split of already existing complexes.

When describing these systems we must be aware of the use of periodic boundary conditions, that may have an effect onto the systems when the given protein (or a cluster of several proteins) is close to the edge of the simulation box. If there are three proteins (or their complexes) at the same position with a similar x- resp. y-coordinates, the proteins near the edges of the simulation box interact with the periodical image of each other. It is obvious that the middle protein interact with both its neighbors. This make some ”ring” interacting of these three proteins in which they are trapped for some time. Only some imbalance in this system, caused either by some random processes within the “ring-forming units” or by approaching another protein, may brake this “ring” (for the better explanation see figure 3.25). An interaction of one protein with the periodical image of the other one is even seen in the system with the 2 PsbI units.

The approach of another protein may, but does not have to, influence the stability and the structure of the given PsbI unit (graphs of RMSD, radius of gyration and tilt of the helix are similar to those of the big systems and are not shown). In some cases, multimerization may lead to stabilization of the radius of gyration resp. the RMSD of the loop of the protein. Also major changes in these quantities are connected with the mutual approach of the proteins. It is not surprising that the helix

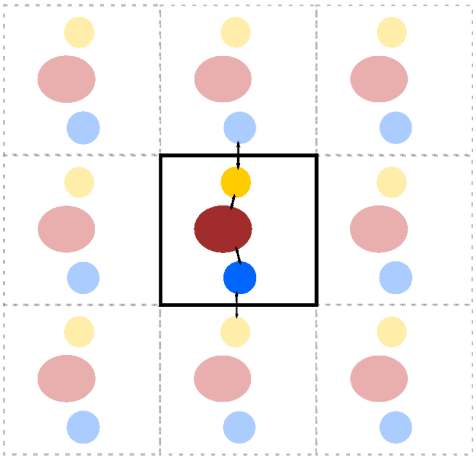


Figure 3.25: *Periodic boundary conditions used in simulations enables mutual interaction (shown by arrows) of yellow and blue object.*

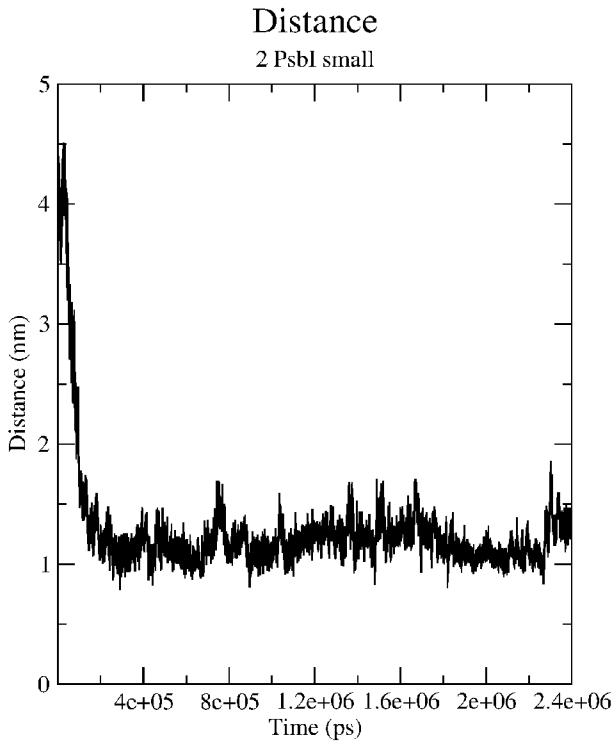
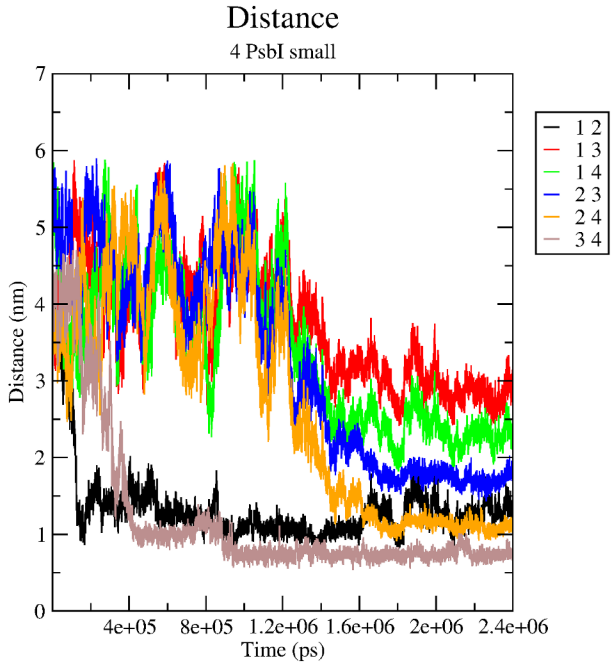


Figure 3.26: *Mutual distance of COM of helices in a small simulation of 2 PsbI units. With respect to the minimum image convention, it cannot be distinguished when protein interact with its real binding partner and when it bind periodical image of the other protein.*

is the less influenced part of the protein and even its tilt with respect to the z-axis is not changed dramatically. All these properties and the behavior is similar to the large system of 4 PsbI units and thus are not discussed further in the detail.

Before describing individual systems, it can be summarized that the multimerization may, but does not have to, lead to stabilization of the loop region of PsbI and that it can be disturbed by the approach of another protein. Equilibration of the loop does not occur in all proteins in the multimer. The proteins tend to form dimers very quickly in which they stay over the rest of the simulation and they can aggregate into a polymers of higher order. Multimerization via helices is much more stable than the one by loops.



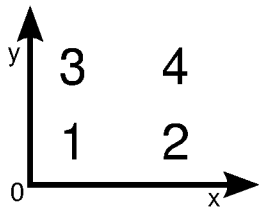


Figure 3.27: *Mutual distance of COM of helices in a small simulation of 4 PsbI units. Their initial arrangement the same as for the large simulation with 4 units. Initial setting of the proteins is shown in the scheme above.*

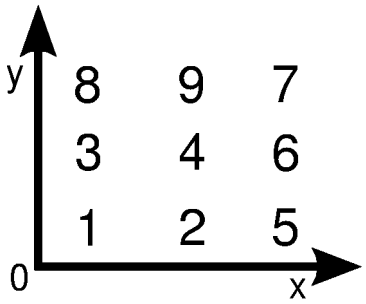


Figure 3.28: *Initial setting of proteins in system containing 9 PsbI units.*

In the system made of two proteins, PsbI dimerizes within the first 150 ns of the simulation (figure 3.26) and stays together during the rest of the simulation. The proteins interact by their helices on the side of the conserved residues and this interaction is preserved over the time. With respect to fluctuations of the tilt of the helix the mutual distance of amino acids being involved in the dimerization changes. As the longer axis of the dimer is (almost) identical to the shorter side of the simulation box, the proteins do not only interact with their partner but may also via the periodic boundary conditions interact with the periodic image of the other protein.

In the system composed of 4 PsbI units, two dimers are formed at approx. 130 and 420 ns. The binding mode of the units in the

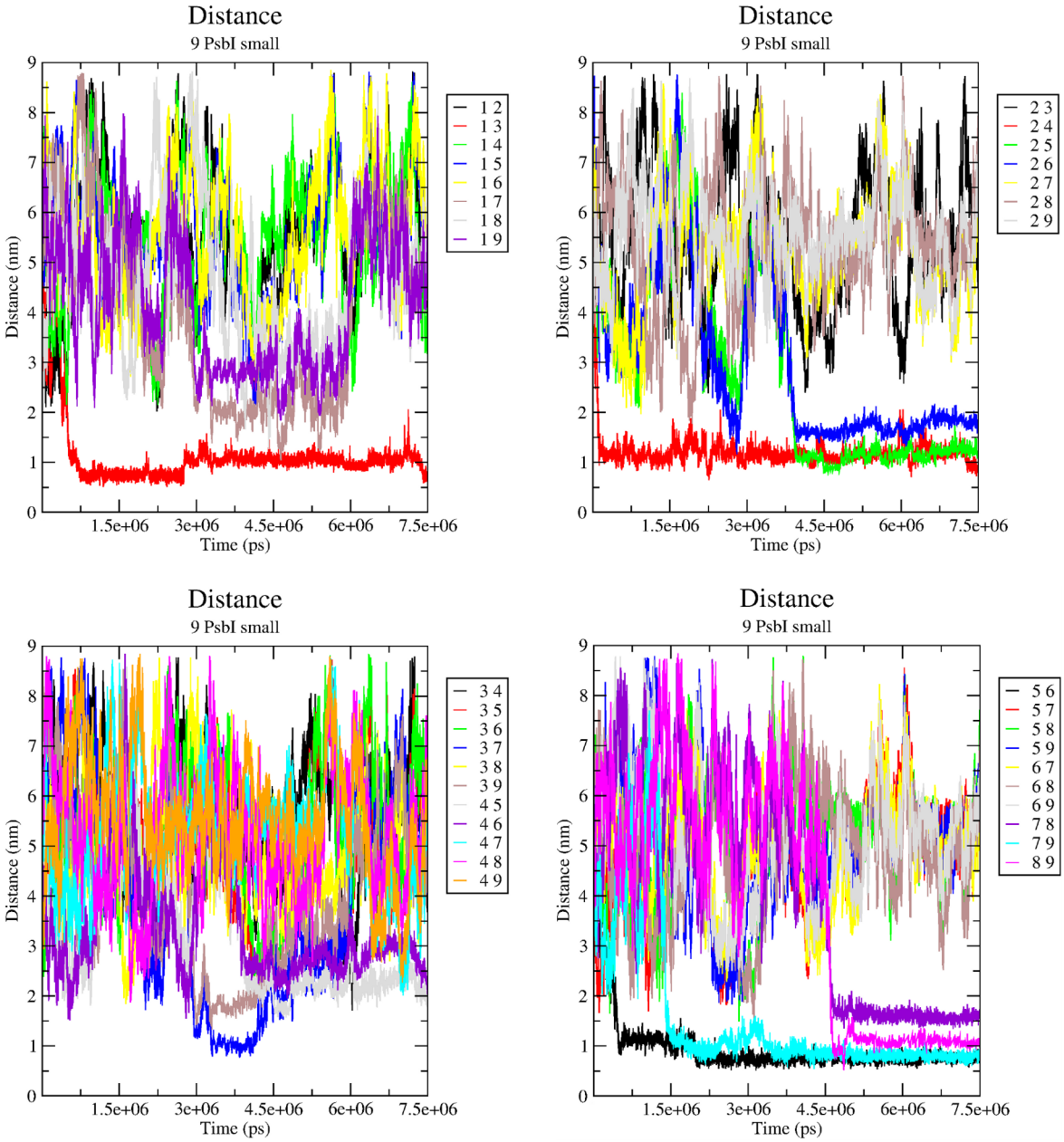


Figure 3.29: *Mutual distance of COM of helices in a small simulation of 9 PsbI units.*

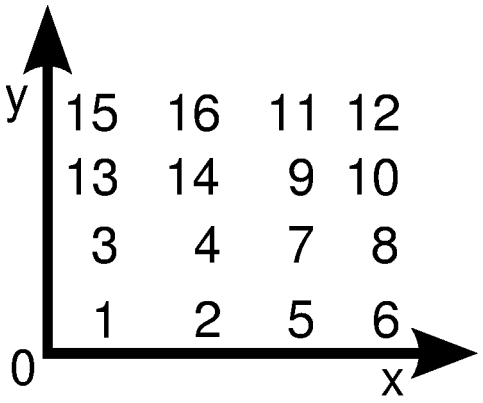


Figure 3.30: *Initial setting of proteins in a system containing 16 PsbI units.*

dimers is not unique. In one case it is the same as described for the system with 2 PsbI units while in the other one the conserved site of one unit interacts with the non-conserved site of the other unit. Tetramerization occurs at about 1.6 μs and is driven by the mutual interaction of the N-termini of the units 2 and 4. The proteins interact via their helices in different modes. Protein 1 interacts with the C-terminal part of its helix only. This trend can be seen in figure 3.27, where the distance of the COM of helix 1 to the COMs of the other proteins is larger than the distances of the others helices with respect to each other.

Similarly to the previous simulations, the behavior of the system containing 9 PsbI proteins (for the initial setting of the simulation see figure 3.28) is driven by an effort of the proteins to group into one multimer. In this system multimer formation is often disturbed by periodic boundary conditions. Not all mutual approaches result in multimerization as can be seen from the graphs showing the mutual distances of the COM of the backbone beads of the helices (see figure 3.29). The inability of two units to establish a dimer could be explained either by the improper orientation of the proteins and/or by proximity of some other protein/complex that influences at least one of the binding partners of the newly forming dimer.

The overall aggregation process starts by formation of four dimers that are stable over the rest of the simulation. The first dimer established is made of units 2 and 4 and is formed at 0.1 μs. The other dimers are 1+3 (since 0.6 μs), 5+6 (0.6 μs) and 7+9 (1.5 μs). In the dimer made by units 2 and 4 the proteins are oriented with their conserved sites towards each other, while in the dimer 5+6 the conserved site is in contact with the non-conserved site of its binding partner. Proteins in the dimer 7+9 are oriented with the conserved site of one monomer towards the non-conserved site of the other protein and in the dimer 1+3 the non-conserved site of one unit interacts with the side of the protein that is perpendicular to the line linking the conserved and the non-conserved sites of the protein (this will be later referred to as a site of the protein). The first attempt of aggregation of dimers into a multimer of a higher order occurs between 1.9 and 2.4 μs, when the dimers 1+3 and 7+9 approach each other and interact by their loops. The other two dimers get to the vicinity of the forming tetramer in the way that all four dimer make a line parallel with the shorter edge of the simulation box, which enables them to interact via periodic boundary conditions. After a short leaving of the dimer 1+3 from the forming multimer, that is probably the result of some instabilities within this octamer, a tetramer 1+3+7+9 is re-established at 2.9 μs. This is connected with the approach of unit 8 to the dimer 5+6. As the result of further rearrangement of the system the tetramer 2+4+5+6 is formed at 3.9 μs. Both tetramers stay relatively close to each other trying to establish a link by their loops. This is another time when the unpaired unit 8 disturbs the multimerization of the system by establishing a helical interaction with the dimer 7+9 at 4.6 μs.

In the first instance this leads to the split of the nonamer into a tetramer and a pentamer (where unit 8 is included) at 5.2 μs. These two multimers get to a mutual proximity at 5.9 μs (as their axis is parallel with the longer edge of the simulation cell, they do not mutually interact via periodic boundary conditions) resulting in a further splitting of the pentamer into a dimer 1+3 and a trimer 7+8+9. In the further course of the simulation mainly the tetramer and the trimer try to aggregate, but no stable interaction is established.

Similarly to the previous system of 9 PsbI proteins, the dimer formation in the simulation containing 16 PsbI proteins (its initial arrangement is shown in figure 3.30) is fast. Most of the dimers are formed within the first half of a microsecond of the simulation. After that dynamics of the system are driven by the effort of the unpaired proteins to multimerize and by accidental interactions of already existing multimers. Like in the previous simulations not all mutual

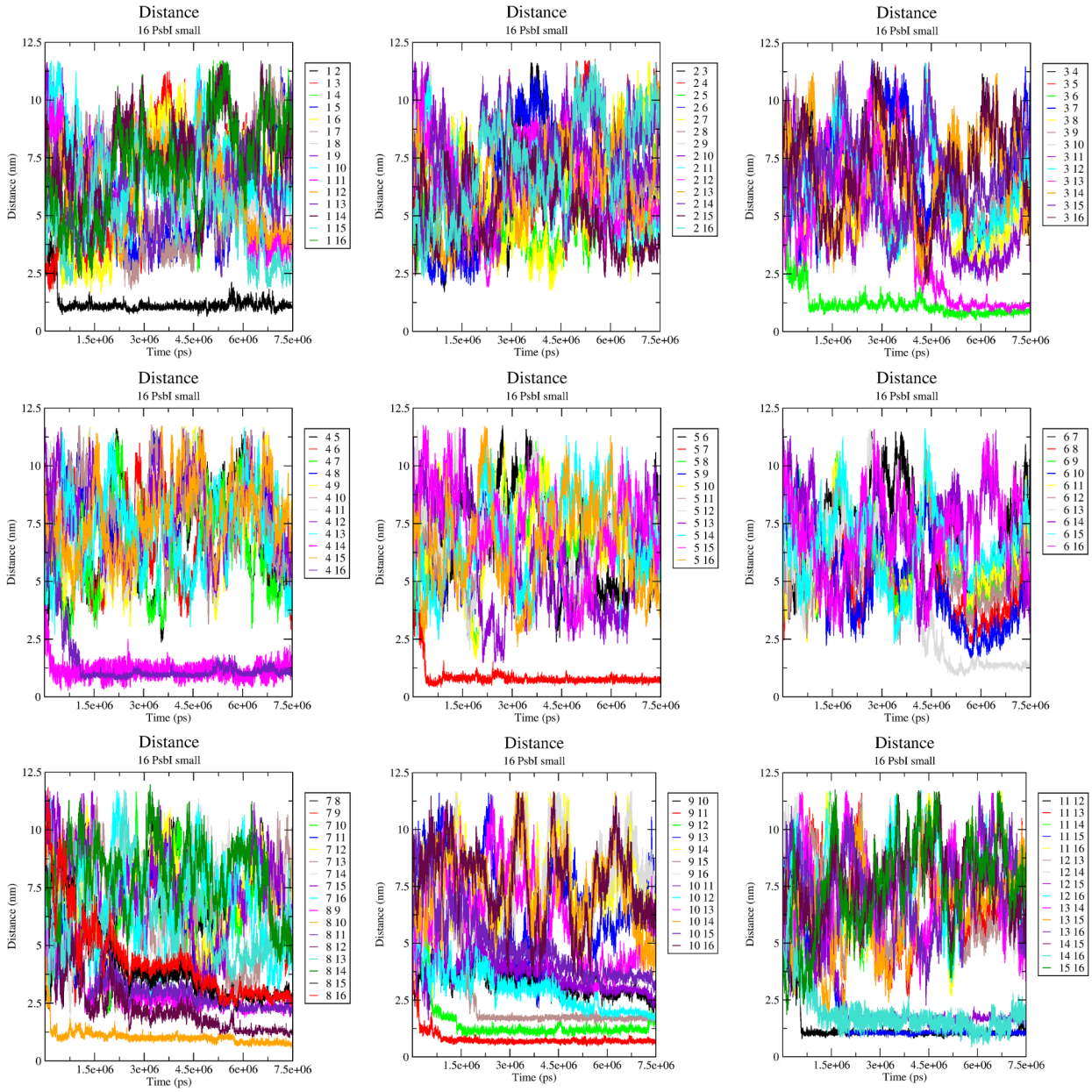


Figure 3.31: *Mutual distance of COM of helices in the small simulation of 16 PsbI units.*

approaches of proteins result in formation of a stable multimer of the higher order. Sometimes multimers via loops are formed. These loop-supported multimers are usually destabilized by the approach of another protein resp. complex, which may result in their splitting. On the other hand, complexes where proteins interact by their helices are stable. Based on stable helical interactions there are two dimers, two trimers and one hexamer in this system at the end of the simulation. Both dimers made of units 1+2 resp. 5+7 are formed independently at approx. 400 ns. Proteins in the dimers are oriented in the same fashion, first by the sides of helices and during some time the helices slightly reorganize to involve at least one conserved site in the interaction.

Trimers are formed in the way, that monomeric PsbI joins an already existing dimer. An example of this behavior is the dimer 4+14 formed at 200 ns, that is at 1.0 μs joined by unit 16 (see figure 3.31). Prior to this, unit 16 approaches the existing dimer at about 0.55 μs but this approach does not lead to firm aggregation, although the units stay close to each other (distance of the COMs of their helices is smaller than 3.7 nm). Then unit 16 returns to the dimer at about 0.7 μs when it is contacted by the loop of unit 14 and embedded into the dimer. This interaction turns into an

interaction on the level of helices at about 1 μs. Although unit 16 attaches to the existing dimer from the side, the resultant trimer is of line-like shape after equilibration, where units are mutually bound in the conserved-non-conserved site manner. An interesting feature of this trimer is that the mutual distance of units 4 and 16 fluctuates less than the one of units 4 and 14 (i.e. the original dimer) and of units 14 and 16. This can be interpreted as a stronger link between units 4 and 16 than between others proteins. During the simulation several other units resp. complexes approach this trimer within a distance where the loop interaction is possible, but a stable multimer is never established.

The second example when a monomer joins an already existing dimer is the case of units 3, 6 and 13. Here, as can be seen from figure 3.31, aggregation starts by formation of the dimer 3+6 that is stabilized by loop interaction after 150 ns. At 750 ns the interaction turns into a helix-helix one. Interaction of units 3 and 6 is first maintained by their conserved sites and during the simulation it drifts to a conserved-non-conserved site interaction. After a short retention close to the dimer at the very beginning of its formation, unit 13 returns to the dimer at 4.0 μs from the side of unit 3. First an interaction on the basis of loops is established that at 5 μs transforms into the interaction on the basis of the helices. This trimer is originally of a triangle-like shape, later it looks more like a line, what is the shape in which the most of the stable multimers are seen. Like in the previous cases, approaches of other proteins do not result in the formation of stable multimers.

Although the last six units can be found as a hexamer at the end of the simulation (see figure 3.31) its formation exhibits similar features with previous complexes. At the very beginning there are two dimers 8+10 and 9+11 formed at approx. 200 ns (both). Unit 12 joins the later dimer at 0.5 μs when its helix interact with the one of unit 11. After equilibration of this trimer, helices 9 and 12 start interacting at 1.3 μs. In the same time interaction of the loop of protein 12 and the helix of protein 8 (that is bound to unit 10) is established for the first time. After leaving the dimer at 1.6 μs, which has to do with the association of the last unit 15 to the trimer 9+11+12 via protein 11, the linkage of dimer 8+10 to the tetramer is renewed at 2.4 μs in the same manner as for the first time. As a result of the random motion within the system, the hexamer is at 5.0 μs sandwiched between the dimer 5+7 trying to interact from the site of unit 15 and trimer 3+6+13 approaching from the side of 8+10. The later interaction outweighs and the trimer joins the hexamer at about 5.8 μs. Due to the periodic boundary conditions dimer 5+7 gets to the site of the trimer 3+6+13 at about 6.3 μs and stays there in an effort to join the complex. No firm linkage is established. This together with approach of dimer 1+2 from the side of unit 15 (that may via the periodic boundary conditions interact with dimer 5+7) leads to destabilization of trimer 3+6+13 within the complex, that departs a bit. Consequently, dimer 1+2 leaves the multimer. At the end of the simulation both, dimer 5+7 and trimer 3+6+13, are within a distance that they may interact with the multimer, but no loop interactions are observed. As a result, the interaction between units 8 and 12 drifts from the loops to the helices at 5.7 μs. Unit 8 is then the only linkage of unit 10 to the rest of the hexamer.

All simulations referred to in this chapter except the 4 PsbI proteins are influenced by periodic boundary conditions that in all cases (except simulations with 2 PsbI proteins) prevent the formation of a multimer of a higher order than are the already existing multimers. This may be one of the reasons why in systems containing 9 and 16 proteins the development is not finished within 7.5 μs. On the other hand, simulations are long enough to describe general processes governing the mutual behavior of PsbI, what was the main task of this project.

PsbI does not like to stay alone in the membrane and the proteins in all systems tend to aggregate. The mutual interaction of units can be twofold: multimers are either stabilized by the loops of the proteins or by their helices. Usually at least one of the binding partners interacts via its conserved site most often with the non-conserved site of its binding partner. Generally, interactions tend to reorient. The only exception is the mutual interaction of the conserved sites. Interaction via helices is more stable than the one by loops. Not all cases when the proteins resp. their complexes get close to each other lead to multimerization, but may induce rearrangement or splitting of an existing multimer.

3.4.7 Structural study of PsbP

Although the crystal structure 1V2B provides a detailed view onto the structure of extrinsic PsbP of PSII, three regions made of amino acids 1 to 15, 89 to 106 and 136 to 140 are unresolved in the crystals. The third missing loop is not expected to be of major interest. Due to its short length (5 amino acids) it is not expected to play an important role in the protein function, but rather connects two secondary structure elements. However, the other two missing loops, mainly the N-terminal one, may be biologically relevant. For this reason an effort to gain a more complete crystal structure of another higher plant studied in our group, spinach (*Spinacia oleracea*), was made by the experimentalists in our laboratory. However, also this structure, deposited in the protein databank with accession code 2VU4 is incomplete, with amino acids 1 to 15, 90 to 107 and 135 to 139 missing. It is not surprising that the unresolved residues are almost the same as in the crystal structure of tobacco.

Molecular modeling followed by MD simulations was used to fill in these gaps. The CyanoP homologue (pdb code: 2XB3) missing the loop between the amino acids 90 and 107 of PsbP from spinach was examined to study the function of the loop mentioned above. This crystal structure is incomplete as well (missing amino acids 1 to 23 and 133 to 137) and so it underwent the same process of homology modeling and energy minimization of the resultant model like PsbP from spinach. Details of the simulation setting are described in paper 1. Although the N-termini of the proteins are expected to be biologically relevant, we gave up the effort to establish their structure, because no structures with high enough homology could be identified and used as templates. This led us to the conclusion, that this loop is not likely to be of an ordered structure and may randomly fluctuate in space in the isolated protein, what was later confirmed by the NMR structure of CyanoP, where this region is highly flexible. When modeled inaccurately, it might disturb the ordered structure of the protein in the homology model as it might stack in an unnatural energetic minimum and so it was better not to have this loop present in the model. The homology modeling of the other loops was easier and successful models were generated. Ten models were computed for each protein, the best of which was chosen with respect to its stereochemical properties and the Modeller objective function and was further examined. The simulation surprisingly shows (for a detailed results description see paper 1) that the loop between amino acids 90 and 107 has a quite rigid structure stabilized by several persitant hydrogen bonds within the loop and even by one persitant hydrogen bond with the rest of the protein. Contrary to PsbP from higher plants, the corresponding loop in the CyanoP homologue is shorter and very flexible. As mentioned in the introduction, CyanoP plays a more peripheral role in photosystem II function and is not specifically attached to the oxygen-evolving complex of PSII. The comparison suggests possible interaction surfaces of PsbP with higher-plant photosystem II. As the large loop it is the only site of major structural difference between the two proteins, these findings suggest that this loop in PsbP may be involved in the interaction of PsbP with PSII, as these interactions must differ in cyanobacteria and higher plants.

For more detailed description of this piece of work see attached paper1:

Kopecky jr V., Kohoutova J., Lapkouski M., Hofbauerova K., **Sovova Z.**, Ettrichova O., Gonzalez-Perez S., Dulebo A., Kaftan D., Kuta Smatanova I., Revuelta J.B., Arellano J.B., Carey J., Ettrich R. (2012): Raman spectroscopy adds complementary detail to the high-resolution X-ray crystal structure of photosynthetic PsbP from Spinacia oleracea. Plos One 7; e46694

4.1 Natural killer cells and missing self hypothesis

Natural killer (NK) cells are large lymphocytes of the innate immune system. The advantage of innate immunity in contrast to the adapted one mediated for instance by small lymphocytes (B-cells and T-cells), is its relatively fast response to the pathogen recognition in order of minutes, while tadapted immunity needs days for full reaction. Penalty payed for the speed is that the innate immunity does not have an immulogical memory (reviewed in e.g. [Horejsi and Bartunkova, 2009].
From the developmental point of view, NK cells are closest to T-cells, with which they share several receptors on their surface, and which are involved in response against similar types of target cells (virally-infected cells and tumour cells for both, NK cells are also involved in stressed cells destruction) and use the same mechanism to destroy target cells (for in-depth review see e.g. [Horejsi and Bartunkova, 2009; Lanier, 2004; Yokoyama, 2005].
NK cells have developed a sophisticated mechanism of target recognition, that is referred to as missing-self hypothesis [Karre et al., 1986]. There are two types of receptors expressed on their surface: activatory and inhibitory ones. Both of them are transmembrane proteins, with an extracellular domain being used for the target cell recognition and a cytoplasmatic part, that participates in the NK cell activation (reviewed in e.g. [Lanier, 2004; Yokoyama, 2005]). Most of but not all the receptors have either an immunoreceptor tyrosine-based inhibitory motif (ITIM) [Daeron et al., 1995] or a short cytoplasmatic domain that is a target for an adaptor molecule (e.g. DAP10) containing an immunoreceptor tyrosine-based activatory motif (ITAM) [Reth, 1989; Cambier et al., 1995] after the receptor activation, based on which their function can be recognized. For both types of receptors phosphorylation of tyrosine in the ITIM resp. ITAM motif is the first step in cell signaling, that activates NK cells resulting in a release of cytokines, porphirins and granzymes that cause apoptosis of the target cell (for review see e.g. [Lanier, 2004; Yokoyama, 2005]). The function of inhibitory receptors is to preserve the killing reaction, while activatory receptors initiate the response and so these receptors act against each other. Potential activation of the NK cells depends on the balance between activatory and inhibitory signals. Obviously, if only activatory receptors are bound, the NK cell attacks the target, while in case of only inhibitory or none receptors bound, the NK cell ignores the target cell. If both, activatory and inhibitory, receptors are bound the behavior of the NK cell depends on which signal outweighs.

4.1.1 NKR-P1

The natural killer receptor P1, NKR-P1 (also known as killer cell lectin-like receptor subfamily B – KLRB or CD161), is the first NK cell receptor discovered [Glimcher et al., 1977] and for a long time it was used as a serological marker of NK cells [Glimcher et al., 1977; Koo and Peppard, 1984]. NKR-P1 is expressed not only by the NK cells, but by some types of T-cells as well [Ballas and Rasmussen, 1990; Vicari and Zlotnik, 1996]. NKR-P1 receptors are type II transmembrane proteins (i.e. they have one transmembrane helix spanning a membrane and the C-terminus of the protein is in the extracellular space), with the C-type lectin-like domain (CTLD) on its C-terminus [Zelensky and Gready, 2005; Kolenko et al., 2011]. In dependance of whether this protein binds Ca^{2+} ions, it may belong either to the group II (bind) or V (not bind) in the CTLDs classification [Drickamer, 1993; Drickamer and Fadden, 2002; Zelensky and Gready, 2004]. *In vivo* occurring NKR-P1 receptor is a 60-kDa homodimer with an intermolecular disulfide bridge in a stalk region [Giorda et al., 1990]. The CTLD is a widely expressed double-looped αβ-fold in metazoa [Zelensky and Gready, 2005]. The first loop is anchored by an anti-parallel β-sheet formed by the N- and C-terminal β-strands. It may (long form) but does not have to (short form) have an N-terminal extension making a β-hairpin with the N-terminal β-strand. The N-terminal β-strand is followed by a loop on which a short β-strand between two helixes is found. This is followed by an other anti-parallel β-sheet made of 3 β-strands, that anchor the second loop, referred to as long loop, that is present in the canonical form of the CTLD and is absent in the compact form. The N-terminal β-strands of the long loop makes an anti-parallel β-sheet with strands from the loop's anchor. At its C-

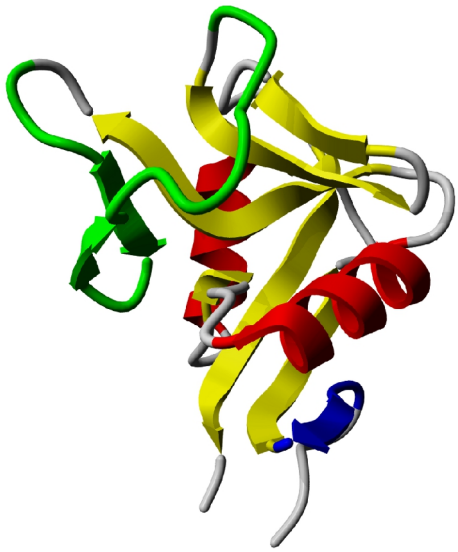


Figure 4.1: *Human receptor CD69 (pdb code: 3HUP) as an example of CTLD.*

terminus the loop is linked by a short loop to the C-terminal β-strand of the outer β-sheet. The NKR-P1 receptors are the long compact form of CTLD (i.e. both the N-terminal β-hairpin and the long loop are present) [Zelensky and Gready, 2005], see figure 4.1.

NKR-P1 is encoded by genes expressed on a so-called NK gene complex [Yokoyama and Seaman, 1993], that is in mice (*Mus musculus*) located on the 6th chromosome [Yokoyama et al., 1991], in rats (*Rattus norvegicus*) on the 4th chromosome [Dissen et al., 1996] and in human (*Homo sapiens*) on the 12th chromosome [Yabe et al., 1993]. Some orthologues of the NKR-P1 were found even in chicken (*Gallus gallus*) and opossum (*Monodelphis domestica*) [Hao et al., 2006]. While in most of the placental mammals (human, dog (*Canis familiaris*), cattle (*Bos taurus*) were examined), there is only one copy of the *Nkr-p1* (and *Clr*, see later) gene present, in rodents (mouse and rat) there is a larger NK gene cluster having more gene copies [Hao et al., 2006]. Namely, four NKR-P1s were reported in rats. Two of them, NKR-P1A and NKR-P1F, are activatory, while the other two, NKR-P1B and NKR-P1G, are inhibitory [Li et al., 2003; Carlyle et al., 2008]. Originally reported PVG rat NKR-P1C receptor [Kveberg et al., 2006], was shown to be a different slicing form of the NKR-P1B in other strands of rats [Kveberg et al., 2009]. These genes make two pair on the chromosome, with NKR-P1A and NKR-P1B are on the centromeric part of the NK-gene cluster, and with the other two, NKR-P1F and NKR-P1G are on the telomeric part of the gene cluster [Voigh et al., 2007; Hao et al., 2006]. Carlyle [Carlyle et al., 2008] speculated about the role of this arrangement in the defense against a viral infection.

There are 6 *Nkrp1* genes, including one pseudogene designed as *Nkrp1e* [Plougastel et al.; 2001], present in the mouse genome [Hao et al., 2006; Plougastel et al.; 2001]. Three of these genes, *Nkrp1a*, *Nkrp1c* and *Nkrp1f* encode an activatory receptor, while the other two, *Nkrp1b* and *Nkrp1g* an inhibitory one [Karlhofer and Yokoyama, 1991; Carlyle et al., 1999; Mesci et al., 2006]. A gene originally reported as *Nkrp1d* in B6 mice [Plougastel et al.; 2001] showed to be a different slicing form of *Nkrp1b* [Carlyle at al., 2006]. The genes on the chromosome are distributed in two group; Nkrp1f makes together with Nkrp1g the telomeric group, while the other make the centromeric group [Plougastel et al.; 2001; Hao et al., 2006].

On chromosomes of both, mice and rat, there are *Nkrp1* genes mixed with C-type lectin-related (*Clr*; also known as osteoclast inhibitory lectin, *Ocil*) receptors, that were shown to be the physiological ligands of NKR-P1 receptors [Iizuka et al., 2003; Carlyle et al., 2004]. Similarly to NKR-P1 receptors, Clr receptors are expressed also as a set of homologues in rodent, but as single protein in other mammals. This distribution of receptors and their ligands may point onto a mutually dependent development of these two groups of proteins [Hao et al., 2006].

In mice it was shown that the NKR-P1B inhibitory receptor binds Clr-b, NKR-P1F binds Clr-g and Clrd/x and NKR-P1G binds Clr-f, Clr-g and Clr-d/x [Carlyle et al., 2004; Iizuka et al., 2003; Aust et al., 2009; Kveberg et al., 2011]. For rats, it is known [Kveberg at al., 2009] that both activatory NKR-P1A and inhibitory NKR-P1B receptors bind the same ligand, Clr-11. Both rat NKR-P1F and NKR-P1G bind Clr-2, Clr-6 and Clr-7. NKR-P1F above this binds also Clr-3 and Clr-4 [Kveberg et al., 2011]. Interestingly, it was also shown [Kveberg et al., 2011], that NKR-P1F and NKR-P1G from mice binds Clrs from rat and that rat NKR-P1F and NKR-P1G binds Clrs from mice.

In human, there is the only copy of NKR-P1, an inhibitory receptor, that binds the only copy of the Clr present in human, designed as LLT1 (lectin-like transcript 1) [Rosen et al., 2005].

4.2 Aims of the project

At the beginning of the project (end of the year 2008) there was no 3D structure of any NKR-P1 known. For further research it was essential to make a homology model of the CTLD of these proteins. The object of modeling was chosen based on proteins with which the collaborating group of Dr. Novak was performing their experiments. These proteins were namely mouse NKR-P1A and NKR-P1C (both activatory) and rat NKR-P1A and NKR-P1B (activatory and inhibitory). To extend our knowledge about the NKR-P1 protein family, I made a sequence analysis of all available NKR-P1 sequence in order to pick conserved regions of a given proteins that may be important either for protein function or for ligand binding. To demonstrate mutual relationship among the given sequences, phylogenetic analysis was performed. As all four sequences in the experimental studies are in the same phylogenetic branch, proteins from two other branches (namely mouse NKR-P1F and human NKR-P1) were added as further targets for homology modeling (and refinement by molecular dynamics simulations) to see the potential differences in evolutionary different proteins. This was my first aim and resulted in paper 2.

Later a crystal structure of mouse NKR-P1A, pdb code 3M9Z [Kolenko et al., 2011], was published. Unfortunately this structure belongs among those few CTLDs where the functionally important long loop is erected and exposed to solvent, what is considered to be an artifact of crystallization [D. Rozbezsky, personal communication]. MD simulations of the given crystal structure showed, that the long loop is unstable in the position presented in crystal structure and that it bends to the opposite site it takes in other CTLD structures [Z. Sovova, unpublished data]. On the other hand, the core of the structure is in agreement with the general CTLD fold. It was further shown that NKR-P1C and not NKR-P1A is the main activatory receptor in mouse. This information together with cross-linking measurements, kindly provided by D. Rozbezsky, Charles University, Praha, motivated us to improve our previously published model in paper 2. This was my second aim and resulted in paper 3.

The position of the long loop in the crystal structure of mouse NKR-P1A is artificial. Hydrogen-deuterium exchange experiments were used to determine its position in solution. To do so, it was necessary to prepare a mutant, where the long loop region is missing. The prediction of this mutant together with verification of its stability was my other task.

NKR-P1 receptors are dimers *in vivo*. However, cutting of CTLD from the stalk region results in the lost of the cysteine participating in the protein dimerization. To get the protein into a more natural dimeric state I tried to design mutations that would result in domain dimerization. Computational modeling together with molecular dynamics and binding energy calculations were used to solve this additional task.

4.3 Modeling and bioinformatics analyses of NKR-P1 proteins

Structural knowledge is an essential characteristic of a protein, that makes design of new experiments on the molecular level easier. In absence of a crystal or an NMR structure, molecular modeling was shown to be a satisfactory tool in protein structure determination.

At the beginning of this project (end of the year 2008) there was no 3D structure of any NKR-P1 receptor known, and the aim of this pivotal study was to compute reliable homology models of various forms of the extracellular domain of this protein and to determine mutual relationship among the NKR-P1 protein family. The last part of this study is dedicated to a sequence analysis performed in order to predict potential binding sites of ligands of this protein.

Phylogeny analysis was performed for various homologues of the CTLD of NKR-P1 from different strains of mice and rats and for human and chicken homologue by five different methods (neighbor-joining, maximum parsimony, maximum likelihood, Fitch-Margoliash method and Bayesian analysis) on both protein and nucleotide sequences of the proteins examined. Various methods were chosen to see if the algorithm used influences the result. The resultant tree (see figure 13 in paper 2) shows separation of the telomeric and the centromeric NKR-P1s of rodents. Both groups of receptors are further clustered according to their function (activatory vs. inhibitory). The non-rodent NKR-P1 makes a separate branch.

Protein sequences used for phylogenetic analysis underwent a sequence analysis in order to identify conserved regions among all the proteins, that may be involved in a ligand binding. There were five conserved regions found. An extra attention was payed to regions designed as III and V that are exposed to the environment and are prone to the interactions with ligands.

Seven CTLDs of NKR-P1 receptors (human, rat NKR-P1A and NKR-P1B, mouse NKR-P1A, NKR-P1C, NKR-P1G, NKR-P1F) were chosen according to the phylogeny analysis to be subjected to molecular modeling. Different template structures were used for various homologue because they have the highest homology with the protein to be modeled. Ten models for each structure were calculated, the best of which was chosen according to the stereochemical properties and the modeller objective function, and was energetically minimized by a short MD simulation (for further details see paper 2). The proteins adopt four different topologies, three of which are adapted by one homologue only. The differences in the topologies occur mainly at the long loop region. This region displays high flexibility but is anchored by conserved sequences, suggesting that its position relative to the rest of the domain might be variable. This loop may contribute to ligand-binding specificity via a coupled conformational transition.

For more detailed information about this project see attached paper 2:

Sovova Z., Kopecky jr. V., Pazderka T., Hofbauerova K., Rozbesky D., Vanek D., Bezouska K., Ettrich R. (2011): Structural analysis of natural killer cell receptor protein 1 (NKR-P1) extracellular domains suggests a conserved long loop involved in ligand specificity. J Mol Model 17; 1353 – 1370

4.4 Refinement of mouse NKR-P1C model

The crystal structure of the CTLD of mouse NKR-P1A (pdb code: 3M9Z) was published in the year 2011. Our model of mouse CTLD of NKR-P1A shows good agreement in reproducing the fold of the core of the domain, however the long loop region in the crystal structure is not attached to its core but is erected and exposed to solvent, what is considered to be an artifact of crystallization, see figure 4.2.

To find the proper position of the long loop region of the mouse NKR-P1C in solution, cross-linking experiments were performed in the collaborating group of Dr. Novak, Charles University, Praha. My task was to make a model of this protein using restraints from experimetal data.

Ten homology models using the crystal structure 3M9Z as the template for the core of the domain and the structure 2YHF [Watson et al., 2011] as the template for the loop were computed. The best model was chosen according to its stereochemical properties and the modeller objective function. The long loop of this model and if necessary lysine residues (targets of the cross-linking agents) were refined by steered-MD in Yasara to adapt a conformation, where the position restraints obtained by the experiment are fulfilled.

Our result confirms that the long loop of the CTLD of mouse NKR-P1C is in solution attached to the core of the domain.

For more detailed information about this project see attached paper 3:

*Rozbesky D., **Sovova Z.**, Marcoux J., Man P., Ettrich P., Robinson C.V., Novak P. (2012): Structural model of lymphocyte receptor NKR-P1C revealed by mass spectrometry and molecular modeling. Anal Chem, accepted, doi: 10.1021/ac302860m*

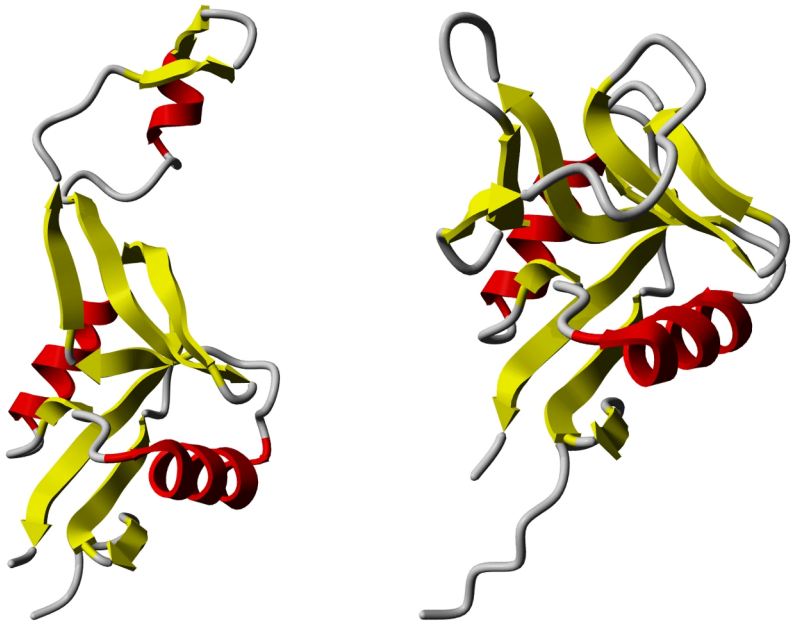


***Figure 4.2:** Comparison of the conformation of the long loop in the crystal structures of mouse NKR-P1A (pdb code: 3M9Z; left) and of human CD69 (pdb code: 3HUP; right).*

4.5 Unpublished results

4.5.1 Methods

Monomer B is one C-terminal amino acid longer than monomer A from the crystal structure 3M9Z [Kolenko et al., 2011] and was used as the input for all mutation studies. All mutations and other protein engineering *in silico*, i.e. point mutations, long loop truncation and setting the initial position of the protein within a dimer, were performed in Yasara software [Krieger et al., 2004]. This output was used as an input for MD simulations in Gromacs 3.3.1 [Lindahl et al., 2001] that was further used for analyses. All the simulations were performed with the Gromacs force field, proteins were hydrated by a solution of SPC water and counterions. Dummy hydrogens were used to enable a timestep of 5 fs and thus to speed up the simulations. Temperature was hold at 300 K by separately connecting protein and solution to the Berendsen temperature bath (τ_t = 0.1 ps), pressure was hold isotropically at 1 bar by connecting the system to the Berendsen pressure bath once per picosecond (compressibility = 4.6e-05 bar^{-1}). Long range electrostatics forces beyond 0.9 nm from the given particle were handled by the PME [Darden et al., 1993], van der Waals interactions were cut beyond 1.4 nm. Periodic boundary conditions were used. Prior to the proper production run, the systems underwent steepest descent (SD) or low-memory Broyden-Fletcher-Goldfarb-Shanno (BFGS) [Byrd et al., 1995] energy minimization (for energy minimization approach used in each simulation, see table 4.1). This was followed by a short 10ps minimization of water with a timestep of 1 fs, position restraints (force = 1000 N) were applied on the protein. After this, systems were neutralized and underwent another 10ps of MD simulation with a tiomestep of 1fs and constraint protein. In the last equilibration simulation, the protein was fully unconstraint and the system was simulated with a timestep of 2 fs for 20 ps. All other run setting were the same as in the production run, that was performed with a timestep of 5fs for times shown in table 4.1.

mutation	minimization	stable	length [ns]
isoleucine	SD	no	10
tryptophan	SD	yes	40
tyrosine	SD	yes	40
methionine	SD	no	5
leucine	BFGS	yes	20
valine	SD	no	10
phenylalanine	SD	yes	20
glutamine	BFGS	yes	40
asparagine	BFGS	no	20
alanine	BFGS	no	10
threonine	SD	no	10
loopless	SD	yes	30

Table 4.1: *Point mutations on position lysine 149 performed in order to find a stable dimer of mouse NKR-P1A. SD stands for steepest descent, BFGS for a low-memory Broyden-Fletcher-Goldfarb-Shanno energy minimization.*

For mutations that led to a stable dimer by the end of the simulation, the binding energy was calculated by the MM/PBSA method [Srinivasan et al., 1998]. The last frame of the previous simulation was used as the input for another simulations were the dummy hydrogens were replaced by the “normal” ones. These simulations with the minimization protocol described above (steepest descent energy minimization was used in all cases) ran with a timestep of 2 fs for 10 ns for both dimers and monomers. The properties of the system were saved every 0.4 ps, the last 1ns of the simulation was used for binding energy calculations. Solvation free energy was calculated by the CHASA web server ([Fleming et al., 2005] roselab.jhu.edu/chasa/runchasa.html), the potential energy and entropy were calculated by tools in Gromacs.

To design the loopless mutant, amino acids between the residues tyrosine 158 and serine 188 were replaced by two alanines. These amino acids were chosen as they are the last ones from the sheet anchoring the long loop region and are close to each other. After 10ns MD simulation and steepest descent energy minimization this monomer was used as an input for dimerization of the loopless mutant. The initial arrangement of the unit was obtained by their 3-D fit onto the crystal dimer. The loopless dimer was equilibred by the above described technique, production ran for 20 ns.

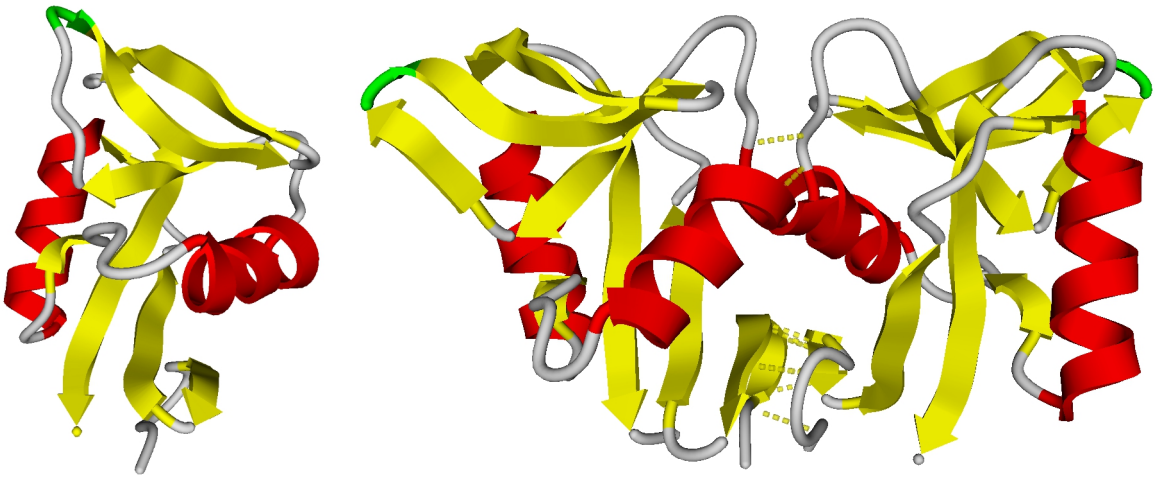


Figure 4.3: *Loopless mutant of mouse NKR-P1A as a monomer (left) and a dimer (right). Alanines replacing the long loop are highlighted in green. Intermolecular hydrogen bonds are shown as yellow disks.*

In the next step, there was an effort to compute a dimer of the complete (i.e. including the long loop region) CTLD of mouse NKR-P1C. The above described mutant of the loopless NKR-P1C was used as the template for the initial arrangement of the monomers in the dimer being constructed. The monomer B of the crystal structure 3M9Z (mouse NKR-P1A) was used as the target for the point mutations. Lysine 149 was chosen as the target for point mutations as in the loopless dimer this amino acid forms mutual hydrogen bond. In case this position would not work, some others amino acids would be tested as the target of the point mutations. For mutations providing stable dimers after 20ns of MD simulation (see table 4.1) the binding free energy was calculated to find the one with the highest binding free energy (and so the most likely dimerizing one). All mutations and setup of the monomers into initial dimer were performed in Yasara. As described above, further molecular dynamics was performed in Gromacs software with the setting described above. Visual inspection was in all cases satisfactory to recognize, if mutated monomers form a dimer of if they departed.
There was no ambition to characterize the behavior of the monomers within the dimer, the aim was to predict mutations that are to be tested experimentally for potential dimerization.

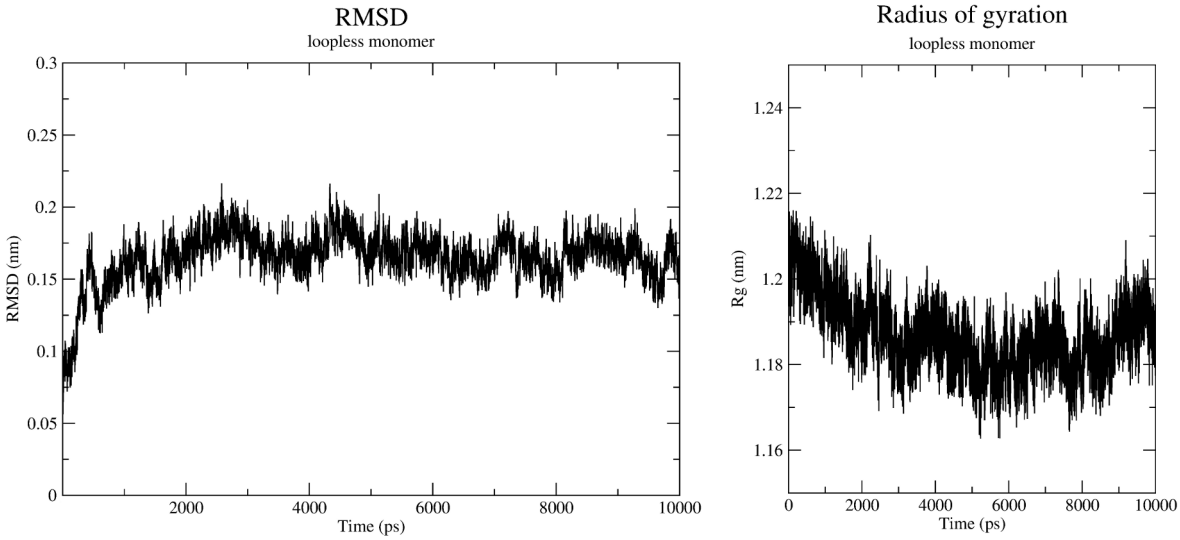


Figure 4.4: *RMSD and radius of gyration for C_α atoms of loopless monomer of NKR-P1A confirm the system reached equilibrium.*

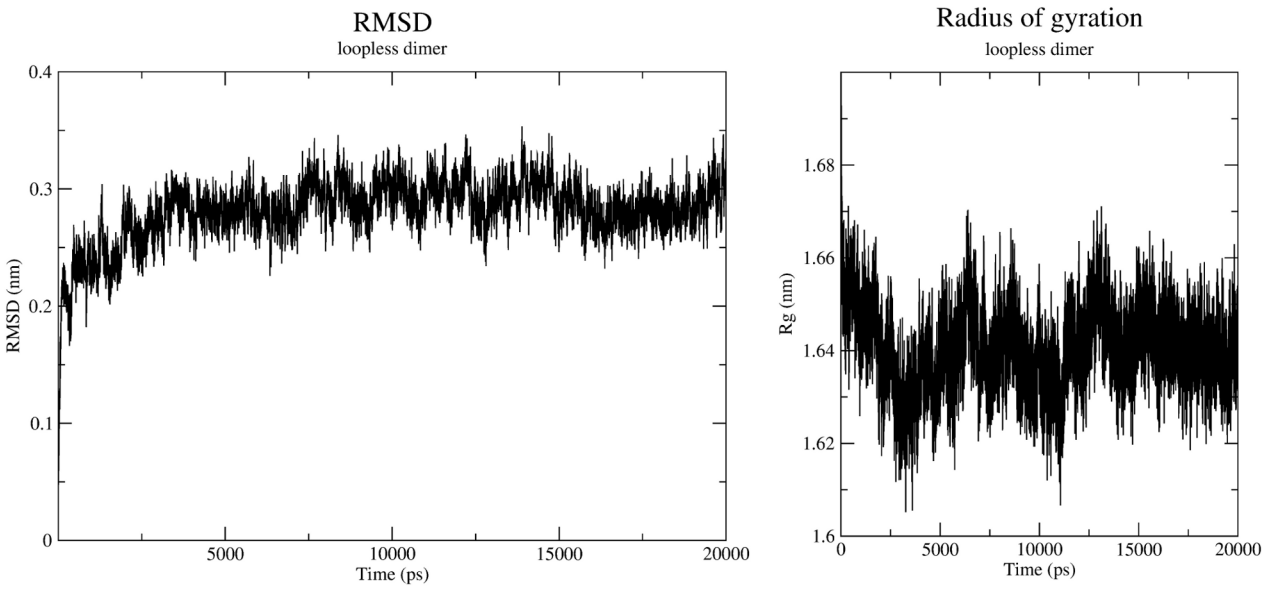


Figure 4.5: *RMSD and radius of gyration for C_α atoms of loopless dimer of NKR-P1A confirm the system reached equilibrium.*

4.5.2 Results

The resultant structure of the loopless of CTLD mouse NKR-P1A after 30 ns of MD simulation is shown in figure 4.3. As can be seen, this structure preserves the CTLD fold what is also confirmed by the calculations of RMSD and radius of gyration of C_α carbons, both of which are in equilibrium, see figure 4.4.

As the existence of the protein in its monomeric state was at the time when these simulations were performed (spring 2009) considered to be unnatural, major attention was payed to the dimer construction once a stable loopless construct of the CTLD of mouse NKR-P1A was constructed.

The stable loopless dimer, see figure 4.3, is formed within 20 ns of MD simulation. Equilibration of the given system is confirmed by RMSD and radius of gyration of C_α, see figure 4.5. After the final energy minimization, some hydrogen bonds between the monomers are present, mainly between their amino acids in the N-terminus region and the region around the mutation. The binding free energy of this loopless mutant, shown in table 4.2, is negative, what confirms that this mutant represents a stable dimer under the given conditions. Contrary to the point mutants of the complete NKR-P1 there is no dominant component in the total binding energy, although the potential energy component is the highest one.

There were 11 mutations of lysine 149 tested to see if the proposed mutation leads to formation of a stable dimer, see table 4.1. The mutation was considered to be stable, if the monomers stayed in the vicinity of each other after system equilibration determined by the RMSD and the radius of gyration (data not shown). Unstable attempts will not be discussed further. All the mutants dimerize in a similar mode to the one of the loopless mutant (data not shown).

	K149Y		**K149F**		**K149L**		**loopless**	
	monomer	*dimer*	*monomer*	*dimer*	*monomer*	*dimer*	*monomer*	*dimer*
potential [J]	4090	9457	4849	9803	4951	9732	7165	4001
ΔG [J]	2795	4726	2418	4403	2803	4981	3668	2178
entrophy [J]	205	1432	1899	1416	1957	686	603	535
total [J]	7090	15615	9166	15622	9711	15399	11436	6714
result [J]	1435		-2710		-4023		-1992	

Table 4.2: *Binding free energies for NKR-P1A mutants examined.*

As can be seen from table 4.1, mutation of lysine 149 to tryptophan, tyrosine, leucine, phenylalanine and glutamine resulted in stable (according the the criterion above) dimers. To be able to calculate binding free energies of the dimer, simulations of the monomer were necessary to be performed. Interestingly, when lysine 149 is mutated to a tryptophan and a glutamine, simulations with explicit hydrogens were not able to run with a timestep of 2 fs. Neither more extensive relaxation of the system, including NVT simulation did not help and so these two mutants were excluded from further examination.
Binding free energies (incl. its individual components) for the other three mutants of lysine 149 to tyrosine, leucine and phenylalanine are listed in table 4.2. These values indicate that only the mutants to leucine and to phenylalanine are stable, while the free energy for the mutants with tyrosine is more favorable, when the two proteins make separate monomers instead of a dimer. In both the cases, entropy is the main contribution of the total binding energy, while the potential energy component is the lower one.

4.5.3 Discussion and conclusion remarks

The proposed mutations leading to dimerization of the CTLD domain of mouse NKR-P1A were experimentally tested in the collaboration group of Dr. Novak, Charles University, Praha. The predicted loop-less mutant forms a stable monomer in solution but it does not spontaneously dimerize as proposed. This could by either explained by that in MD simulation the initial setting of monomers was somehow unnatural for the protein, probably some energetic barrier was violantely overcome or by the mutual interaction of more protein units that restrain the establishment of the stable dimer. The complete CTLDs on the other hand make large aggregates during the refolding and it was impossible to purify just the dimers from the solution. A variant with more proteins units were not tested by MD simulations.

5. References

Akerlund H.E., Jansson C. (1981) Localization of a 34 000 and a 23 000 M_r polypeptide to the lumenal side of the thylakoid membrane. FEBS Lett. 124; 229 – 232

Akerlund H.E., Jansson C., Andersson B. (1982) Reconstruction of photosynthetic water splitting in inside-out thylakoid vesicle and identification of a participating polypeptide. BBA 681; 1 – 10

Allen M.P., Tildesley D.J. (1987) Computer simulation of liquids. Oxford University Press, Oxford

Aust J.G., Gays F., Mickiewicz K.M., Buchanan E., Brooks C.G. (2009) The expression and function of the NKRP1 receptor family in C57BL/6 mice. J Immunol 183; 106 – 116

Baena-Gonzalez E., Aro E.M. (2002) Biogenesis, assembly and turnover of photosystem II units. Philos Trans R Soc Lond B Sci Biol 357; 1451 – 1459

Baker D., Sali A. (2001) Protein structure prediction and structural genomics. Science 294; 93 – 96

Ballas Z.K., Rasmussen W. (1990) NK1.1+ thymocytes. Adult murine CD4-, CD8- tyhymocytes contain an NK1.1+, CH3+, CD5hi, TCR-V beta 8+ subset. J Immunol 145; 1039 – 1045

Balsera M., Arellano J.B., Revuelta J.L., Hermose J.A. (2005) The 1.49 Å Resolution Crystal Structure of PsbQ from Photosystem II of *Spinacia oleracea* Reveals a PPII Structure in the N-terminal Region. JMB 350; 1051 – 1060

Barber J. (1985) in Topics in Photosynthesis. Vol. 6: Photosynthetic mechanisms and the environment Eds: Barber J., Baker N.R. pp 91 – 134, Elsevier, Amsterdam

Barber J., Andersson B. (1992) Too much of a good thing: light can be bad for photosynthesis. Trends Biochem Sci 17; 61 – 66

Barber J., Iwata S. (2005) Refined X-ray structure of photosystem II and its implications. In: Wydrzynski T., Satoh K. eds. Photosystem II: the light-driven water:plastoquinone oxidoreductase. Springer: The Netherlands, 469 – 489

Bautista J.A., Tracewell C.A., Schlodder E., Cunningham F.X., Brudvig G.W., Diner B.A. (2005) Construction and characterization of genetically modified *Synechocystis* sp. PCC6803 photosystem II core complex containing carotenoids with shorted pi-conjugation than beta-carotene. JBC 280; 38839 – 38850

Berendsen H.J.C., Postma J.P.M., van Gunsteren W.F., DiNola A., Haak J.R. (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81; 3684 – 3691

Boehm M., Romero E., Reisinger V., Yu J., Komenda J., Eichacker L.A., Dekker J.P., Nixon P.J. (2011) Investigating the Early Stages of Photosystem II Assembly in *Synechocystis* sp. PCC 6803. JBC 286(17); 14812 – 14819

Boekema E.J., van Breeman J.F.L., van Roon H., Dekker J.P. (2000) Conformational changes in in photosystem II supercomplexes upon removal of extrinsic subunits. Biochemistry 39; 12907 – 12915

Bondarava N., Beyer P. Krieger-Liszakay A. (2005) Function of the 23 kDa extrinsic protein of photosystem II as a manganese binding protein and its role in photoactivation. BBA 1708; 63 – 70

Born M., von Karman T. (1912) Uber Schwingungen in Raumgittern. Physik Z 13; 297 – 309

Bricker T.M., Frankel L.K. (2003) Carboxylate groups on the manganase-stabilizing protein are required for efficient binding of the 24 kDa extrinsic protein to photosystem II. Biochemistry 42; 2056 – 2061

Bricker T., Roose J.L., Fagerlund R.D., Frankel L.K., Eaton-Rye J.J. (2012) The extrinsic proteins of Photosystem II. BBA 1817; 121 – 142

Britt R.D. (2001) Oxygen production. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Broser M., Gabdulkhakov A., Kern J., Guskov A., Muh F., Saenger W., Zouni A. (2010) Crystal Structure of Monomeric Photosystem II from *Thermosynechococcus elongatus* at 3.6-Å Resolution. JBC 285; 26255 – 26262

Burda K., Kruk J., Borgstadt R., Stanek J., Strzalka K., Schmid G.H., Kruse O. (2003) Mössbauer studies of the non-heme iron and cytochrome *b*559 in a *Chlamydomonas reinhardtii* PSI− mutant and their interactions with α-tocopherol quinone. FEBS Lett 535(1 – 3); 159 – 165

Burnap R., Sherman L.A. (1991) Deletion mutagenesis in *Synechocystis* sp. PCC 6803 indicates that the Mn-stabilizing protein of photosystem II is not essential for O_2 evolution. Biochemistry 30; 440 – 446

Bussi G., Donadio D., Parrinello M. (2007) Canonical sampling through velocity rescaling. J Chem Phys 126; 014101 – 014108

Byrd R.H., Lu P., Nocedal J. (1995) A limited memory algorithm for bound constrained optimization. SIAM J Scientif Statistic Comput 16; 1190 – 1208

Cambier J., Daeron M., Fridman W.H., Gergely J., Kinet J.P., Klausner R., Lynch R., Malissen B., Pecht I., Reinherz E., Ravetch J., Reth M., Samelson L., Sandor M., Schreiber A., Seed B., Terhorst C., van de Winkel J., Weiss A. (1995) New nomenclature for the Reth motif (or ARH1/TAM/ARAM/YXXL). Immunol Today 16; 110

Carlyle J.R., Martin A., Mehra A., Attisano L., Tsui F.W., Zuniga-Pflucker J.C. (1999) Mouse NKR-P1B, a novel NK1.1 antigen with inhibitory function. J Immunol 162; 5917 – 5923

Carlyle J.R., Jamieson A.M., Gasser S., Clingan C.S., Arase H., Raulet D.H. (2004) Missing self-recognition of Ocil/Clr-b by inhibitory NKR-P1 natural killer cell receptors. PNAS 101; 3527 – 3532

Carlyle J.R., Mesci A., Ljutic B., Belanger S., Tai L.-H., Rousselle E., Troke A., Proteau M.-F., Markigiannis A.P. (2006) Molecular and genetic basis for strain-dependent NK1.1 alloreactivity of mouse NK cells. J Immunol 176; 7511 – 7524

Carlyle J.R., Mesci A., Fine J.H., Chen P., Belanger S., Tai L.-H., Makrigiannis A.P. (2008) Evolution of the Ly49 and Nkrp1 recognition systems. Seminars Immunol 20; 321 – 330

Cauchy M.A. (1847) Methode generale pour la resolution des systemes dequations simultanees. Comptes Rendus Hebd Acad Sci 25; 536 – 538

Chaloub R. M., Silva L. M. T., Rodrigues M. A. R., Dos Santos C. P. (2003) Phase transition of thylakoid membranes modulates photoinhibition in the cyanobacterium *Anabaena siamensis*. Photosynth. Res. 78; 143 – 152

Chapman D.J., De-Felice J., Barber J. (1983) Growth temperature effects on thylakoid membrane lipid and protein content of pea chloroplasts. Plant Physiol 72; 225 – 228

Chothia C., Lesk A.M. (1986) The relation between the divergence of sequence and structural in proteins. EMBO J 5; 823 – 826

Cornell W.D., Cieplak P., Bayly C.I., Gould I.R., Merz jr. K.M., Ferguson D.M., Spellmeyer D.C., Fox T., Caldwell J.W., Kollman P.A. (1995) A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J Am Chem Soc 117; 5179 – 5197

Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. (2004) WebLogo: A sequence logo generator. Genome Res. 14; 1188 – 1190

Cullis P.R., Hope M.J., Tilcock C.P.S. (1986) Lipid polymorphism and role of lipids in membranes. Chem Phys Lipids 40; 127 – 144

Daeron M., Latour S., Malbec O., Espinosa E., Pina P., Pasmans S., Fridman W.H. (1995) The same tyrosine-based inhibitory motif, in intracytoplasmic domain of FcγRIIB, regulates negatively BCR-, TCR-, and FcR-dependant cell activation. Immunity 3; 635 – 646

Darden T., York D., Pedersen L. (1993) Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. J Chem Phys 98; 10089 – 10092

Daum B., Kuhlbrandt W. (2011) Electron tomography of plant thylakoid membranes. J. Exp. Bot. 62; 2393 – 2402

Dekker J.P., Boekema E.J. (2005) Supramolecular organization of thylakoid membrane proteins in green plants. BBA 1706; 12 – 39

Dietrich C., Bagatolli L.A., Volovyk Z.N., Thompson N.L., Levi M., Jacobson K., Gratton E. (2001) Lipid rafts reconstituted in model membranes. Biophys J 80; 1417 – 1428

Diner B.A., Ries D.F., Cohen B.N., Metz J.G. (1988) COOH-terminal processing of polypeptide D1 of the photosystem II reaction center of Scenedesmus obliquus is necessaty for the assembly of the oxygen-evolving complex. JBC 263; 8972 – 8980

Dissen E., Ryan J.C., Seaman W.E., Fossum S. (1996) An autosomal dominant locus, Nka, mapping to the Ly-49 region on a rat natural killer (NK) gene complex, controls NK cell lysis of allogeneic lymphocytes. J Exp Med 183; 2197 – 2207

Dobakova M., Tichy M., Komenda J. (2007) Role of the PsbI protein in photosystem II assembly and repair in the cyanobacterium *Synechocystis* sp. *PCC 6803*. Plant Physiol 145; 1681 – 1691

Dormann P., Hoffmann-Benning S., Balbo I., Benning C. (1995) Isolation and characterization of an arabidopsis mutant deficient in the thylakoid lipid digalactosyl diacylglycerol. Plant Cell 7; 1801 – 1810

Dormann P. (2007) Galactolipids in plant membranes. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Douce R., Joyard J. (1990) Biochemistry and function of the plastid envelope. Annu. Rev. Cell Biol. 6; 173 – 216

Drickamer K. (1993) Evolution of Ca2+-dependent animal lectins. Prog Nucleic Acid Res Mol Biol 45; 207 – 232

Drickamer K., Fadden A.J. (2002) Genomic analysis of C-type lectins. Biochem Soc Symp 69; 59 – 72

Enami I., Kikuchi S., Fukuda T., Ohta H., Shen J.R. (1998) Binding and functional properties of four extrinsic proteins of photosystem II from a red alga, *Cyanidium caldarium*, as studied by release-reconstruction experiments. Biochemistry 37; 2787 – 2793

Epand R.F., Savage P.B., Epand R.M. (2007) Bacterial lipid composition and the antimicrobial efficacy of cationic steroid compounds (Ceragenins). BBA 1768; 2500 – 2509

Ferreira K.N., Iverson T.M., Maghlaoui K., Barber J., Iwata S. (2004) Architecture of the photosynthetic oxygen-evolving center. Science 303; 1831 – 1838

Filippov A., Oradd G., Lindblom G. (2004) Lipid lateral diffusion in order and disorder phases in raft mixtures 86; 891 – 896

Fleming P.J., Fitzkee N.C., Mazei M., Srinivasan R., Rose G.D. (2005) A novel method reveals that solvent water favors polyproline II over β-strand conformation in peptides and unfolded proteins: conditional hydrophobic accessible surface area (CHASA). Protein Sci 14; 111 – 118

Gawrish K. (2005) The dynamics of membrane lipids. In: The structure of biological membranes. Ed. Yeagle P.L. CRC press, Boca Raton, USA, pp. 147 – 171

Ghanotakis D.F., Topper J.N., Yocum C.F. (1984) Structural organization of the oxidizing side of photosystem II. Exogenous reductants reduce and destroy the Mn-complex in photosystem II membranes depleted of the 17 and 23 kDa polypeptides. BBA 767; 524 – 531

Giorda R., Rudert W.A., Vavassori C., Chambers W.H., Hiserodt J.C., Trucco M. (1990) NKR-P1, a signal transduction molecule on natural killer cells. Science 249; 1298 – 1300

Giroud C., Siegenthaler P.A. (1988) Development of oat prothylakoids into thylakoid during greening does not change transmembrane galactolipid asymmetry but preserves the thylakoid bilayer. Plant Physiol. 88; 412 – 417

Glimcher L., Shen F.W., Cantor H. (1977) Identification of a cell-surface antigen selectively expressed on the killer cell. J Exp Med 145; 1 – 9

Gombos Z., Kanervo E., Tsvetkova N., Sakamota T., Aro E. M., Murata N. (1997) Genetic enhancement of the ability to tolerate photoinhibition by introduction of unsaturated bonds into membrane glycerolipids. Plant. Physiol 115; 551 – 559

Gounaris K., Barber J., Harwood J.L. (1986) The thylakoid membranes of higher plant chloroplasts. Biochem. J. 237; 313 – 326

Govindjee, Kern J.F., Messinger J., Whitmarsh J. (2010) Photosystem II. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Gruner S.M. (2005) Nonlamellar lipid phases. In: The structure of biological membranes. Ed. Yeagle P.L. CRC press, Boca Raton, USA, pp. 173 – 199

Guidotti G. (1972) Membrane proteins. Annu. Rev. Biochem. 41; 731 – 752

van Gunsteren W.F., Bakowies D., Baron R., Chandrasekhar I., Christen M., Daura X., Gee P., Geerke D.P., Glattli A., Hunenberger P.H., Kastenholz M.A., Oostenbrink C., Schenk M., Trzesniak D. van der Vegt N.F.A., Yu H.B. (2006) Biomolecular modeling: goals, problems, perspectives. Angew Chem Int Ed 45; 4064 – 4092

Guskov A., Kern J., Gabdulkhakov A., Broser M., Zouni A., Saenger W. (2009) Cyanobacterial photosystem II at 2.9-Å resolution and the role of quinones, lipids, channels and chloride. Nat Struct Mol Biol 16; 334 – 342

Hao L., Klein J., Nei M. (2006) Heterogeneous but conserved natural killer receptor gene complex in four major orders of mammals. PNAS 103; 3192 – 3197

Hartel H., Dormann P., Benning C. (2001) Galactolipids not associated with the photosysthetic apparatus in phosphate-deprived plants. J. Photochem. Photobiol. 61; 46 – 51

Hashimoto A., Ettinger W.F., Yamamoto Y., Theg S.M. (1997) Assembly of newly imported oxygen-evolving complex subunits in in isolated chloroplasts: sites of assembly and binding mechanism. Plant Cell 9; 441 – 452

Hauser H., Poupart G. (2005) Lipid strucure. In: The structure of biological membranes. Ed. Yeagle P.L. CRC press, Boca Raton, USA, pp. 1 – 51

Hazel J.R. (1995) Thermal adaptation in biological membranes: Is homeoviscous adaptation the explanation? Annu Rev Physiol 57; 19 – 42

He Q.F., Vermaas W. (1998): Chlorophyll a availability affects *psbA* translation and D1 precursor processing in vivo in *Synechocystis* sp. PCC6803. PNAS 95; 5830 – 5835

Heinz E. (1977): Enzymatic reactions in galactolipid biosynthesis. In: Lipids and lipid polymers in higher plants. Eds: Tevini M., Lichtenthaler K.H.; Springer-Verlag, Berlin; pp. 102 – 120

Hess B., Kutzner C., van der Spoel D., Lindahl E. (2008) GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. JCTC 4; 435 – 447

Hinner M.J., Marrink S.J., de Vries A.H. (2009): Location, Tilt, and binding: a molecular dynamics study of voltage-sensitive dyes in biomembranes. J Phys Chem B 113; 15807 – 15819

Hockney R.W. (1970) The potential calculation and some applications. Methods Comput Phys 9; 136 – 211

Horejsi V., Bartunkova J (2009) *in*: Zaklady imunologie (Basics of immunology), Triton, Praha; textbook in czech

Ifuku K., Sato F. (2001) Importance of the N-terminal sequence of the extrinsic 23 kDa polypeptide in photosystem II in ion-retention in oxygen evolution. BBA 1546; 196 – 204

Ifuku K., Nakatsu T., Kato H., Sato F. (2004) Crystal structure of the PsbP protein of photosystem II from Nicotiana tabacum. EMBO reports 5; 362 – 367

Ifuku K., Yamamoto Y., Ono T.A., Ishihara S., Sato F. (2005a) PsbP protein, but not PsbQ protein, is essential for the regulation and stabilization of photosystem II in higher plants. Plant Physiol 139; 1175 – 1184

Ifuku K., Nakatsu T., Shimamoto R., Yamamoto Y., Ishihara S., Kato H., Sato F. (2005b) Structure and function of PsbP protein of photosystem II from higher plants. Photosynth Res 84; 251 – 255

Ifuku K., Ishihara S., Shimamoto R., Ido K., Sato F. (2008) Structure, function, and evolution of the PsbP protein family in higher plants. Photosynth Res 98; 427 – 437

Iizuka K., Naidenko O.V., Plougastel B.F.M., Fremont D.H., Yokoyama W.M. (2003) Genetically linked C-type lectin-related ligands for the NKRP1 family of natural killer cells. Nat Immunol 4; 801 – 807

Ikuechi M., Inoue Y. (1988a) A new 4.8-kDa polypeptide intrinsic to the PS II reaction center, as revealed by modified SDS-PAGE with improved resolution of low-molecular-weight proteins. Plant Cell Physiol 29; 1233 – 1239

Ikuechi M., Inoue Y. (1988b) A new photosystem II reaction center component (4.8 kDa protein) is encoded by chloroplast genome. FEBS Lett 241; 99 – 104

Ikuechi M., Shukla V.K., Parkasi H.B., Inoue Y. (1995) Directed inactivation of the *psbI* gene does not affect photosystem II in the cyanobacterium *Synechocystis* sp. *PCC 6803*. Mol Gen Genet 249; 622 – 628

Inagaki N., Yamamoto Y., Satoh K. (2001) A sequential two-step proteolytic process in the carboxyl-terminal truncation of precursor D1 protein in *Synechocystis* sp PCC6803. FEBS Lett 509; 197 – 201

Ishihara S., Takabayashi A., Endo T., Ifuku K., Sato F. (2007) Distinct functions for the two PsbP-like proteins PPL1 and PPL2 in the chloroplast thylakoid lumen of Arabidopsis. Plant Physiol 145; 688 – 679

Iwai M., Katoh H., Katayama M., Ikuechi M. (2004) PSII-Tc protein plays an important role in dimerization of photosystem II. Plant Cell Physiol 45; 1809 – 1816

Iwai M, Katayama M., Ikuechi M. (2006) Absence of the *psbH* gene product destabilizes the pohotosystem II complex and prevents association of the photosystem II-X protein in the thermophilic cyanobacterium *Thermosynechococcus elongatus* BP-1. Photosynthetic Res. 87; 313 – 322

Iwai M., Suzuki M., Kamiyama A., Sakurai I., Dohmae N., Inoue Y., Ikuechi M. (2010) The PsbK subunit is requested for the stable assembly and stability of other small subunits in the PSII complex in the thermophilic cyanobacterium *Thermosynechococcus elongatus* BT-1. Plant Cell Physiol 51; 554 – 560

Israelachvili J.N. (1973) Theoretical considerations on the asymmetric distribution of charged phopholipid molecules on the inner and outer layers of curved bilayer membranes. BBA 323; 659 – 663

Israelachvili J.N., Mitchell, D.J., Ninham B.W. (1977) Theory of self-assembly of lipid bilayers and vesicles. BBA 470; 185 – 201

Israelachvili J.N., Marcelja S., Horn R.G. (1980) Physical principles of membrane organization. Quar Rev Biophys 13; 121 – 200

Jackson S.A., Hinds M.G., Eaton-Rye J.J. (2012) Solution sructure of cyanoP from Synechocystis sp. PCC6803: New insights on the structural basis for functional specialization amongst PsbP family proteins. BBA 1817; 1331 – 1338

Janero D.R., Barrnett R. (1981) Cellular and thylakoid-membrane glycolipids of *Chlamydomonas reinhardtii* 137^{+}. J. Lipid. Res. 22; 1119 – 1125

Jorgensen W.L., Maxwell D.S., Tirado-Rives J. (1996) Development and testing the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118; 11225 – 11236

Kabsch W., Sander C. (1985) Identical pentapeptides with different backbones. Nature 317; 19 – 25

op den Kamp J.A.F. (1979): Lipid asymmetry in membranes. Ann Rev Biochem 48; 47 – 71

Karlhofer F.M., Yokoyama W.M. (1991) Stimulation of murine natural killer (NK) cell by a monoclonal antibody specific for the NK1.1 antigen. IL-2-activated NK cells possess additional specific stimulation pathways. J Immunol 146; 3662 – 3673

Karre K., Ljunggrren H.G., Piontek G., Kiessling R. (1986) Selective rejection of H-2-deficient lymphoma variants suggests alternative immune defence strategy. Nature 319; 675 – 678

Kashino Y., Lauber W.M., Carroll J.A., Wang Q., Whitmarsh J., Satoh K., Pakrasi H.B. (2002) Proteomic analysis of a highly active Photosystem II preparation from the cyanobacterium Synechocystis sp. PCC6803 reveals the presence of novel polypeptides. Biochemistry 41; 8004 – 8012

Kawakami K., Umena Y., Iwai M., Kawabata Y., Ikuechi M., Kamiya N., Shen J.R. (2011) Roles of PsbI and PsbM in photosystem II dimer formation and stability studied by deletion mutagenesis and X-ray crystallography. BBA 1807; 319 – 325

Kenyon C.N, Rippka R., Stanier R.Y. (1972) Fatty acid composition and physiological properties of some filamentous blue-green algae. Arch. Microbiol. 83; 216 – 236

Kirchhoff H., Haferkamp S., Allen J.F., Epstein D.B.A., Mullineaux C.W. (2008) Protein diffusion and macromolecular crowding in thylakoid membranes. Plant Physiol. 146; 1571 – 1578

Kok B., Forbush B., McGloin M. (1970) Cooperation of charges in photosynthetic oxygen evolution. Photochem. Photobiol 11; 457 – 475

Kolenko P., Rozbezsky D., Vanek O., Kopecky jr. V., Hofbauerova K., Novak P., Pompach P., Hasek J., Skalova T., Bezouska K., Dohnalek J. (2011) Molecular architecture of mouse activating NKR-P1 receptors. J Struct Biol 175; 434 – 441

Komenda J., Reisinger V., Muller B.C., Dobakova M., Granvogl B., Eichacker L.A. (2004) Accumulation of the D2 protein is a key regulatory step for assembly of the photosystem II reaction center complex in *Synechocystis PCC6803*. JBC 279; 48620 – 48629

Komenda J., Tichy M., Eichacker L.A. (2005) The PsbH protein is associated with the inner antenna CP47 and facilitates D1 processing and incorporation into PSII in the cyanobacterium *Synechocystis* PCC 6803 . Plant Cell Physiol. 46(9); 1477 – 1483

Komenda J., Kuvikova S., Granvogl B., Eichacker L.A., Diner B.A., Nixon P.J. (2007) Cleavage after residue Ala352 in the C-terminal extension is an early step in the maturation of the D1 subunit of photosystem II in *Synechocystis PCC6803*. BBA 1767; 829 – 837

Komenda J., Nickelsen J., Tichy M., Prasil O., Eichacker L.A., Nixon P.J. (2008) The cyanobacterial homologue of HCF136/YCF48 is a component of an early photosystem II assembly complex and is important for both the efficient assembly and repair of photosystem II in *Synechocystis* sp PCC6803. JBC 283; 22390 – 22399

Komenda J., Sobotka R., Nixon P.J. (2012) Assembly and maintaining the photosystem II complex in chloroplasts and cyanobacteria. Curr Opin Plant Biol 15; 245 – 251

Koo G.C., Peppard J.R. (1984) Establishment of monoclonal anti-Nk-1.1 antibody. Hybridoma 3; 301 – 303

Korn E.D. (1969) Cell membranes: structure and synthesis. Annu. Rev. Biochem. 38; 263 – 288

Krieger E., Darden T., Nabuurs S.B., Finkelstein A., Vriend G. (2004) Making optimal use of empirical energy functions: force-field parametrization in crystal space. Proteins 57; 678 – 683

Kveberg L., Back C.J., Dai K.-Z., Inngjerdingen M., Rolstad B., Ryan J.C., Vaage J.T., Naper C. (2006) The novel inhibitory NKR-P1C receptor and Ly49s3 identify two complementary, functionally distinct NK cell subsets in rats. J Immunol 176; 4133 – 4140

Kveberg L., Dai K.-Z., Westgaard I.H., Daws M.R., Fossum S., Naper C., Vaage J.T. (2009) Two major groups of rat NKR-P1 receptors can be distinguished based on chromosomal localization, phylogenetic analysis and Clr ligand binding. Eur J Immunol 39; 541 – 551

Kveberg L., Dai K.-Z., Inngjerdingen M., Brooks C.G., Fossum S., Vaage J.T. (2011) Phylogenetic and functional conservation of the NKR-P1F and NKR-P1G receptors in rat and mouse. Immunogenetics 63; 429 – 436

Koynova R. (1997) Liquid crystalline metastability of phosphatidylglycerols. Chem Phys Lipids 89; 67 – 73

Kunstner P., Guardiola A., Takahashi Y., Rochaix J.D. (1995) A mutant strain of *Chlamydomonas reinhardtii* lacking the chloroplast photosystem II psbI gene grows photoautotrophically. J Biol Chem 270; 9651 – 9654

Kuwabara T., Suzuki K. (1995) Reversible changes in conformation of the 23-kDa protein of photosystem II and their relationship to the susceptibility of the protein to a proteinase from photosystem II membranes. Plant Cell Physiol 36; 495 – 504

Kyte J., Doolittle R.F. (1982) A simple method for displaying the hydrophatic character of a protein. J. Mol. Biol. 157; 105 – 132

Lanier L.L. (2004) NK cell recognition. Annu Rev Immunol 23; 225 – 274

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23; 2947 – 2948

Lennard-Jones J.E. (1924) On the determination of molecular fields. Proc R Soc Lond A 106; 463 – 477

Lewis R.N.A.H., McElhaney R.N. (2005) The mesomorphic phase behavior of lipid bilayers. In: The structure of biological membranes. Ed. Yeagle P.L. CRC press, Boca Raton, USA, pp. 53 – 120

Li J., Rabinovich B.A., Hurren R., Shannon J., Miller R.G. (2003) Expression, cloning and function of the rat NK activating and inhibitory receptors NKR-P1A and -P1B. Int Immunol 15; 411 – 416

Lindblom G., Rilfors L. (1989) Cubic phases and isotropic structures formed by membrane lipids – possible biological relevance. BBA 988; 221 – 256

Lindahl E., Hess B., van der Spoel D. (2001) Gromacs 3.0: a package for molecular simulation and trajectory analysis. J Mol Model 7; 306 – 317

Lins R.D., Hunenberger P.H. (2005) A new GROMOS force field for hexopyranose-based carbohydrates. J Comput Chem 26; 1400 – 1412

Loll B., Kern J., Saenger W., Zouni A., Biesiadka J. (2005) Towards complete cofactor arrangement in the 3.0 Å resolution structure of photosystem II. Nature 438; 1040 – 1044

Mabbitt P.D., Rautureau G.J.P., Day C.L., Wilbanks S.M., Eaton-Rye J.J., Hinds M.G. (2009) Solution Structure of Psb27 from Cyanobacterial Photosystem II. Biochemistry 48; 8771 – 8773

MacKerell jr. A.D., Bashford D., Bellott M., Dunbrack jr. R.L., Evanseck J.D., Field M.J., Fischer J., Gao J., Guo H., Ha S., Joseph-McCarthy D., Kuchnir L., Kuczera K., Lau F.T.K., Mattos C., Michnick S., Ngo T., Nguyen D.T., Prodhom B., Reiher III W.E., Roux B., Schlenkrich M., Smith J.C., Stote R., Straub J., Watanabe M., Wiorkiewicz-Kuczera J., Yin D, Karplus M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 102; 3585 – 3616

MacKerell jr. A.D. (2004) Empirical force fields for biological macromolecules: overview and issues. J Comput Chem 25; 1584 – 1604

Marrink S.J., Risselada H.J., Mark A.E. (2005) Simulation of gel phase formation and melting in lipid bilayers using a coarse grained model. Chem Phys Lip 135; 223 – 244

Marrink S.J., Risselada H.J., Yefimov S., Tieleman D.P., de Vries A.H. (2007) The MARTINI force field: coarse grained model for biomolecular simulations. J Phys Chem B 111, 7812 – 7824

Mason L.H., Willettebrown J., Anderson S.K., Gosselin P., Shores E.W., Love P.E., Ortaldo J.R., McVicar D.V. (1998) Characterization of an associated 16-kDa tyrosine phosphoprotein required for Ly-49D signal transduction. J Immunol 160; 4148 – 4152

Matsumoto K., Sakai H., Ohta K., Kameda H., Sugawara F., Abe M., Sakaguchi K. (2005) Monolayer membranes and bilayer vesicles characterized by α- and β-anomer of sulfoquinovosyldiacylglycerol (SQDG). Chem Phys Lipids 133; 203 – 214

Meades G.D. jr., McLachlan A., Sallans L., Limbach P.A., Frankel L.K., Bricker T. (2005) Association of the 17-kDa extrinsic protein with photosystem II in higher plants. Biochemistry 44; 15216 – 15221

van de Meene A.M.L., Hohmann-Marriott M.F., Vermaas W.F.J., Roberson R.W. (2006) The three-dimensional structure of the cyanobacterium *Synechocystis* sp. PCC6803. Arch Microbiol 184; 259 – 270

van Meer G., Voelker D.R., Feigenson G.W. (2008) Membrane lipids: where they are and how they behave. Nat Rev Mol Cell Biol 9; 112 – 124

Mereschkowsky C. (1905) Über Natur und Ursprung der Chromatophoren im Pflanzenreiche. *Biol Centralbl* 25: 593 – 604

Mesci A., Ljutic B., Makrigiannis A.P., Carlyle J.R. (2006) NKR-P1 biology. From prototype to missing self. Immunol Res 35; 13 – 26

Michoux F., Takasaka K., Boehm M., Nixon P.J., Murray J.W. (2010) Structure of CyanoP at 2.8 Å: Implications for the evolution and function of the PsbP subunit of photosystem II. Biochemistry 49; 7411 – 7413

Miyao M., Murata N. (1983a) Partial reconstruction of the photosynthetic oxygen evolution system by rebinding of 33-kDa polypeptide. FEBS Lett 164; 375 – 378

Miyao M., Murata N. (1983b) Partial disintegration and reconstruction of the photosynthetic oxygen evolution system. Binding of 24kilodalton and 18 kilodalton polypeptide. BBA 725; 87 – 93

Miyao M., Murata N. (1984) Calcium ions can be substituted for the 24-kDa polypeptide in photosynthetic oxygen evolution. FEBS Lett 168; 118 – 120

Monticelli L., Kandasamy S.K., Periole X., Larson R.G., Tieleman D.P., Marrink S.J. (2008) The MARTINI coarse-grained force field: extension to proteins. JCTC 4; 819 – 834

Morein S., Andersson A.S., Rilfors L., Lindblom G. (1996) Wild-type *Escherichia coli* cells regulate the membrane lipid composition in a "window" between gel and non-lamellar structures. JBC 271; 6801 – 6809

Murakami R., Ifuku K., Takabayashi A., Shikanai T., Endo T., Sato F. (2002) Characterization of an *Arabidopsis thaliana* mutant with impaired psbO, one of two genes encoding extrinsic 33-kDa proteins in photosystem II. FEBS Lett 523; 138 – 142

Murata N., Wada H., Gombos Z. (1992) Modes of fatty-acid desaturation in cyanobacteria. Plant Cell Physiol. 33; 933 – 941

Nanba O., Satoh K. (1987) Isolation of a photosystem II reaction center consisting of D-1 and D-2 polypeptides and cytochrome *b*-559. PNAS 84; 109 – 112

Newton I. (1687) Philosophiae naturalis principia mathematica. London, book in latin

Nickelsen J., Rengstl B., Stengel A., Schottkowski M., Soll J., Ankele E. (2010) Biogenesis of the cyanobacterial thylakoid membrane system – an update. FEMS Microbiol Lett 315; 1 – 5

Nielsen L.E., Kadavy D.R., Rajagopal S., Drijber R., Nickerson K.W. (2005) Survey of extreme solvent tolerance in Gram-positive cocci: membrane fatty acid changes in *Staphylococcus haemolyticus* grown in toluene. Appl Environ Microbiol 71; 5171 – 5176

Nishihara M., Yokota K., Kito M. (1980) Lipid molecular species composition of thylakoid membranes. BBA 617; 12 – 19

Nixon P.J., Michoux F., Yu J., Boehm M., Komenda J. (2010) Recent advances in understanding the assembly and repair of photosystem II. Ann. Bot. 106(1); 1 – 16

Oostenbrink C., Villa A., Mark A.E., van Gunsteren W.F. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameters set 53A5 and 53A6. J Comput Chem 25; 1656 – 1676

Orr G. R., Raison J. K. (1987) Compositional and thermal properties of thyalakoid polar lipids of *Nerium oleander* L. in relation to chilling sensitivity. Plant. Physiol. 84; 88 – 92

Ort D.R., Kramer D. (2009) Photosynthesis. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Palinska K.A. (2008) Cyanobacteria. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Patel G.B., Sprott G.D. (2005) Archaeal membrane lipids. In: Encyclopedia of life sciences (ELS). John Wiley and Sons, Ltd: Chichester

Plougastel B., Matsumota K., Dubbelde C., Yokoyama W.M. (2001) Analysis of a 1-Mb BAC conting overlapping the mouse *Nkrp1* cluster of genes: cloning of three new *Nkrp1* members, *Nkrp1d*, *Nkrp1e*, and *Nkrp1f*. Immunogenetics 53; 592 – 598

Ponder J.W., Case D.A. (2003) Force fields for protein simulations. In: Advances in protein chemistry. Vol. 66; 27 – 81

Rawyler A., Siegenthaler P.A. (1981) Transmembrane distribution of phospholipids and their involvement in electron transport, as revealed by phospholipase A_2 treatment of spinach thylakoids. BBA 635; 348 – 358

Rawyler A., Siegenthaler P.A. (1985) Transversal localization of monogalactosyldiacylglycerol and digalactosyldiacylglycerol in spinach thylakoid membranes. BBA 815; 287 – 298

Rawyler A., Unitt M.D., Giroud C., Davies J.P., Mayor J.P., Harwood J.L., Siegenthaler P.A. (1987) The transmembrane distribution of galactolipids in chloroplast thylakoid is universal in a wide variety of temperate climate plants. Photosynth. Res. 11; 3 – 13

Reth M. (1989) Antigen receptor tail clue. Nature 338; 383 – 384

de las Rivas J., Balsera M., Barber J. (2004) Evolotion of oxygenic photosynthesis: genome-wide analysis of OEC extrinsic proteins. Trends Plant Sci 9; 18 – 25

Rodgers J.M., Sorensen J., de Meyer F.J.M., Schiott B., Smit B. (2012) Understanding the phase bahavior of coarse-grained model lipid bilayers through computational calorimetry. J Phys Chem B 116; 1551 – 1569

Roffey R.A., Theg S.M. (1996) Analysis of important cerbonyl-terminal truncations of the 23-kilodalton subunit of the oxygen-evolving complex suggests that its structure is an important determinant for thylakoid transport. Plant Physiol 111; 1329 – 1338

Rog T., Vattulainen I., Bunker A., Karttunen M. (2007) Glycolipid membranes through atomistic simulations: effect of glucose and galactose head groups on lipid bilayer properties. J Phys Chem B 111; 10146 – 10154

Roose J.L., Pakrasi H.B. (2004) Evidence that D1 processing is required for manganase binding and and extrinsic protein assembly into photosystem II. JBC 279; 45417 – 45422

Roose J.L., Wegener K.M., Parkasi H.B. (2007) The extrinsic proteins of Photosystem II. Photosynth Res 92; 369 – 387

Rosen D.B., Bettadapura J., Alsharifi M., Mathew P.A., Warren H.S., Lanier L.L. (2005) Cutting edge: Lectin-like transcript-1 is a ligand for the inhibitory human NKR-P1A receptor. J Immunol 176; 7796 – 7799

Sakurai I., Shen J.R., Leng J., Ohashi S., Kobayashi M., Wada H. (2006) Lipids in oxygen-evolving photosystem II of cyanobacteria and higher plants. J. Biochem. 140; 201 – 209

Sali A., Blundell T.L. (1993) Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234; 779 – 815

Sato N., Murata N. (1982) Lipid biosynthesis in the blue-green alga, *Anabaena variabilis* II. Fatty acids and lipid molecular species. BBA 710; 279 – 289

Sato N. (2010) Phylogenomic and structural modeling analyses of PsbP superfamily reveal multiple small segment additions in evolution of photosystem II-associated PsbP protein in green plants. Mol. Phylogenet Evol 56; 176 – 186

Schwenkert S., Umate P., Dal Bosco C., Volz S., Mlcochová L., Zoryan M., Eichacker L.A., Ohad I., Herrmann R.G., Meurer J. (2006) PsbI affects the stability, function, and phosphorylation patterns of photosystem II assemblies in tobacco. J Biol Chem 281; 34227 – 34238

Sen A., Mannock D.A., Collins D.J., Quinn P.J., Williams W.P. (1983) Thermotropic phase properties and structure of 1,2-distearoylgalactosylglycerols in aqueous systems. Proc. R. Soc. Lond. B 218; 349 – 364

Shen J.R., Inoue Y. (1993) Binding and functional properties of two new extrinsic components, cytochrome c-550 and a 12-kDa protein, in cyanobacterial photosystem II. Biochemistry 32; 1825 – 1832

Shipley G.G., Green J.P., Nichols B.W. (1973) The phase behavior of monogalactosyl, digalactosyl, and sulphoquinovosyl diglycerides. BBA 331; 531 – 544

Siegenthaler P.A., Giroud C. (1986) Transversal distribution of phospholipids in prothylakoid and thylakoid membranes from oat. FEBS Lett. 201; 215 – 219

Singer S.J., Nicolson G.L. (1972) The fluid mosaic model of the structure of cell membranes. Nature 175; 720 – 731

Singhal G.S., Renger G., Sopory S.K., Irrgang K.D., Govindjee (1999) In: Concepts in photobiology: photosynthesis and photomorphogenesis. Narosa Publishers, New Dehli and Kluwer Academic, Dordrecht; 11 – 51

Shi L.X., Hall M., Funk C., Schroder W.P. (2012) Photosystem II, a growing complex: Updates on newly discovered components and low molecular mass proteins. BBA 1817; 13 – 25

van der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A.E., Berendsen H.J.C. (2005) GROMACS: fast, flexible, and free. J Comput Chem 26; 1701 – 1718

van der Spoel D., Lindahl E., Hess B., Kutzner C., van Buuren A.R., Apol E., Meulenhoff P.J., Tieleman D.P., Sijbers A.L.T.M., Feenstra K.A., Berendsen H.J.C. (2006) Gromacs user manual, version 4.0

Sundby C., Larsson C. (1985) Transbilayer organization of the thylakoid galactolipids. BBA 813; 61 – 67

Suorsa M., Sirpio S., Allahverdiyeva Y., Paakkarinen V., Mamedov F., Styring S., Aro E.M. (2006) PsbR, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. JBC 281; 145 – 150

Suzuki T., Minagawa J., Tomo T., Sonoike K., Ohta H., Enami I. (2003) Binding and functional properties of the extrinsic proteins in oxygen-evolving photosystem II particle from a green alga, *Chlamydomonas reinhardtii* having his-tagged CP47. Plant Cell Physiol 44; 76 – 84

Srinivasan J., Cheatham III T.E., Cieplak P., Kollman P.A., Case D.A. (1998) Continuum solvent studies of the stability of DNA, RNA and, phosphoramidate-DNA helices. J Am Chem Soc 120; 9401 – 9409

Sveshnikov D., Funk C. Schroder W.P. (2007) The PsbP-like protein (sll1418) of Synechocystis sp. PCC6803 stabilizes the donor side of photosystem II. Photosynth Res 93; 101 – 109

Takahashi T., Inoue-Kashino N., Ozawa S., Takahashi Y., Kashino Y., Satoh K. (2009) Photosystem II complex in vivo is a monomer. JBC 284; 15598 – 15606

Thornton L., Ohkawa H., Roose J., Kashino Y., Keren N., Pakrasi H. (2004) Homologs of plant PsbP and PsbQ proteins are necessary for regulation of Photosystem II activity in the cyanobacterium Synechocystis PCC6803. Plant Cell 16; 2164 – 2175

Tilcock C.P.S. (1986) Lipid polymorphism. Chem. Phys. Lipids 40; 109 – 125

Tomo T., Emani E., Satoh T. (1993) Orientation and nearest neighbor analysis of *psbI* gene product in the photosystem II reaction center complex using bifunctional cross-linkers. FEBS Lett 323; 15 – 18

Tomo T., Akimoto S., Tsuchiya T., Fukuya M., Tanaka A., Mimuro M. (2009) Replacement of chlorophyll with di-vinylchlorophyll in the antenna and reaction center complexes of the cyanobacterium *Synechocystis* sp. PCC 6803: Characterization of spectral and photochemical properties. BBA 1787; 191 – 200

Umena Y., Kawakami K., Shen J.R., Kamiya N. (2011) Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. Nature 473; 55 – 60

Unitt M.D., Harwood J.L. (1985) Sidedness studies of thylakoid phosphatidylglycerol in higher plants. Biochem. J. 228; 707 – 711

Verlet L. (1967) Computer "experiments" on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. Phys Rev 159; 98 – 103

Vermeer L.S., de Groot B.L., Reat V., Milon A., Czaplicky J. (2007) Acyl chain order parameter profiles in phospholipid bilayers: computational from molecular dynamics simulations and comparison with ^{2}H NMR experiments. Eur Biophys J 36; 919 – 931

de Vitry C., Olive J., Drapier M., Recouvreur M., Wollman F.A. (1989) Post-translational events leading to the assembly of photosystem II protein complex a study using photosynthesis mutants from *Chlamydomonas reinhardtii*. J Cell Biol 109; 991 – 1006

Vodrazka Z. (2002) In: Biochemie. Academia, Praha textbook in Czech

Voigt S., Mesci A., Ettinger J., Fine J.H., Chen P., Chou W., Carlyle J.R. (2007) Cytomegalovirus evasion of innate immunity by subversion of the NKR-P1B:Clr-B missing-self axis. Immunity 26; 617 – 627

Vicari A.P., Zlotnik A. (1996) Mouse NK1.1+T cells: a new family of T cells. Immunol Today 17; 71 – 76

Wada H., Murata N. (1989) *Synechocystis* PCC 6803 mutants defective in desaturation of fatty acids. Plant Cell Physiol. 30; 971 – 978

Watson A.A., Lebedov A.A., Hall B.A., Fenton-May A.E., Vagin A.A., Dejnirattisai W., Felce J., Mongkolsapaya J., Palma A.S., Liu Y., Feizi T., Screaton G., Murshudov G.N., O'Callaghan C.A. (2011) Structural flexibility of the macrophage dengue virus receptor CLEC5A. Implications for ligand finding and signaling. JBC 286; 24208 – 24218

Webb M.S., Green B.R. (1991) Biochemical and biophysical properties of thylakoid acyl lipids. BBA 1060; 133 – 158

Webber A.N., Packham L., Chapman D.J., Barber J., Gray J.C. (1989a): A fifth chloroplast-encoded polypeptide is present in the photosystem II reaction center complex. FEBS Lett 242; 259 – 262

Webber A., Packman L.C., Gray J.C. (1989b) A 10kDa polypeptide associated with the oxygen-evolving complex of Photosystem II has a putative C-terminal non-cleavable thylakoid transfer domain. FEBS Lett 242; 435 – 438

White D.A. (1973) The phospholipid composition of mammalian tissues. In: Form and function of phospholipids. Eds: Ansell G.B., Hawthorne J.N., Dawson R.M.C; Elsevier, Amsterdam; pp 441 – 482

Wolf M.G., Hoefling M., Aponte-Santamaria C., Grubmueller H., Groenhof G. (2010) g membed: Efficient insertion of a membrane protein into an equilibrated lipid bilayer with minimal perturbation. J Comput Chem 31; 2169 – 2174

Xu H., Vavilin D., Vermaas W. (2001) Chlorophyll b cam serve as the major pigment in functional photosystem II complexes of cyanobacteria. PNAS 98; 14168 – 14173

Yabe T., McSherry C., Bach F.H., Fisch P., Schall R.P., Sondel P.M., Houchins J.P. (1993) A multigene family on human chromosome 12 encodes natural killer-cell lectins. Immunogenetics. 37; 455 – 460

Yao D., Kieselbach T., Komenda J., Promnares K., Prieto M.A., Tichy M., Vermaas W., Funk C. (2007) Localization of the small CAB-like proteins in photosystem II. JBC 282; 267 – 276

Yeagle P.L. (2005) The role of cholesterol in the biology of cell. In: The structure of biological membranes. Ed. Yeagle P.L. CRC press, Boca Raton, USA, pp. 243 – 254

Yi X., Hargett S.R., Liu H., Frankel L.K., Bricker T.M. (2007) The PsbP protein is required for photosystem II complex assembly/stability and photoautotrophy in Arabidopsis thaliana. J Biol Chem 282; 24833 – 24841

Yokoyama W.M., Ryan J.C., Hunter J.J., Smith H.R.C., Stark M., Seaman W.E. (1991) cDNA cloning of mouse NKR-P1 and genetic linkage with Ly-49. Identification of a natural killer cell gene complex on mouse chromosome 6. J Immunol 149; 3229 – 3236

Yokoyama W.M., Seaman W.E. (1993) The Ly-49 and NKR-P1 gene families encoding lectin-like receptors on natural killer cells: the NK gene complex. Annu Rev Immunol 11; 613 – 635

Yokoyama W. M. (2005) Natural killer cell immune responses. Immunol Res 32; 317 – 325

Zak E., Norling B., Maitra R., Huang F., Andersson B., Parkasi H.B. (2001) The initial steps of biogenesis of cyanobacterial photosystems occur in plasma membrane. PNAS 98; 13443 – 13448

Zelensky A.N., Gready J.E. (2004) C-type lectin-like domains in Fugu rubripes. BMC Genomics 5; 51 – 72

Zelensky A.N., Gready J.E. (2005) The C-type lectin-like domain superfamily. FEBS J 272; 6179 – 6217

Zhao W., Rog T., Gurtovenko A.A., Vattulainen I., Karttunen M. (2008) Role of phosphatidylglycerols in the stability of bacterial membranes. Biochimie 90; 930 – 938