ABSTRACT

Title of Dissertation: COMPUTATIONAL SIMULATIONS ON MEMBRANES AND A TRANSMEMBRANE PROTEIN

Xiaohong Zhuang, Doctor of Philosophy, 2016

Dissertation directed by: Associate Professor Jeffery B. Klauda
Department of Chemical and Biomolecular Engineering

To accurately model the transmembrane proteins, accurate descriptions of its natural environment, i.e., lipids, are critical. The all-atom CHARMM36 lipid force field (C36FF-AA) is tested with molecular dynamics (MD) simulations. Through comparison to experiments, we conclude that the C36FF-AA is accurate for use with bilayers of varying head and chain types over biologically relevant temperatures. The united-atom chain model of the C36FF (C36FF-UA) of common lipids is developed to improve simulation efficiency. It shows good agreement between the simulated bilayer properties obtained by C36FF-UA and experiments, and also between the simulated results from UA and AA lipid models. Besides the single-component membrane, multiple-components 18:2 linoleoyl-containing soybean membrane models have been developed. The structural properties of pure linoleoyl bilayers agree well with experiments, based on which the soybean membrane models also result in reasonable structural properties.

Accurate lipid force field greatly facilitates the study of transmembrane proteins. Lactose permease of *Escherichia coli* (*E. coli*) belongs to major facilitator superfamily (MFS) which is the largest and most diverse family of transporters and serves as a model for secondary active transporters (SATs) in this dissertation. LacY structures of the cytoplasmic-open, occluded-like,

and recently periplasmic-partially-open state have been determined, however, the crystal structure of LacY in the periplasmic-open state is still not available. The periplasmic-open LacY structure is important for understanding the complete proton/sugar transport process of LacY as well as other similar SAT proteins. MD simulations are performed to test the accuracy of the previously developed periplasmic-open LacY$^{\text{IM-EX}}$ model (*JMB* 404:506), and two other periplasmic-open LacY models, LacY$^{\text{SW}}$ and LacY$^{\text{FP}}$ models (*JMB* 407:698). The simulated results indicate that LacY$^{\text{IM-EX}}$ is the only structure that remains stable in the periplasmic-open state. The MD dummy spin label simulations (MDDS) have also been performed and the results show that the orientation of the spin labels significantly affect the distance measurement so that the proper interpretation of DEER requires the aid of MDDS simulations. Self-guided Langevin dynamics (SGLD) simulations are performed to search periplasmic-open LacY. The results show that no outward-facing is obtained with nanosecond-averaged results, but if we study individual structures, conformational sampling is obtained with certain SGLD parameters that enhance natural helical motions. This SGLD approach might hold promise for studying conformational changes of other SAT proteins.

COMPUTATIONAL SIMULATIONS ON MEMBRANES AND A TRANSMEMBRANE PROTEIN

by

Xiaohong Zhuang

Dissertation submitted to the Faculty of the Graduate School of
the University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctorate of Philosophy
2016

Advisory Committee:

Associate Professor Jeffery B. Klauda, Chair
Professor Sergei Sukharev
Assistant Professor Amy J. Karlsson
Associate Professor Panagiotis Dimitrakopoulos
Patrick and Marquerite Sung Professor Srinivasa R. Raghavan

DEDICATION

I dedicate my Ph.D. dissertation to my parents and my husband
for their great love and support.

ACKNOWLEDGEMENTS

Many thanks to Dr. Jeffry B. Klauda, for his wise guidance, great patience, terrific understanding, and timely help. With his mentorship, I have learned much that prepare me to be a qualified scientific researcher.

I acknowledge Dr. Norbert Kučerka for sharing the x-ray and neutron scattering data and SIMtoEXP software; Dr. H. Ronald Kaback for sharing the DEER experimental data; Dr. Wonpil Im and Dr. Shahidul Islam for sharing and discussing on the MDDS simulation.

I also acknowledge the staffs of Department of Chemical and Biomolecular Engineering at the University of Maryland for their assistance throughout my Ph.D. program, and the exporters at Division of Information Technology at the University of Maryland for helping me out with the software installations and module applications.

I would like to thank my doctoral committee members for their valuable time and insightful discussions on my research work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Abbreviation	**Expansion**
C36FF-AA	All-Atom CHARMM36 Force Field
C36FF-UA	United-Atom CHARMM36 Force Field
D_{B}	Overall bilayer thickness
DBSCAN	Density-Based Spatial Clustering of Applications with Noise
D_{C}	Half of hydrophobic thickness
D_{cut}	Cutoff Distance of lipid clustering
D_{HH}	Headgroup-to-headgroup thickness
DEER	Double Electron-Electron Resonance
FRET	Fluorescence Resonance Energy Transfer
IM-EX	Implicit-Explicit membrane
K_{A}	Area compressibility modulus
LacY	Lactose Permease of *E. coli*
MD	Molecular Dynamics
MDDS	Molecular Dynamics Dummy Spin-label
MFS	Major Facilitator Superfamily
NFF	Neutron Form Factor
$N_{\mathrm{HB}}^{\mathrm{inter}}$	Number of Inter-lipid Hydrogen Bonds
$N_{\mathrm{HB}}^{\mathrm{intra}}$	Number of Intra-lipid Hydrogen Bonds
NMR	Nuclear Magnetic Resonance
NPT	isothermic-isobaric ensemble
PBC	Periodic Boundary Conditions
PME	Particle Mesh Ewald
SA/lip	Surface Area per Lipid
SAT	Secondary Active Transporter
S_{CD}	Deuterium Order Parameter
T_1	Spin-lattice relaxation time
τ_2	Second order correlation time
ω_c	Carbon Larmor frequency
XFF	X-ray Form Factor

Chapter 1 Introduction

1.1 Membrane

Membrane lipids function as boundaries to cells and internal organelles and regulate the molecular traffic across the boundaries [1]. Together with sterols, phospholipids are the main structural components of membranes [1], and their studies are necessary to understand the role and mechanisms on how membrane lipids regulate molecular transport as well as protein function [2]. When amphipathic phospholipids are mixed with water at the correct water to lipid ratio, bilayers spontaneously form, in which the nonpolar acyl chains remain in the center and the outer polar head groups interacting with the aqueous phase [1].

Cell membrane bilayers (without cholesterol) are usually in the liquid crystalline phase (L_α) [2]. As the temperature decreases below certain melting temperature, a phase transition from L_α to the gel phase (L_β) occurs [3]. Even though the physiological temperature of the human body is nearly constant, microorganisms live in varying temperatures, which influences cell membrane properties and structure. The variation of these properties is commonly related to the change in membrane fluidity. As the temperature increases, the membrane fluidity increases [4].

1.1.1 Single-component bilayer

Studying bilayer structural properties is important to understand roles and mechanisms of how membrane lipids regulate molecular transport as well as protein function [2]. Many properties of pure phospholipid bilayers with multiple chain types (Figure 1) mimic the cell membranes effectively even though they do not provide all the complexities [5], such as the sterols, glycolipids, and proteins which usually included in the biomembrane that affect the structural properties. Thus, this model is often used to study various biomembrane topics when the biological membrane is too complex for experimentation [5]. Experimentally, Kučerka et al. [6-10] measured single lipid membrane structural parameters, such as surface area per lipid (SA/lip), overall bilayer thickness, hydrophobic thickness, and headgroup-to-headgroup thickness, while Leftin and Brown [11] measured deuterium order parameters and the NMR spin-lattice relaxation time. Molecular dynamics (MD) simulations have also been extensively used to determine lipid membrane structural properties. With an accurate lipid force field, simulation results can assist the interpretation of x-ray scattering [12], NMR [13], and combined x-ray and neutron scattering [7, 9] results to characterize lipid bilayer structure and dynamics. Moreover, MD simulations are able to provide insight into bilayer structural properties that are difficult to determine experimentally.

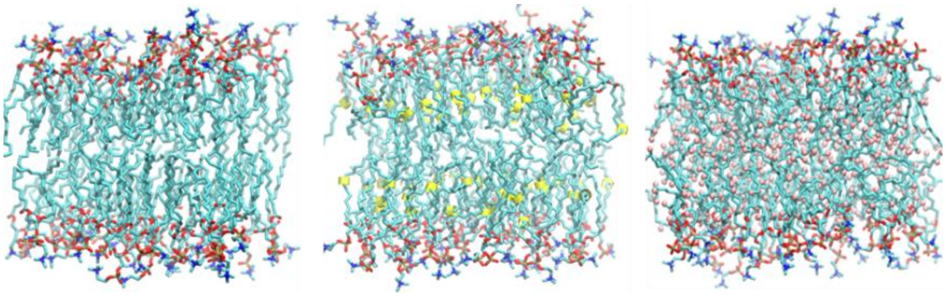

Figure 1 Various types of lipid bilayer system snapshots at the end of the simulations 1,2-dimyristoyl-sn-glycero-3-PC (DMPC) at 30 °C, 50 ns (left), 1-stearoyl-2-oleoyl-sn-glycero-3-PC (SOPC) at 50 °C, 100 ns (middle), and 1,2-diphytanoyl-sn-glycero-3-PC (DPhPC) at 30 °C, 160 ns (right). The traditional color scheme is applied: cyan for carbon, blue for nitrogen, red for oxygen, and orange for phosphate. Yellow represents the double bond in SOPC and pink is for branch and terminal methyl groups in DPhPC. The hydrogen atoms and water molecules are not shown for clarity.

There are various force fields for MD simulations of the phospholipids. The all-atom CHARMM36 force field (C36FF-AA) developed by Klauda et al. [14, 15] allows for constant particle number, pressure, and temperature (*NPT*) simulations of pure lipid bilayers, accurately yielding various structural and dynamic properties of lipid membranes. MD simulations with the C36FF-AA have been used to study the lipid bilayers with cholesterol [16, 17], glycolipids [18], membrane-associated proteins [19-21], micelles [21-23], and Gram-negative bacterial outer [24, 25] and inner [26, 27] membranes. Previously, it has been reported that the C36FF-AA is an accurate force field for PC bilayer simulation at varied temperatures [28]. We propose that C36FF-AA may be capable of providing better structural properties than other force fields regardless of lipid types; e.g., while the experimental SA/lip for POPE is 58.0 $Å^2$ at 35 °C [7], the AMBER Lipid14 predicts 55.5±0.2 $Å^2$ at 37 °C [29], SLIPIDS gives 56.3±0.4 $Å^2$ at 37 °C [30], and C36FF-AA yields 59.2±0.3 $Å^2$ at 37 °C [14] that is closest to the experimental data. Moreover, C36FF-AA [14] results in more accurate deuterium order parameter than Berger force field [31] for DPPC and POPC [32]. Recently, the properties of DMPC, POPC, and POPE obtained from C36FF-AA [14], Lipid14 [29], Slipids [33], and GROMOS54a7 [34] were compared in [35], showing that C36FF yields more accurate lipid volume, bilayer thickness, and lipid diffusion coefficient than the other force fields.

Therefore, MD simulations were performed to test accuracy of all-atom CHARMM36 force field for a variety of lipid types of phospholipid (different head groups and tail types for saturated, mono-, mixed-, and poly-unsaturated lipids) over a wider temperature range (30-80 °C for saturated and 30-40 °C for unsaturated lipids) by comparing the simulation structural properties with and validate the experimental structural properties data. In total, 76 lipid types were simulated. Moreover, the large data set from our simulations allows investigating certain trends in specific properties as a function of temperature, chain length, and unsaturation level. The in-depth analysis of lipid membrane simulations at the atomic level allow for a complete understanding of lipid hydrogen bonding, cluster formation, and internal lipid motions.

Table 1 The simulated lipid's tail information
List of the tail name abbreviation, chain length, saturation level, and double bond position of the first atom. The double bonds in all the unsaturated lipids have cis geometry.

Saturated				
DL	DM	DP	DS	
12:0	14:0	16:0	18:0	
Mono-unsaturated				
DX	DO	DG	DE	DN
16:1 (Δ9)	18:1 (Δ9)	20:1 (Δ11)	22:1 (Δ13)	24:1 (Δ15)
Mixed mono/di-unsaturated (*sn*-1/*sn*-2)				
PO	SO	PL	SL	
16:0/18:1 (Δ9)	18:0/18:1 (Δ9)	16:0/18:2 (Δ9,12)	18:0/18:2 (Δ9,12)	
Poly-unsaturated (*sn*-1/*sn*-2)				
SA	SD	DA		
18:0/20:4 (Δ5,8,11,14)	18:0/22:6 (Δ4,7,10,13,16,19)	20:4/20:4 (Δ5,8,11,14)		

The head group types considered in this study are phosphatidic acid (PA), phosphocholine (PC), phosphoethanolamine (PE), phosphoglycerol (PG), and phosphoserine (PS). As listed in Table 1, the lipid tail types vary with saturated chain lengths ranging from 12 to 18, which are dilauroyl (DL, 12:0), dimyristoyl (DM, 14:0), dipalmitoyl (DP, 16:0), and distearoyl (DS, 18:0). Mono-unsaturated lipids vary with chain length ranging from 16 to 24 which are dipalmitoleoyl (DX), dioleoyl (DO), digadoleoyl (DG), dierucoyl (DE), and dinervonoyl (DN). Mixed mono/di-unsaturated lipids include 1-palmitoyl-2-oleoyl (PO), 1-stearoyl-2-oleoyl (SO), 1-palmitoyl-2-linoleoyl (PL), and 1-stearoyl-2-linoleoyl (SL). Poly-unsaturated lipids include 1-stearoyl-2-arachidonoyl (SA), 1-stearoyl-2-docosahexaenoyl (SD), and diarachidonoyl (DA). The full name of a lipid can be obtained by combining the tail and head group names of the lipid in the following format; e.g., DXPE is 1, 2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine. Table 1 summarizes all the acyl chain types and a few representative lipid structures are shown in Figure 2. The list of all the simulation systems that have been performed are shown in Table A.1, and the detail results will be provided in Chapter 4.

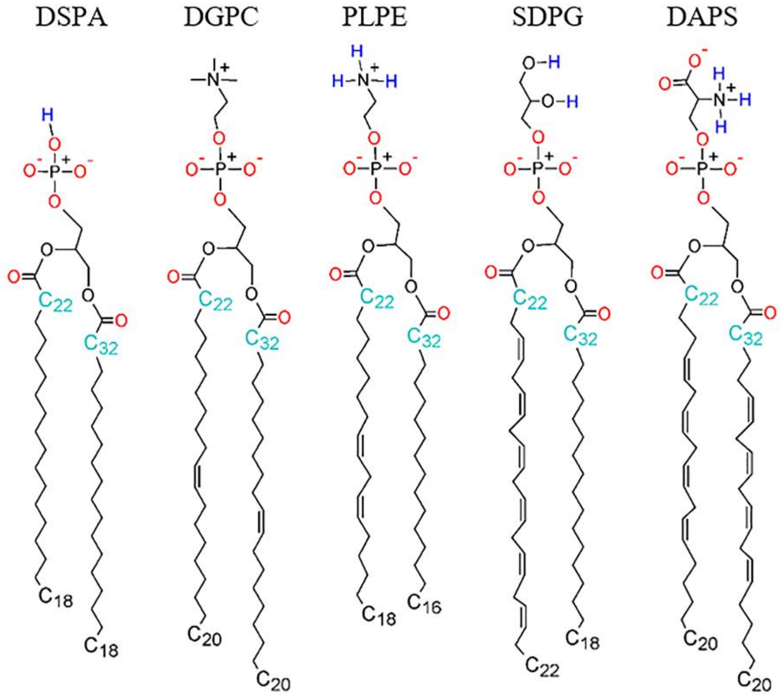

Figure 2 Chemical structures of a few representative lipids types.
The blue hydrogen and red oxygen atoms indicate the hydrogen bonding proton and acceptors used in the hydrogen bond analysis, and the carbon atoms used for reorientational correlation analysis are shown in cyan.

The total of 72 lipid molecules in each bilayer (36 lipids per leaflet) is applied in all the single-component lipid bilayer system (except for the pure 18:2 linoleic lipid bilayer which will be introduced in the later section). In order to check the system size effects on the analyzed properties, we performed addition simulations of DOPA, DOPG, and DOPS with 128 lipid molecules. As shown in Table A.2, there are no or minor effects on almost all structural properties. We did not calculate lipid diffusion constants in this study due to well-known finite-size effect carefully studied in [36].

1.1.2 Multiple-component bilayer

1.1.2a E. coli Membrane

The biological membranes are more complex than a simplified homogenous lipid bilayer. Specifically, a realistic membrane contains a variety of lipid components, such as glycerol lipids, sterols, and transmembrane proteins, etc. An accurate representation of the membrane is essential when the lipid-protein interactions need to be taken into account, for instance, the *E. coli* membrane which was used in our studies of a transmembrane protein.

Table 2 The top6 membrane composition
The lipid number in the top/bottom leaflet are shown as ideal (initially built), but the lipids that overlap with proteins are removed. The modified columns show the exact lipids numbers applied in the explicit membrane MD simulations for all the protonation states. 1-palmitoyl-2-cis-9,10-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine (PMPE), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 1-pentadecanoyl-2-cis-9,10-methylene-hexadecanoic-acid-sn-glycero-3-phosphoethanolamine (QMPE), 1-palmitoyl- 2-cis-9,10-methylene-hexadecanoic-acid-sn-glycero-3-phosphatidylglycerol (PMPG), 1-palmitoyl-2-palmitoleoyl-snglycero-3-phosphoglycerol (PSPG), and 1-oleoyl-2-palmitoleoyl-snglycero-3-phosphoethanolamine (OSPE). Please refer to reference [27] for the chemical structure of each lipid.

Lipids	*sn*-2	*sn*-1	Ideal top/bot.	Modified top/bot.
PMPE	cy17:0	16:0	74/74	72/74
POPE	18:1	16:0	20/20	19/20
QMPE	cy17:0	15:0	20/20	19/19
PMPG	cy17:0	15:0	16/16	16/15
PSPG	16:1	16:0	14/14	14/14
OSPE	16:1	16:0	12/12	11/12

A multiple-component lipid membrane model (top6) was developed previously to match the composition of the inner membrane of *E. coli.* This is based on the composition published by Pandit and Klauda [27] in which the ratio of PE:PG equals 4.2:1 and the membrane contains lipids with a cyclopropane moiety (cyC17:0). The lipids used are listed in Table 2 and it is known that *E. coli* cannot produce sterol, so these membranes lack these sterol lipids. This membrane model was applied to study the lactose permease of *E. coli*, LacY which will be introduced in the later section.

1.1.2b Soybean Membrane

In comparison, considerably less work has been done on the plasma composition of higher plants. In the past, studies have been hindered because highly enriched fractions of plasma membranes could not be produced sufficiently for analysis. However, recent studies have provided more concrete characterizations through different isolation techniques, resulting in plasma membranes fractions that may exceed 80% purity [37]. Additionally, lipid analysis techniques have continued to improve over time. Mass spectrometry, for instance, has been largely successful in recent years due to its high sensitivity and high specificity [38].

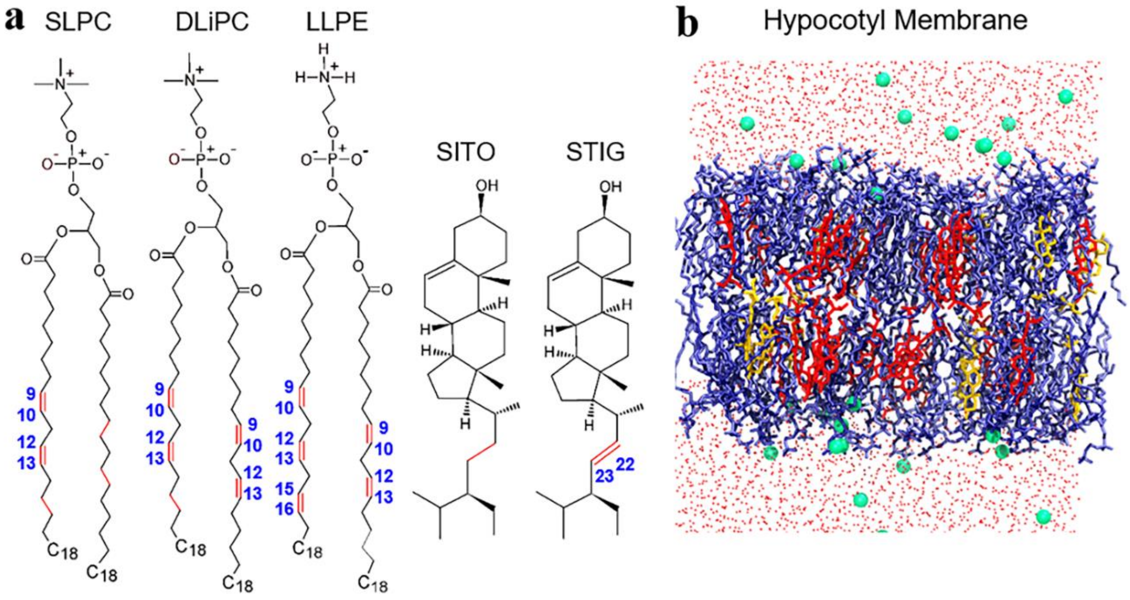

Figure 3 Chemical structures of linoleoyl lipids and snapshot of soybean membrane
a) The lipids are SLPC, DLiPC, LLPE, and sterol lipids of β-sitosterol (SITO), stigmasterol (STIG). The positions of double bond carbons are shown in blue texts. b) The snapshot of hypocotyl at the end of the simulation. The glycerol phospholipids are shown in blue lines; the two types of sterols SITO and STIG are in red and yellow lines respectively, the water is shown in red dot and the potassium ions are shown in green sphere.

Plant membranes have also been modeled but less frequently than membranes of other organisms. Previously, MD simulation techniques were used to investigate the interaction between isoprene and a phospholipid membrane [39]. The results of the study found that plants release isoprene to stabilize the membrane and avoid undergoing a heat-induced phase transition [39]. Soy membranes, however, have not been extensively studied through computational models.

Table 3 Lipid composition of soybean hypocotyl membrane

Lipid	*sn*-1/*sn*-2	Exp[37, 40, 41]	# lipid/leaflet
PLPC	16:0 – 18:2	11.5%	12
DLiPC	18:2 - 18:2	22.1%	22
PLPE	16:0 - 18:2	13.3%	13
DLiPE	18:2 - 18:2	7.3%	7
LLPE	18:2 - 18:3	5.1%	5
PLPI	16:0 - 18:2	8.5%	9
L2PI	18:2 - 18:2	3.6%	4
Sitosterol	N/A	21.0%	21
Stigmasterol	N/A	7.5%	7

Table 4 Lipid composition of soybean root membrane

Lipid	*sn*-1/*sn*-2	Exp[37, 40, 41]	# lipid/leaflet
PLPA	16:0 - 18:2	10.0%	10
LLPA	18:2 - 18:3	10.0%	10
PLPC	16:0 - 18:2	11.0%	11
LLPC	18:2 - 18:3	11.0%	11
PLPE	16:0 - 18:2	10.2%	10
LLPE	18:2 - 18:3	10.2%	10
LLPS	18:2 - 18:3	4.3%	4
PLPI	16:0 - 18:2	4.7%	5
Sitosterol	N/A	18.6%	19
Stigmasterol	N/A	10.0%	10

The compositions of soybean plasma membranes vary depending on the species, stage of development, and the part of the plant. The two parts of the plant that were examined in this study were the hypocotyl and the root of a developing seedling. The hypocotyl is the stem of the germinating seedling, found below the seed leaves and the embryonic root. The root membrane under study was that of the seedling soybean root. The compositions of these membranes were weighted and averaged from past experimental studies [37, 40, 41] and modeled using lipid bilayers consisting of 100 lipids on the top and bottom leaflets.

The purpose of the soybean membrane study is to model and characterize the properties of the plasma membranes of these two regions of the soy plant through MD simulations for use in future studies. The pure 1-stearoyl-2-linoleoyl-sn-glycero-3-phosphocholine (SLPC, 18:0/18:2) and dilinoleylphosphatidylcholine (DLiPC, di-18:2) bilayer systems are studied (Figure 3a). The hypocotyl and root membranes each contains two common types of sterols (sitosterol and stigmasterol) and 7 or 8 eight types of linoleoyl glycerol phosphate lipids, with the head groups such as PA, PC, PE, PI, and PS (for root membrane only). The ratio of the sterols and lipids of each head group are very similar in the hypocotyl and root membrane models. The composition for hypocotyl membrane is shown in Table 3 and root membrane in Table 4. The built hypocotyl membrane based on the given composition is shown in Figure 3b, and the detail results will be provided in Chapter 5.

1.2 Transmembrane Protein-Lactose Permease of *E. coli* (LacY)

1.2.1 Function of LacY

Normal functioning of cells requires energy, and disaccharides (e.g. lactose) are one example energy source. In *E. coli*, lactose is transported into the cell via a transmembrane protein (e.g. lactose permease, LacY) that catalyzes the coupled translocation of β-galactosides and a proton [42]. LacY is classified as a secondary active transporter (SAT) that utilizes the free energy from the transporting of proton along the gradient to drive the accumulation of β-galactosides against their concentration gradient [43] via a symport mechanism (proton and sugar transport in the same direction). It belongs to major facilitator superfamily (MFS) which is the largest and most diverse transporters that transport many types of substrates, such as ions, peptides and disaccharides [44] as in LacY. Moreover, MFS can be found in membranes of many living organisms [45, 46].

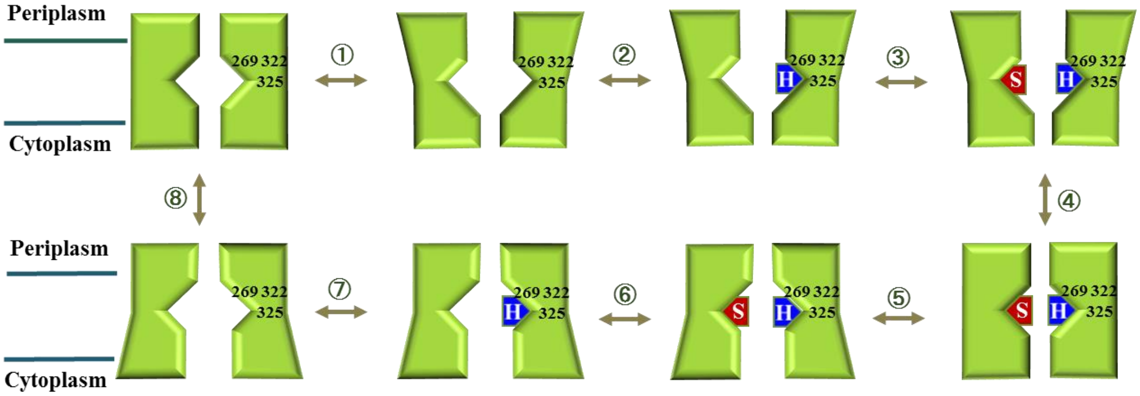

Figure 4 A schematic model showing the sugar/proton symport mechanism [47-49]. In the plot, H (blue) represents proton, S (red) represents galactoside sugar. The residues (Glu269, His322, and Glu325) involved in proton transport are shown.

Based on the complete sugar/proton transport cycle in LacY proposed by Guan and Kaback [47], a similar schematic model of process with the known occluded transition states added by Madej et al. [49] is shown in Figure 4, which involved eight transition steps: 1) Sugar induces LacY to change from *apo*-occluded to outward-facing state; (2) Proton enters LacY from periplasmic side and protonates Glu269; (3) Sugar enters LacY from periplasmic side; (4) The outward-facing LacY become occluded during which the proton is transferred to Glu325 through His322 (5) The occluded LacY change to the inward-facing state; (6) Sugar is released on the cytoplasmic side; (7) Proton is released on the cytoplasmic side; (8) LacY transits back to the *apo*-occluded state [47-49].

1.2.2 Structures of LacY

LacY consists of 417 residues, in which twelve transmembrane helices are connected by hydrophilic loops, short N-terminus and C-terminal tail [50]. It is pseudo-symmetric with six helices in N-domain and C-domain respectively, and interior hydrophilic cavity [51], like any other MFS members [44]. The earlier obtained LacY crystal structures are in the cytoplasmic-open (inward-facing), such as 2V8N [52] shown in Figure 5, or 1PV7 [53], in which the cytoplasmic side is open, and the periplasmic side of LacY is tightly packed. The occluded-like structure has also been determined (4OAA [54] shown in Figure 5), in which both cytoplasmic and periplasmic sides are closed. The recently nanobody-binded periplamic-partially-open LacY 5GXB [55] was also obtained.

The crystal structure of periplasmic-open (outward-facing) state LacY is not available. The residues involved in proton translocation are near the center of the LacY hydrophilic cavity [56]. According to the alternating access mechanism [57], the periplasmic side of LacY should be able to open and close during the translocation of sugar across the membrane [58-60].

A few experimental techniques, such as single-molecule fluorescence (Förster) resonance energy transfer (sm-FRET) [61] and double electron-electron resonance (DEER) [62], have been used to understand structural changes in LacY. The sm-FRET results show that upon binding of a galactopyranoside, distance of labeled residues decrease on the cytoplasmic side and increase on the periplasmic side; the opening/closing of the cytoplasmic and periplasmic side of LacY may not be coordinated, which suggests two domains of LacY do not behave as rigid bodies [61]. DEER has been used to measures the distances between the pairwise couplings electron spins of the spin labels. It is also observed that galactopyranoside induce a closing on the cytoplasmic side and opening on the periplasmic side [62].

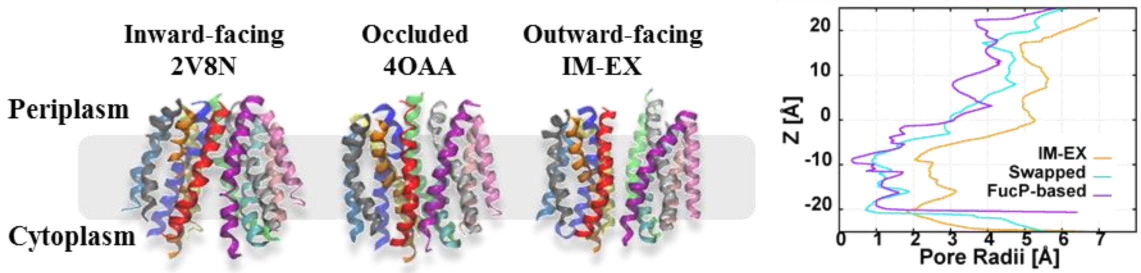

Figure 5 Side view snapshots and pore radii of the LacY in different conformational states. The LacY structure of cytoplasmic-open 2V8N LacY crystal, the occluded 4OAA LacY crystal, and the periplasmic-open LacY$^{IM\text{-}EX}$ model (left panel). Pore radii of three periplasmic-open (outward-facing) LacY models (right panel).

The exact structure of periplasmic-open LacY is important for understanding the complete proton/sugar transport cycle of LacY and also the mechanism of other substrate transporters. However, LacY structures are only known for the cytoplasmic-open [52, 53] and occluded-like [54] state. Molecular simulation methods have been applied to determine the periplasmic-open LacY$^{IM\text{-}EX}$ (Figure 5) structure by Pendse et al. [48]. Most of the residue pair (C_α-C_α) distance change of the outward-facing LacY$^{IM\text{-}EX}$ model has the reasonable agreement with DEER. The major disagreement is that the cytoplasmic side does not close as much as DEER data suggests. In this study, the MD simulations are performed to further test the accuracy of LacY$^{IM\text{-}EX}$ model, as well as two other available periplasmic-open LacY models, LacY^{SW} and LacY^{FP} models obtained by Radestock et al. [51], the detail results of which will be provided in Chapter 6. The pore radii of three models are shown in Figure 5 (right panel). Another key disagreement between LacY$^{IM\text{-}EX}$ model and DEER are that distance change between the spin-labeled neighboring residues are very similar in simulated C_α-C_α distance [48], but they are significantly different in DEER experimental data [62], which motivate us to perform MDDS simulations to test the effect of spin labels (size and internal flexibility) on the LacY residue distance measurement, the detail results of which will be provided in Chapter 7.

Chapter 2 Computational Methods

2.1 Principles of Molecular Dynamics Simulation

The computer simulation has advantages of low cost, high efficiency, accessibility-of details in the atomic scale. For rare cases when the light particles (He, H_2, and D_2) have very high translational and rotational motion frequency ($h\nu > K_B T$), the quantum effect needs to considered, while most materials can be assumed to obey law the classical (Newtonian) mechanics with great accurary [63]. Therefore, MD simulation allows simulating the physical movements of atoms (position, and velocity) by numerically solving Newton's equations of motion using an interatomic potential (i.e. force field).

Besides experiments, MD simulations have also been extensively used to determine lipid membrane equilibrium structural properties and also transport properties (diffusivity, viscosity, and thermal conductivity, etc.). Only the equilibrium properties are covered in this work, while the transport properties are not calculated due to lack of experimental data and also slightly less relevant to the membrane study. With an accurate lipid force field, simulation results can assist the interpretation of x-ray scattering [12], NMR [13], and combined x-ray and neutron scattering [7, 9] results to characterize lipid bilayer structure and dynamics. Moreover, MD simulations are able to provide insight into bilayer structural properties that are difficult to determine experimentally.

For complex biological molecules, molecular simulation programs are typically used to perform MD simulations. Some example programs are Assisted Model Building with Energy Refinement (AMBER) [64], Chemistry at HARvard Macromolecular Mechanics (CHARMM) [65], GROMACS [66], Large-Scale Atomic/Molecular Massively

Parallel Simulator (LAMMPS) [67], and the Nanoscale Molecular Dynamics (NAMD) [68], among which CHARMM [65] and NAMD [68] are the two programs applied in my study.

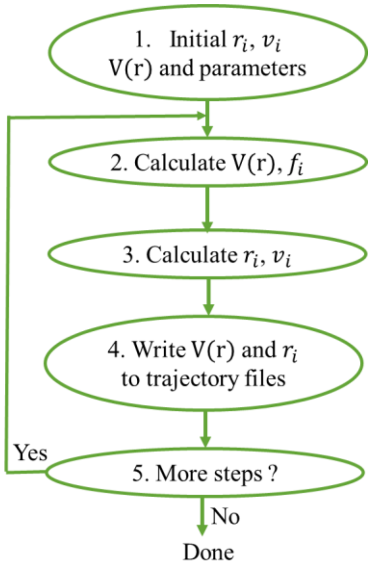

Figure 6 Simulation flowchart [63, 69]

The common process steps in the MD simulation program are shown in Figure 6, which in words are as follows [63, 69]: 1. Initialize the system by providing the initial positions (coordinates) and velocities of all the particles in the system, and also reading the initial state parameters, such as number of particles, number of time steps, and initial temperature, etc., and also the desired environmental parameters, such as desired temperature and pressure, etc. 2. Calculate interatomic interaction potential energy and then the force based on the selected force field. 3. Numerically integrate Newton's equation of motion based on the specified algorithm. 4. Write the position and velocity data of all the particles in this time step to the trajectory file 5. If the number of the time step is less than the specified desired simulation time, repeat Step 2-4. Otherwise, break the loop, and the simulation process is complete. The obtained trajectories will then be analyzed to obtain the averaged system properties. To better understand the MD simulation, more details on steps 1-3 will be discussed in each of the following subsections.

2.1.1 Initialization

The term initialization mainly refers to specify the initial positions and velocities of particles in the system. Even though the initial state should not affect the properties at the equilibrium state, to avoid unnecessary longer simulation time, we still want to select the initial positions of particles to be compatible with the overall structure of the system that we want to build, and are confined by the shape and size of box based on the crystal structure of the simulated objects, with the requirement that the atomic core of the particles should not overlap [63]. Based on initial temperature and density, a uniform distribution of velocity values within the certain interval is generated. Then each velocity component (x, y, and z) of each particle is randomly drawn from the velocity uniform distribution. Therefore, generally, the velocities of particles are randomly distributed initially, and they would form a Maxwell-Boltzmann distribution when the equilibrium is reached [63]. The desired number of steps (simulation time) is based on the time the system reaches equilibrium since we aim to calculate the equilibrium properties.

2.1.2 Calculate Force based on Force Field

Before getting into the calculation, it is necessary to introduce the term-force field. A force field contains two components: function to calculate the energy from position coordinates and parameters in the function. A key functioning step of simulation is to calculate interaction function and force based on the varied force field. There are many types of force fields available. The selection of force field depends on the type of simulated system and the interested parameters that we want to calculate.

All force fields potential functions contain two classes of interactions. First class is the bonded interactions which include covalent bond-stretching, angle-bending, dihedral or torsion potentials when rotating around bonds, and out-of-plane “improper torsion”. And the second class of the non-bonded interactions which include Lennard-Jones repulsion and dispersion and Coulomb electrostatics. For our all-atom molecular dynamics simulations, as we aim to test the accuracy of all-atom CHARMM36 force field (C36FF-AA), we have been applying C36FF-AA with the potential function is shown below, which has the additional cross Urey-Bradley term representing the in-plane deformations and asymmetric bond stretching between atoms 1 and 3 in some cases [70].

$$V(r_1,\ldots r_N)=\sum_{bonds}K_b(b-b_0)^2+\sum_{angles}K_\theta(\theta-\theta_0)^2+\sum_{cross\ \mathrm{UB}}K_{UB}\left(r_{1,3}-r_{1,3}^0\right)^2+\sum_{improper}K_{im}(1-\cos(2\phi)) \quad (1)$$

$$+\sum_{dihedrals}\left[\sum_{j}K_{\varphi,j}\left(1+\cos(n_j\varphi-\delta_j)\right)\right]+\sum_{\substack{nonbonded\\ pairs\,i,j}}\varepsilon_{ij}\left[\left(\frac{r_{\min,ij}}{r_{ij}}\right)^{12}-\left(\frac{r_{\min,ij}}{r_{ij}}\right)^{6}\right]+\sum_{\substack{nonbonded\\ pairs\,i,j}}\frac{q_iq_j}{\varepsilon_D r_{ij}}$$

Then based on calculated potential energy, the force for particle *i* is calculated:

$$f_i=-\frac{\partial}{\partial r_i}V(r_1,\ldots r_N) \quad (2)$$

Calculation of interaction force of among all particles based on applied force field is usually most time-consuming step of the simulation process. For the pairwise addition interactions, i.e. considering the force on particle i due to all other (nearest imaged neighbor) particles (which is done in CHARMM force field). Originally, the computing scale time would be O(N^2) to evaluate N(N-1)/2 pairs of interactions. However, some techniques have been developed to speed up the force calculation. For instance, with the switching function [71] with reasonable cutoff distance applied, the time scale can be reduced to O(N).

2.1.3 Integrate Equation of Motion

As in almost all systems, it is valid to approximate the movement of the particles in the system using Newton's Equation of motion: $f_i(t) = m_i \frac{\partial^2 r_i}{\partial t^2}$. Taylor expansion of the position function of a particle $r(t)$ around time t related to velocity $v(t)$ and time step Δt, as shown in equation (3), is used to derive the varied algorithms to numerically integrate equation of Newtonian motion, including low-order (Verlet-style) algorithms and high-order algorithms.

$$r(t + \Delta t) = r(t) + v(t)\Delta t + \frac{f(t)}{2m}\Delta t^2 + \cdots \tag{3}$$

The energy conservation is an important factor for a good algorithm. There are a short time (i.e. constant for a few time steps) and a long time (i.e. constant for many time steps) energy conservations.

The low-order algorithms tend to have moderate short-time but good long-time energy conservation, while the high-order algorithms tend to have good short-time but inaccurate long-time energy conservation. There still is no algorithm available that can meet both short and long time energy conservation, which means that there is no algorithm that can predict both the short and long time trajectory of the particles. Therefore, from MD simulations, instead of precise predictions based on the initial condition, we obtained the statistical predictions which generally represent the true system very well [63]. The low-order Verlet-style integration algorithms are applied for my simulations. Velocity Verlet [72] is applied in the NAMD [68], while CHARMM [65] provides Velocity Verlet [72] and Leapfrog Verlet [73] and a few other algorithms. Leapfrog Verlet [73] is used in all the CHARMM scripts involved in my work.

2.1.4 Periodic Boundary Condition

The periodic boundary condition (PBC) is applied to avoid surface artifacts, so a molecule that exits the primary cell to the right reappears on the left. PBC is important because single simulation box is generally small to obtain good statistical data, while with PBC now we can assume that the periodic system has an infinitely large system size. All systems are large enough to avoid any finite size issues that have been noted previously [36].

2.1.5 Interaction Between Molecules

Since molecules interact with each other, the nonbonded interactions are computationally demanding, if we consider over all neighbors in the infinite periodic system. To increase efficiency, for van der Waals interactions, instead of sudden cut-off which brings more error, a switch function (in our simulations, a switch distance of 10 Å) is applied so that the interaction is switched off 2 Å before the cut-off (12 Å). For the electrostatic interactions, the particle mesh Ewald summation (PME) is applied which splits the summation into short- and long-range interactions separated by cut-off. The short-range part is directly calculated while the long-range part is treated by assigning charges to a grid that is solved in reciprocal space through Fourier transforms [69].

2.1.6 Temperature and Pressure Control

The cut-offs and rounding errors can cause drifts in the energy. A thermostat is applied to maintain the desired set temperature by adding heat through various algorithms during the integration. In our MD simulations, Langevin dynamics was applied to non-hydrogen atoms to maintain the constant temperature. Similarly, the system can be maintained in the set total pressure through scaling the simulation box size in x, y, and z dimensions. Nosé-Hoover Langevin-piston algorithm [74, 75] was used to maintain a constant pressure in our MD simulations.

2.1.7 Time Step

Millions of pairs of non-bonded interactions have to be calculated at every time step, therefore it is the most cost computation over the simulation. Larger time step will increase the computation efficiency, however, it generally introduces more error, e.g., 1-fs time step introduces error in bond vibrations (because bond length vibration is about the time scale of 1 fs [69]). In our simulations, we are not interested in the detail about the bond vibration for hydrogens, so we apply SHAKE algorithm [76] to constraint the hydrogen-involved bond length, which allows us to use 2-fs as the time steps.

2.2 Molecular Dynamics Simulation Setup Protocols

2.2.1 Common MD Simulation Setup

CHARMM-GUI *Membrane Builder* [77-81] was used to build either homogenous or heterogamous membrane with or without protein. These simulations used the standard TIP3P water model [82, 83] with the C36FF [14, 15]. The van der Waals interactions were smoothly switched off between 10 and 12 Å by a forced-based switching function [71]. All the bond lengths involving hydrogen atoms were constrained using the SHAKE algorithm [76]. Particle mesh Ewald [84] was used for electrostatic interactions with an interpolation order of 6 and a direct space tolerance of 10^{-6}. NAMD [68] was used to perform the MD simulations at varied temperatures. The simulation temperatures were above the gel-transition temperatures of individual lipids. After standard CHARMM-GUI six minimization and equilibration steps (0.685 ns) [81], the simulations were run for 100 or 150 ns (depends on the time the system reach equilibrium) with a time-step of 2 fs, and the data was collected every 4 ps. All simulations were run in the *NPT* ensemble with a pressure set to 1 bar. Langevin dynamics was applied to non-hydrogen atoms to maintain the constant temperature with a Langevin coupling coefficient of 1 ps^{-1}, while Nosé-Hoover Langevin-piston algorithm [74, 75] was used to maintain constant pressure with a piston period of 50 fs and a piston decay of 25 fs.

2.2.2 Membrane-only System

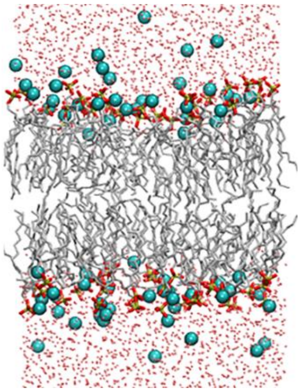

Figure 7 An example built DLPA bilayer system together with water and ion molecules. The red dots represent water, and the cyan spheres are K$^+$ ions. The orange surrounded by red dots represent the phosphate groups located in the water and bilayer interface, and the silver lines represent the hydrophobic acyl chains of DLPA.

For single-component lipid bilayer, each bilayer system was built in a tetragonal box containing a total of 72 lipid molecules (36 per leaflet) and about 45 water molecules per lipid to make sure that lipids were fully hydrated. K$^+$ ions were added to neutralize anionic PA, PG, and PS bilayer systems. The example of the built system of DLPA bilayer is shown in Figure 7. For the united-atom simulations, CHARMM-GUI *Membrane Builder* [77-81] is used to build the all-atom lipid bilayer system first, and then the tails are modified to generate the united-atom bilayer system. This method is also applied to build the soybean membrane model that contains new lipid structures which are not currently available in the CHARMM-GUI *Membrane Builder*.

2.2.3 LacY-Spin-label System

DEER studies on proteins mostly use the 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl methanethiosulfonate spin label (MTSSL, Figure 8a), which can selectively attach to the sulfhydryl group of either natural or engineering mutated cysteine residues and release the sulfuric acid [85], during which the disulfide bond is formed between MTSSL and cysteine. The DEER experiment provides distance distribution instead of a single value of the distance for each residue pair due to the flexibility of structures [86, 87]. Molecular dynamics dummy spin-labels (MDDS) simulations [88] are performed to predict the spin label distance distribution in order to verify the effect of size and orientation of spin labels on the accuracy of DEER measurement. MDDS simulation has been used to study some SAT proteins, such as conformational change of LeuT [88], and the conformational change and ion-coupling of another Mhp1 [89]. MD simulations with the all-atom MTSSL are computationally expensive. Moreover, multiple pairs of spin labels in the vicinity cannot be simulated simultaneously, because the MTSSL will interact with each other. Therefore, instead of using a full spin label MTSSL (Figure 8a), dummy OND spin labels (DSL) as shown in Figure 8b are applied, in which nitroxide oxygen ON(D) atom is used to replace the MTSSL side chain, as DSLs are parametrized as a neutron atom, not interacting with each other, and having very weak interaction with surrounding protein residues [88, 90]. Besides, the targeted residues are mutated to ALA to avoid steric clashes with ON(D) spin labels [88]. The dummy spin label with side chain (Ala-O) is well parametrized so that it behaves same as full MTSSL [88].

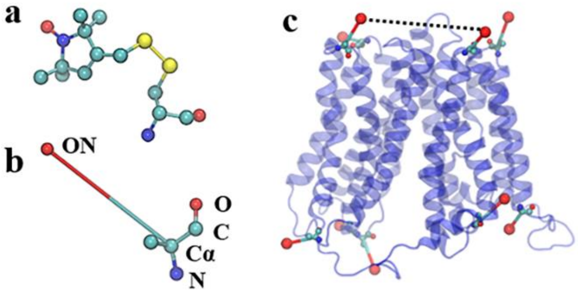

Figure 8 The snapshots of the two types of spin labels and MDDS simulation setup: a) MTSSL spin-labeled residue used in DEER b) Dummy spin-labeled residue with Ala-O side chain used in the MDDS simulation c) A side view snapshot of the built inward-facing LacY with nine dummy spin labels for the MDDS simulation.

The structures used for the MDDS simulation are selected based on equilibrated pore radii. The last frame structure at each 2 ns (10-20 ns) of the inward-facing and occluded structures are sampled from implicit Self-guided Langevin dynamics (SGLD) simulations with 2V8N as initial structure, and each 10 ns (10-100 ns) of outward-facing from explicit MD simulation with E269-protonated LacY$^{IM\text{-}EX}$ as initial structure. Total six inward-facing structures, six occluded structures, and ten outward-facing structures are sampled for the MDDS simulations. To direct compare the structure and spin label distances, the MDDS simulations with crystal structures (2V8N [52], 1PV7 [53] and 4OAA [54]) are also performed. In each LacY structure, the nine target residues are mutated to ALA (R73, V105, S136, N137, I164, T310, Q340, S375, and S401), and the dummy spin labels are added to these residues to match with the experimentally-measured spin labels. The MDDS simulations of LacY in a vacuum are performed using Langevin dynamics at 37 °C for 10 ns with a dielectric constant of 80. The time step is 0.5 fs, and the data is collected every 250 fs. The friction constants are set to 10.0 ps^{-1} for spin labels (ON), and 5.0 ps^{-1} for all the rest atoms. A constraint force constant of 1 kcal·mol^{-1}·Å^{-2} is applied to the backbone of the protein throughout the MDDS simulations. An example of built structure of inward-facing LacY with dummy spin labels is shown in Figure 8c. For each run, the spin label distance distribution is the histogram plots of the spin label pair distances obtained from 20,000 frames. The mean distances are obtained by fitting the averaged distance distribution data of each conformational state to Gaussians.

2.2.4 Membrane-LacY System

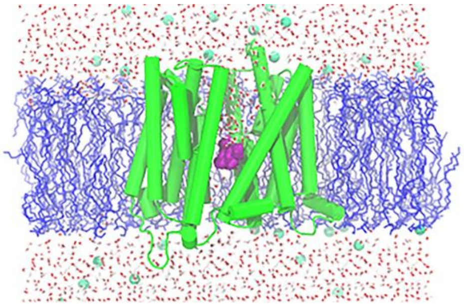

Figure 9 A side view snapshot of built LacY$^{IM\text{-}EX}$ in the explicit membrane and ββ-(Galp)$_2$. The sugar ββ-(Galp)$_2$ is shown in purple and K^+ is shown in the cyan sphere.

With the available PDB of the outward-facing LacY structural models, the CHARMM-GUI *Membrane Builder* [77-79, 81] is used to build a protein, membrane, water, and neutralizing potassium ions heterogeneous systems in a tetragonal box. Then ββ-(Gal*p*)$_2$ is appended to the system. Since the binding of β-galactosides induces LacY to open on the periplasmic side, [62] we believe that the LacY tends to remain in the outward-facing state when it binds to β-galactosides. An example of the final built system is shown in Figure 9. The lipid membrane is modeled to match the composition of the inner membrane of *E. coli*. This is based on the composition published by Pandit and Klauda [27] in which ratio of PE:PG equals 4.2:1 and the membrane contains lipids with a cycle-propane moiety (cyC17:0). A total of 312 lipid molecules 156 per leaflet, and two water layer of thickness 15 Å are built. Additional pore water is placed to avoid protein clash toward the empty pore center, but it cannot overlap with ββ-(Gal*p*)$_2$. The LacY models with either Glu269, His322 and Glu325 (Figure 9b) protonated, which are three essential residues for proton translocation [53]. The NAMD program [68] is used to equilibrate the lipid bilayer systems using two types of equilibrations. The standard *Membrane Builder* six-step equilibration process [77-79, 81] gradually turnoff both backbone and side chain restraints over 685 ps at a temperature of 37 °C. When the LacY is unstable, an extra 10-ns restrained equilibration may be applied, in which the side chain restraints gradually turnoff over 685 ps, while the constant backbone restraints of 10 $\text{kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$ is applied and continues for additional 5 ns, finally the backbone restraints gradually turnoff over 5 ns. After the completion of equilibrations, the MD simulation runs are performed for 100 or 150 ns.

2.3 Self-Guided Langevin Dynamics Simulation

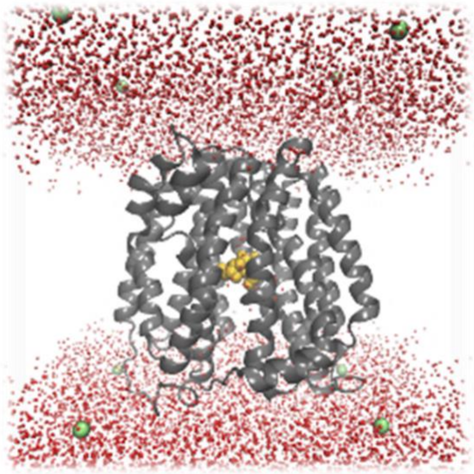

Figure 10 Snapshot of SGLD simulation in the implicit membrane setup for LacY. The following color schemes are applied: LacY is shown in grey, sugar ββ-$(\text{Gal}p)_2$ in yellow, water molecules in red, chloride ions in green.

The SGLD simulation can accelerate the conformational change while preserving the conformational distribution well when the guiding parameters are within the suggested range (guiding factor less than 1 is recommended for avoiding deviation from the canonical ensemble) [91]. Therefore, it is applied to enhance search efficiency of LacY in different conformational states without the loss of the conformational accuracy [73, 91, 92].

The wild-type inward-facing LacY crystal structure 2V8N [52] from RCSB protein data bank is used in the SGLD simulation in the implicit membrane. The CHARMM-GUI *Membrane Builder* [77-79, 81] is used to build the membrane, water, and neutralizing ions heterogeneous systems in a tetragonal box. To generate implicit membrane environment, the membrane is removed from the system, and a planar (with normal in the z direction) exponential potential restraint is applied to LacY, water and ions in order to maintain the hydrophobic interaction. The number of ions is adjusted to keep system neutral. To increase the simulation efficiency, the XY dimensions of water layers are reduced to A=B=90 Å. Then the disaccharide β-D-Gal*p*-(1-1)-β-D-Gal*p* is appended to the system. The pore water that overlap with ββ-(Gal*p*)$_2$ is removed. The final built system is shown in Figure 10, which is similar to Figure 9 but without the membrane bilayer and slightly smaller X/Y dimensions. The CHARMM program [65] is used with the backbone of the protein fixed, the system is minimized for 1000 steps and equilibrated for 1ns with Langevin dynamics and Hoover Langevin-piston algorithm [74, 75]. Then, the SGLD simulations with leap-frog Verlet algorithm [73] are performed for 20 ns at 37 °C with no restraints. The time step is 2 fs, and the data is collected every ps. The varied guiding factor (SGFT) between 0.5 and 1.25, and the varied friction constant (FBETA, collisions per picosecond) between 0.5 and 25 ps^{-1} are applied to test their ability to enhance conformational sampling.

Chapter 3 Analysis Method

3.1 Lipid Bilayer Properties

3.1.1 Experimental Approaches

Small angle x-ray and neutron scattering (SAXS/SANS) can be applied to obtain x-ray and neutron form factors as well as the bilayer thicknesses [93]. Without a structural model, headgroup-to-headgroup thickness (D_{HH}) can be obtained by measuring the distance between the peaks in the electron density profile (EDP) resolved from x-ray scattering, and the overall bilayer thickness (D_{B}) can be obtained from neutron scattering with protonated lipid and deuterated water [6]. With the experimentally estimated lipid volume (V_{L}) [94] and D_{B}, the SA/lip can be obtained by $SA/\mathrm{lip} = 2V_{\mathrm{L}}/D_{\mathrm{B}}$ [93]. Alternatively, for all experimental data used in this work, more accurate thicknesses and SA/lip can be obtained by fitting scattering density profile (SDP) structural model proposed by Kučerka et al. to both x-ray and neutron scattering data [95-97]. Our MD simulations were also compared with previously published experimental deuterium segmental order parameters (S_{CD}) [11, 98] and the nuclear spin-lattice (Zeeman) relaxation times (T_1) [11] measured by nuclear magnetic resonance (NMR).

3.1.2 Simulation Calculation Methods

3.1.2a X-ray and Neutron Form Factors

In our simulation of homogenous membranes, the atom density profile (ADP) was calculated by CHARMM [65] from which the EDP was obtained. The SIMtoEXP program [99] was used to obtain the form factors ($|F(q)|$) where q is a total scattering vector by

Fourier transformation of the electron/neutron density data. The reduced chi square (χ^2) were used to describe the level of agreement [99]

$$\chi^2 = \frac{\sqrt{\sum_{i=1}^{N_q} (|F_s(q_i)| - k_e|F_e(q_i)|)^2 / \left(\Delta F_e(q_i)\right)^2}}{\left(N_q - 1\right)} \tag{4}$$

where N_q is the number of data points, $F_s(q_i)$ is the simulated form factor, $F_e(q_i)$ is the experimental form factor, k_e is the scaling factor, and $\Delta F_e(q_i)$ is the experimental uncertainty of data i.

3.1.2b Surface Area per Lipid

For single-component bilayer, the SA/lip was directly calculated by the area of the simulation box divided by the number of lipids per leaflet. Block averages were used to obtain the statistical errors of SA/lip. For multiple-component bilayer, the X and Y coordinates of the representative atoms of each lipid (O3 for sterols, and C2, C21, and C31 for glycerol phosphate lipids) of the primary cell and the surrounding 8 images are imported into Qhull [100] which constructs the convex polygons for each atom (with all the interior angles less than 180°). The total area of the polygons is equal to the box size, from which the averaged estimated SA/lip for each lipid component will be obtained. The area compressibility modulus, K_A was calculated using,

$$K_A = k_B T \langle A \rangle / \sigma^2_{\langle A \rangle} \tag{5}$$

where k_B is the Boltzmann constant, $\langle A \rangle$ is the average surface area, T is the temperature, and $\sigma^2_{\langle A \rangle}$ is the variance of the area.

3.1.2c Bilayer Thicknesses

To calculate various membrane thicknesses, the electron density of each atom is combined into the following groups: a polar head group (i.e. acid/proton for PA, choline for PC, ethanolamine for PE, glycerol for PG, and serine for PS), phosphate, glycerol, carbonyl, methylene (CH2), methine (CH) (if there is any double bound), methyl (CH3), K^+ (for PA, PG, and PS), and water [101]. D_{HH} is the distance between two peaks in the EDP along the membrane normal (i.e., the z-axis) as shown in Figure A.1, which can be directly obtained. D_B is the distance between the midpoints of the volume probability profile of water molecules, and the hydrophobic thickness ($2D_C$) is the distance between the midpoints of the volume probability profile of bilayer's hydrocarbon acyl chains [6]. With the assumption that the volume of each component (on average) is independent of the membrane normal [102-104], which is the same assumption made in SIMtoEXP [99], D_B and D_C can also be directly calculated from the EDP. The volume probability relation is shown in Eq. (6). The sum of component probabilities equals 1 in each bin [99]:

$$1 = \sum_{i=1}^{N_c} p_i(z) = \sum_{i=1}^{N_c} V_i n_i(z) \quad (6)$$

where p_i is the probability of component i, N_c is the number of component in the lipid, V_i is component volume, and n_i is number density of component i. When the EDP of hydrocarbon acyl chain (including CH_3, CH_2, CH) reaches maximum, and the hydrocarbon acyl chain is the only component (i.e., $V_C = 1/n_{C,\max}$), D_C can be calculated as half of the distance between the positive and negative z values that correspond to $n_{C,\max}/2$ (i.e., $C_{\max}/2$ in EDP, where $C_{\max}$ is electron density of hydrocarbon acyl chain, as shown in Figure A.1). Similarly, when the EDP of water reaches maximum at the edge region of the lipid bilayer, and water is the only component except the small trace of K^+ ions, D_B can be calculated as the distance between the positive and negative z value that corresponds to $n_{\mathrm{W,max}}/2$ (i.e., $W_{\mathrm{max}}/2$ in EDP).

3.1.2d Deuterium Order Parameter

The simulated S_{CD} is defined as the ensemble average of the second order Legendre polynomial P_2 [105, 106]:

$$S_{CD}^{(i)} = |\langle P_2(cos\theta_i)\rangle| = |\langle 3\,cos^2\,\theta_i - 1\rangle/2| \tag{7}$$

where θ_i is the angle between C-H bond on the i^{th} carbon and the bilayer normal, and the angular bracket denotes a time and ensemble average. S_{CD} is a measure of the overall order of bilayer lipids, i.e., a higher S_{CD} value indicates more ordered lipid chains.

3.1.2e Hydrogen Bonding

One of the main advantages of simulation over experiment is the accessibility of further details in the atomic scale, e.g., hydrogen bonding to understand the fundamentals of lipid bilayer structural properties. The number of intra-lipid ($N_{\mathrm{HB}}^{\mathrm{intra}}$) and inter-lipid ($N_{\mathrm{HB}}^{\mathrm{inter}}$) hydrogen bonds were calculated with the following hydrogen bond definition: distance between proton and the acceptor pairs less than 2.4 Å and the angle of donor-proton-acceptor greater than 150° including the periodic boundary conditions. The H and H-acceptors involved in hydrogen bonding for each head group are shown in Figure 2; the phosphate (PO_4^-) and carbonyl (CO) groups are the common H-acceptors for PA, PE, PG, and PS lipids, and PS lipids have additional H-acceptor carboxylate (COO^-) group. Note that there are no hydrogen bonds between PC lipids as it has no proton donor.

3.1.2f Lipid Lateral Clustering

Lateral clustering analysis was performed to visualize the effects of hydrogen bonding and other inter-lipid interactions. The clustering of lipids was analyzed using python scikit-learn [107] implementation of density-based spatial clustering of applications with noise (DBSCAN) algorithm [108]. A maximum distance of 5.5 Å (6.5 Å is used for single-component PC and PG lipid bilayer) between the center of mass of head group atoms (phosphate and the above atoms as shown in Figure 2), and minimum cluster size of three lipids per cluster were applied [109].

3.1.2g Correlation Time and Spin-Lattice Relaxation Time

As the bilayers reach thermal equilibrium, the systems are relatively static macroscopically. However, all the atoms are in constant motions due to the thermal fluctuations [110]. To characterize the dynamics of C-C vector spins in the equilibrated system, the second rank reorientational correlation function is calculated to find the correlation between the C-C vectors [111]:

$$C_2(t) = \langle P_2(\hat{u}(0) \cdot \hat{u}(t)) \rangle = \left(3\,\langle (\hat{u}(0) \cdot \hat{u}(t))^2 \rangle - 1\right)/2 \tag{8}$$

where $\hat{u}(0)$ and $\hat{u}(t)$ are the normalized C-C unit vectors at time 0 and *t*, respectively. P_2 is a second order Legendre polynomial, and $C_2(t)$ is time and ensemble averaged correlation between $\hat{u}(0)$ and $\hat{u}(t)$ [111]. Time *t* varies among the simulation time of the equilibrated portion with a step of 4 ps. $C_2(t)$ describes how much the spin of a C-C vector at time *t* correlates to the spin of a C-C vector at time zero. As *t* increases, the correlation function $C_2(t)$ decays first and then reaches a plateau. There can be multiple stages of correlations, short or long range order [110], so that simulated $C_2(t)$ are fitted to two exponential decay functions and a constant:

$$C_2(t) = a_0 + a_1 e^{-t/\tau_1} + a_2 e^{-t/\tau_2} \tag{9}$$

where the plateau a_0 is a long-time value of the correlation function. a_1 and a_2 are coefficients of two exponential decay functions, τ_1 and τ_2 are two correlation times corresponding to fast and slow spin-lattice relaxations of lipids, respectively. τ_2 for C_{22}-C_{32} can be used to describe the correlation time of lipid wobble, which is restricted rotation of the long axis of lipids [112, 113].

Another parameter that describes the NMR relaxation is the spin-lattice relaxation time (T_1). By assuming the pure dipolar relaxation between the ^{13}C nucleus and its attached protons, the spin-lattice relaxation rate ($1/NT_1$) are related to spectral density [114].

$$R_1 = \frac{1}{NT_1} = \frac{1}{10}\left(\frac{\hbar\gamma_c\gamma_h\mu_0}{4\pi r_{c-h}^3}\right)^2 [J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)] \tag{10}$$

where N is the number of protons attached to the carbon, $\hbar$ is Plank's constant divided by 2π, and γ_c and γ_h are the gyromagnetic ratio for C and H (in radian·$s^{-1}T^{-1}$), respectively. μ_0 is vacuum permeability, $r_{c\text{-}h}$ is the effective C-H bond length, 1.117 Å [115], ω_C and ω_H are the Larmor frequencies of C and H (in radian·s^{-1}), respectively, and the ratio $\omega_H/\omega_C = 3.9764$, $J(\omega)$ is the spectral density of the second rank reorientational correlation function:

$$J(\omega) = \int_0^\infty C_2(t)\cos(\omega t)dt \tag{11}$$

Based on calculated simulated $C_2(t)$ and the NMR carbon Larmor frequency, Eq. (11) was calculated by numerical integration (trapezoidal rule), and the corresponded simulated spin-lattice relaxation rate was calculated by Eq. (10), from which spin-lattice relaxation time T_1 was obtained for comparison with experimental T_1.

3.2 Protein Conformational Analyses

The pore size of the cavity can be used to describe the level of opening/closing of LacY structure on the cyto- and periplasm. The pore radius of LacY is calculated by using HOLE program [116], which searches the maximum radii of spheres in the cavity without overlapping any LacY atom with known van der Waals radii [117]. Metropolis Monte Carlo simulated annealing procedure [118, 119] is applied to adjust the point on the xy-plane (orthogonal to the pore vector which is in the z direction). Then the process continues by taking small displacement in the direction of the approximate vectors to find more optimal center points on the new plane until the specified maximum radii 7 Å is reached which determines the end of the pore. Based on z axis of protein, the cutoff of 7 Å radius is near the transition from being within the protein and in the bulk phase, therefore the z range corresponding to the cutoff radius of 7 Å would contain almost all LacY structure. The z coordinate of the center points of spheres and their radii are stored for each frame, then the final pore radii are 1-ns block averaged with the z coordinate bin size 0.25 Å.

The C_α distance of residue pair is commonly analyzed to describe the conformation of a protein. DEER data [62] provides the distance of the spin labels attached to the nine residue pairs near the end of helices of LacY. With the available inward-facing LacY crystal structure and the DEER spin label distance, the residue pair distances of outward-facing LacY are predicted. In our analysis, we calculate the distance of those same residue pairs from our simulations and then compare the difference. However, due to large uncertainty of DEER spin label distance measurement, the estimations involve large inaccuracy (more detail will be provided in Section 7.2).

The root mean standard displacement (RMSD) calculations are also commonly applied. By selecting the initial LacY structure as the reference, RMSD can be calculated to describe the stability of LacY structure by calculating the magnitude of conformational change of LacY at each time step. Besides the initial structure, the LacY structure of different states, or different LacY models can also be used as a reference structure to compare the conformational difference over time. Trajectory overlap can be performed to visualize the conformational difference of two LacY structures.

3.3 General Tools

Python combined with the Shell scripts was used for the general data analyses. Visual Molecular Dynamics (VMD) [120] was used to create snapshots of the bilayers and LacY. Gnuplot [121] was used to generate the structure property plots.

Chapter 4 Results on Single-Component Membranes

This chapter covers two types of CHARMM36 force fields specifically used here for simulating the single-component or pure lipid bilayer in two separate sections, but can be used for more complex models with multiple lipids and proteins. The all-atom model which explicitly simulate all atoms, and united-atom atom model which lumps each carbon and its attached hydrogens (CH, CH2, and CH3) into a single representative atoms (CH1E, CH2E, and CH3E) corresponding to each are presented. Section 4.1 demonstrates the simulation results with C36FF-AA model in order to test the accuracy of it and also to show the capability of the model to provides the reasonable in-depth analyses results. The further detail related to C36FF-AA is published in [28, 109]. Since C36FF-AA is been proved to be an accurate force field, to improve the simulation efficiency while keeping the similar accuracy, the C36FF-UA is developed and tested, the results of which are demonstrated in Section 4.2. The development of C36FF-UA model lays groundwork for more complex membrane models, which are too complex and computational expensive to study with the all-atom force field.

4.1 All-atom CHARMM36 Force Field

4.1.1 Test the accuracy

Biomembrane functionalities strongly depend on their structures. Therefore, considerable efforts have been made to estimate the membrane structural parameters, such as surface area per lipid (SA/lipid), volume per lipid (V_L), hydrophobic thickness (D_C), overall bilayer thickness (D_B), and deuterium order parameters (S_{CD}). Instead of analyzing the data separately, Kučerka et al. [104] proposed to simultaneously analyze both x-ray and neutron scattering (SAXS/SANS) data to determine the lipid structural parameters. By this method, experimental data on common lipid bilayers were measured to estimate their SA/lipid and thicknesses by fitting the scattering density profile (SDP) model [104] to the scattering data.

MD simulation is also used to determine the lipid membrane structural properties. The all-atom CHARMM36 force field (C36 FF-AA) developed by Klauda et al. [14] allows for accurate constant molecule number, pressure, and temperature (*NPT*) simulations of pure lipid bilayers. Moreover, it is capable of matching various structural and dynamical properties of lipid membranes, which results in more realistic membrane models, including lipid molecules containing branched chains [122]. To further verify the accuracy of the C36 FF-AA, we provide MD simulation results of varied types of glycerol phosphate lipid bilayers (Figure 2) over a wide range of temperatures and comparisons with experimental data obtained by Kučerka et al. [6] and NMR data obtained by Douliez et al. [98] and Brown et al. [11].

4.1.1a X-ray and Neutron Form Factors

Comparison of x-ray and neutron scattering with MD simulations is most direct when the form factors are used [101, 123]. These experimental methods provide detailed characteristics of the overall and component bilayer structure and thus are an excellent tool to assess the accuracy of the C36FF. This section will briefly discuss the results of our MD simulations in comparison with all available form factor data for saturated and unsaturated lipids. To quantitatively describe the level of agreement between simulated and scaled x-ray scattering experimental form factors, the χ^2 values [99] in Eq. (4) were obtained from the SIMtoEXP program for the first two lobes only (Figure 11a and Table A.3 to A.4); a χ^2

value lower than 8 should be considered the excellent agreement, and lower than 30 reasonably good agreement.

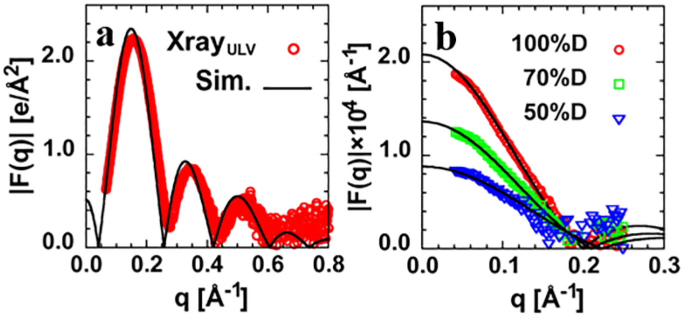

Figure 11 Comparison of MD-based form factor to experiment.
a) MD-based form factor (the black curve) of DOPG at 30 °C to experimental x-ray form factor (red circle) [8]. b) Comparison of the MD-based form factor of DLPG at 60 °C to experimental deuterium neutron scattering data [8] in three contrasts. %D is %D2O. The black curves are the MD-based form factor at each corresponding deuterium concentration.

Saturated lipids Similar to DLPC in [28], the simulated XFF of the DLPE and DLPG agree well with experimental x-ray scattering data at a temperature range of 30-60 °C (Figure A.2a-f, Table A.3). The simulated XFF of 12:0, 14:0, and 16:0 PC lipids at 60 °C agree well with experiment at the first lobe, but show minor deviation in the second lobe (Figure A.2g-i), while 18:0 PC shows minor deviation in both lobes (Figure A.2j). Moreover, as chain length increases, the χ^2 of saturated PC lipids and the deviation in the second lobes slightly increases (Table A.3, Figure A.2g-j). The difference between the simulated and experimental XFF in the first lobe is due to inaccurate lipid density, while the deviation in the second lobe is caused by inaccuracies in either the lipid or water density [28, 124]. Based on these, we can conclude that the water density maybe slightly inaccurate for the saturated PC lipids at relatively high temperature (60 °C), while the lipid density may be slightly inaccurate for DSPC at 60 °C. Overall, the simulated form factors of the saturated lipids demonstrate great agreement with experiments and also results in χ^2 in the excellent agreement range.

Unsaturated lipids χ^2 of the XFF of mono-unsaturated PC lipids are generally greater than saturated lipids, and the mixed-unsaturated lipids have intermediate values (Table A.4), which implies that the agreement level increases as the saturation level increases. As shown in Figure A.2n-r, the simulated XFF of the 18:1 PC lipids agree well with experiment, while the 16:1, 20:1, 22:1, and 24:1 PC lipids show minor deviations in both the first and second lobes. The mono-unsaturated PC lipids with long chain (22:1, and 24:1 PC) results in more deviation in the first lobe. These imply that the water density is slightly off in the mono-unsaturated PC lipids, and the lipid density may be slightly inaccurate for lipids with long acyl chains, which agree with the conclusion from the saturated PC lipids. The mixed-unsaturated PO/SO PC and PG lipids match experimental form factors well in all lobes (Figure A.2s-t,w-x), while PO/PO PE lipids show minor deviation in the second lobe (Figure A.2u-v).

Except for POPS, the NFF of all lipids (with varied tail types and temperatures) show excellent agreement with experiments (Figure 11b). For POPS, the simulated NFF deviates from experimental NFF in the region of $q < 0.2$ (Figure A.3), suggesting possible deviation in its D_B and also SA/lip as neutron scattering data is more sensitive to D_B and SA/lip [10, 124] (The comparisons will be provided and discussed below).

Further error analysis χ^2 calculated by Eq. (4) may not agree with the comparison result by directly visualizing the form factor plot because χ^2 is strongly affected by the experimental uncertainty $\Delta F_e(q_i)$ and number of data points N_q. e.g., the experimental uncertainty for DOPC is 0.01, while it is 0.05 for DXPC which results in much higher χ^2 for DOPC even though the actual agreement of DOPC between simulation and experiments is very well. The comparison plots show that the disagreement level looks similar for

DXPC and POPS (Figure A.2n and Figure A.3a), and the significant higher χ^2 of POPS (561) may be due to different number of data points in the first two lobes of POPS (~300), and DXPC and other lipids (~400). For DNPC (Figure A.2r) and some PG lipids (Figure A.2a, m, and x), the deviation between the simulated and experimental form factor is probably due to experimental errors as their XFF does not cross $|F(q)|=0$ for each lobe. The minor deviation in the second lobe may also be due to low resolution resulted from unilamellar vesicle (ULV) sample in the second lobe, while the use of oriented bilayer sample that has higher resolution in the second (and third) lobe may result in better agreement (Figure A.2q, s) [28].

Overall, the agreement between experimental and MD-simulated form factors by C36FF is excellent for almost all lipid types, but there are a few deviations that suggest either experimental errors or minor inaccuracies in the water model or the lipid force field.

4.1.1b Surface Area per Lipid

Temperature and head group dependence For all lipids, the SA/lip increases as temperature increases (Figure 12a and Table A.5-A.7), which agrees with the results of our previous study on PC lipids [28]. The slope k_T generally increases as chain length increases, indicating that the temperature effect is stronger for lipids with longer acyl chains. Considering only the head group, the SA/lip is affected by the steric size of the head group and its interactions via hydrogen bonding. Among the five lipid types, PC, PG, and PS have larger head groups than PE lipids, and PA lipids have the smallest head group. For all tail types, the PG lipids have the greatest SA/lip among five head groups due to its intra-lipid hydrogen bonding (the details of hydrogen bonding will be provided in Section 4.1.2a). PC lipids have a large head group and no hydrogen bonding, yielding intermediate SA/lip. PA, PS, and PE lipids have relative small SA/lip, in which PA is close to or slightly greater than PS, and followed by PE lipids (Figure 12a-c). The PS and PE lipids form more intermolecular hydrogen bonds than other head group lipids, which results in lower SA/lip. PS lipids have more hydrogen bonds but a larger head group size than PE lipids, yielding only a slightly higher SA/lip (than PE lipids) in most systems.

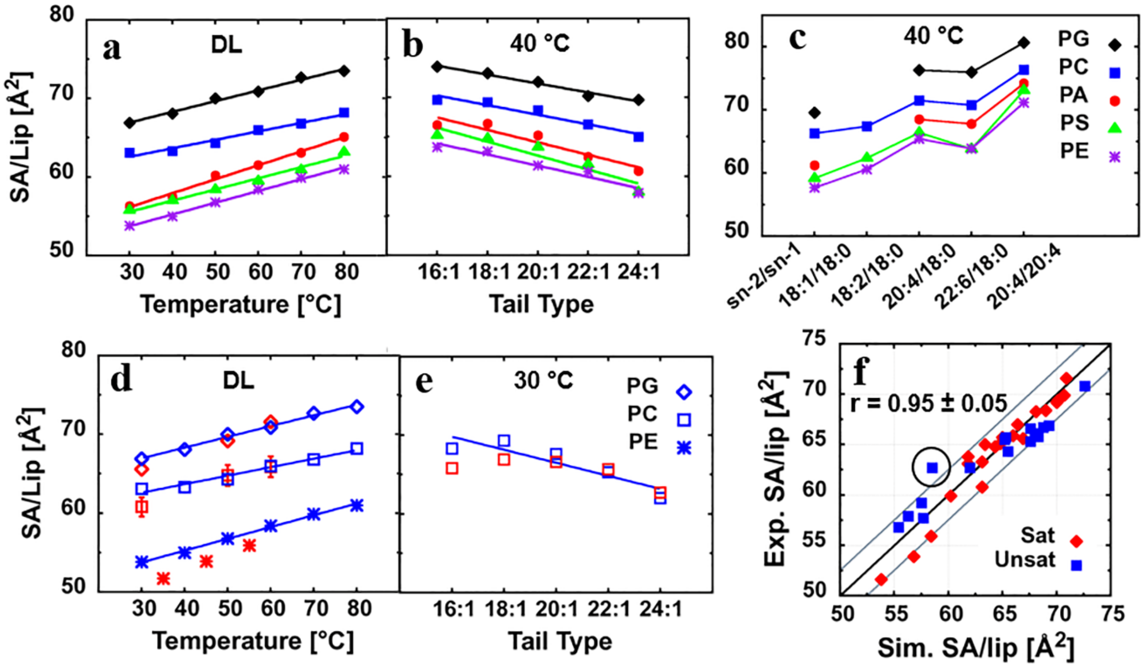

Figure 12 The variations of SA/lip and the comparison to experiment.
The variations of SA/lip as a function of a) temperature, b) mono-unsaturated lipid tails, and (c) poly-unsaturated lipid tails. d) Comparison of simulated SA/lip with experimental data [6-8, 10] at different temperatures. The blue represents simulation data, and the red represents experimental data. The slopes are −0.14 ± 0.01 (simulation) and −0.20 ± 0.01 (experiment) for PG lipids, −0.11± 0.01 (simulation) and −0.17 ± 0.02 (experiment) for PC lipids, and −0.15 ± 0.01 (simulation) and −0.21 ± 0.01 (experiment) for PE lipids. The error bars show the uncertainty of experimental values of DLPC. e) Comparison of simulated SA/lip with experimental data for mono-unsaturated lipid tails. The error bar is not shown when its uncertainty is within the symbol size. f) Overall comparison of simulated and experimental SA/lip [6-10] with the correlation coefficient and its standard deviation. The black line is x=y, and two gray lines are the boundaries of 95% confidence interval of experimental SA/lip. The circle indicates the most deviated SA/lip of POPS with 62.7 Å^2 (experiment) and 58.5 ± 0.3 Å^2 (simulation). The error bars of the simulated SA/lip are not shown as they are within the symbol size.

Tail dependence For both saturated and mono-unsaturated lipids, except for DSPC lipids at 80 °C, the SA/lip decreases as the chain length increases (Figure 12b, Table A.5) due to more ordered chain packing with longer acyl chains. As shown in Figure 12c, the SA/lip of SL (18:0/18:2) is higher than SO (18:0/18:1), indicating that the SA/lipid increases as the level of unsaturation of chain increases because of the kinks (causing less ordered packing) formed by additional double bonds. For SA (18:0/20:4) and SD (18:0/22:6) lipids, the SA/lip slightly decrease in most systems (Figure 12c, Table A.5-A.7), suggesting that the chain length effect is stronger than the double bond effect when both chain length and unsaturation increase. However, at 30 °C, the SA/lip of PC and PG lipids increase, implying that double bonds may be the dominant affect at lower temperatures, while the chain length affect is dominant at higher temperatures. Comparing SA to DA (20:4/20:4), as more double bonds than chain length are added, a double bond affect is much stronger, yielding much higher SA/lip at both 30 °C and 40 °C.

Comparison to experimental data As shown in Figure 12d and Figure A.4, for saturated PG, PC, and PE lipids, the simulated SA/lip agree with experimental data better at intermediate temperatures (50 and 60 °C), and minor deviations occur at low and high temperatures. The slope k_T of simulated SA/lip is lower than experimental k_T for all three head group types (Figure 12d). This may be due to the inaccurate thermal expansion of TIP3P water model, for which the density of pure water agrees well with experiment at intermediate temperatures (40 and 50 °C), but higher and lower than the experiment at low and high temperature, respectively [28]. The over hydration in the head group results in looser packing of lipids at lower temperatures. Mono-unsaturated PC lipids agree better at higher chain length, while the simulated SA/lip is higher than experimental data at lower chain length, which may also due to higher water density. There is an unexpectedly low SA/lip at 16:1 and also 24:1 that deviate from a linear trend for both simulation and experiment (Figure 12e), based on which the fitting of the 2nd order polynomial to the SA/lip (and also thicknesses) of mono-unsaturated PC lipids is proposed by Kučerka et al. [10]. In this study, however, we did not apply a non-linear model because of the SA/lip from 18:1 to 22:1 are quite linear and we do not have 14:1 SA/lip to fully support the non-linearity. Overall, the simulated and experimental SA/lip are strongly correlated with the Pearson correlation coefficient of 0.95 ± 0.05 (Figure 12f). Besides the SA/lip of POPS that shows the highest deviation (4.2Å^2), the simulated SA/lip of all other lipids are within the 95% confidence interval from the experimental values. The average absolute deviation (AAD) between the simulated and experimental SA/lip is 1.3 Å^2, and the average error (%AAD) is 2.1%.

4.1.1c Thicknesses

Three types of thicknesses are calculated: (i) the overall bilayer thickness (D_B) that is a measure of water penetration into the bilayer, (ii) the hydrocarbon acyl region thicknesses ($2D_C$) that is a measure of the bilayer hydrophobic thickness, and (iii) the headgroup-to-headgroup thickness (D_{HH}) that is a measure of the distance between the hydrophilic regions of the bilayer. Each thickness provides information on various regions of the lipid membrane and can be compared with SDP model-fitted experimental thicknesses.

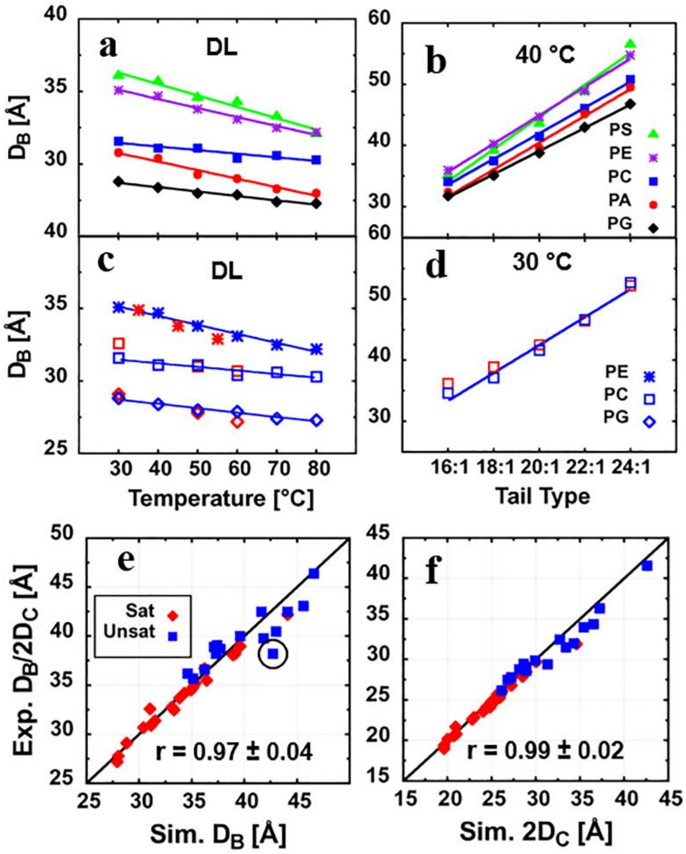

Figure 13 The variations of D_B and the comparison to experiment.
The variations of D_B by a) temperature and b) mono-unsaturated lipid tails. Comparison of simulated D_B with experimental data [6-8, 10] as a function of c) temperature and d) mono-unsaturated lipid tail length. The blue represents the simulation data, and the red represents the experimental data. Overall comparison of simulated and experimental data [6-10] of e) D_B and f) $2D_C$, shown with correlation coefficients and their standard deviations. In e), the circle indicates the D_B of POPS at 30 °C: 38.2 Å (experiment) and 42.7 Å (simulation).

Temperature, head group, and tail dependence As temperature increases, the thicknesses generally decreases (Figure 13a, Figure A.5, and Table A.8-A.16), which agrees with the PC lipid case reported in [28]. The temperature dependence of PC lipids is weaker than the other lipids. The D_B, $2D_C$, and D_{HH} of the PS and PE lipids are greater than PC and PA lipids, and the PG lipids have the smallest values (Figure 13a-b and Figure A.5). The high intermolecular hydrogen bonds in PS and PE lipids may also contribute to their large thicknesses due to the tight packing. Even though PC lipids also have a large head group, three methyl groups in choline may make PC head group less hydrophilic and has a tendency to bend toward the hydrophobic region (Figure 2). The PG lipids have lower thicknesses than PA lipid (which has the smallest head group) because the PG head group bends toward phosphate due to its high intramolecular hydrogen bonding, causing loose lipid packing (i.e., larger SA/lipid) and thus smaller thicknesses. More specifically, the D_{HH} of PS is greater than PE lipids for all tail types, while this is not always the case for D_B and $2D_C$. This implies that head group size and intermolecular hydrogen bonds may play more important roles in D_{HH}. The PC lipids have higher D_B but lower $2D_C$ than PA lipids (Figure A.5), suggesting that the head group size plays more roles in D_B, while hydrogen bonds may play more roles in $2D_C$. As the chain length increases, the thicknesses increase as expected (Figure 13b and Figure A.5). Comparing SO to SL, the thicknesses decrease as unsaturation level increase (Figure A.5 and Table A.8-A.16). By changing the tail from SA (20:4/18:0) to SD (22:6/18:0), the thickness increases, implying that the chain length has stronger effects than the double bond for the same increment in chain length. Comparing SA to DA (20:4/20:4), as the increment in the double bond is twice more than the chain length change, the thickness decreases.

Comparison to experimental data Except for DSPC at 60 °C and POPS at 30 °C, the simulated D_B agree well with experiment (Figure 13c-d, Figure A.6, and Table A.8-A.10). Simulated D_C of saturated PE lipids are slightly lower than experimental data at low temperature, while the PG and PC lipids match experimental data very well (Figure A.6a-c). Simulated D_{HH} of PG lipids agrees with experimental data well, while the PE and PC lipids are higher than experimental data, which is likely due to over hydration in the head group region (Figure A.6d-f). The D_{HH} results agree with the observations that simulated

XFF of PG lipids agree with experimental XFF well, while the simulated XFF of PC lipids are slightly shifted left (Figure A.2). The thicknesses of simulated POPS at 30 °C shows most deviations, in which D_B, $2D_C$, D_{HH} are 4.5, 1.9, and 2.8 Å higher than the experiment, respectively, which also agrees with the fact that the XFF is shifted to the left and the noticeable disagreement in NFF. Overall, the simulated and experimental D_B, D_C, and D_{HH} agree reasonably well, and they are strongly correlated with the Pearson correlation coefficient of 0.97 ± 0.04 (D_B), 0.99 ± 0.02 (D_C), and 0.96 ± 0.04 (D_{HH}), respectively (Figure 13e-f, Figure A.7, and Table A.8-A.16). The AAD and %AAD are 0.9 and 2.4% (D_B), 0.7 and 2.3% (D_C), and 1.9 Å and 5.3% (D_{HH}), respectively.

4.1.1d Deuterium Order Parameter (S_{CD})

Another commonly measured experimental bilayer property is S_{CD}, which describes the order of each carbon in the acyl chain. This provides details of the molecular structure of lipids with respect to the bilayer normal. Except for *sn*-2 of poly-unsaturated lipids due to the presence of double bonds, the S_{CD} of carbon 4 to 7 are invariant and have meaningful values (i.e., not too small) for each lipid system (Figure A.8a-f), so the temperature and chain length dependence is described by the averaged S_{CD} of carbon 4 to 7.

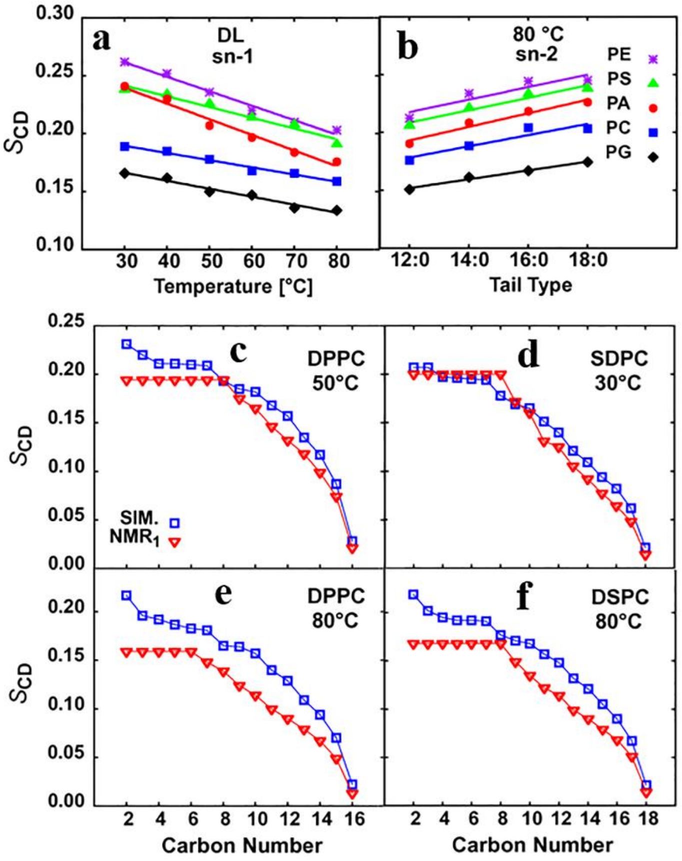

Figure 14 The simulated C4-C7 averaged S_{CD} and comparison of S_{CD} to experiments. a) *sn*-1 at different temperatures and b) *sn*-1 with varied mono-unsaturated lipid tail length. Comparison of simulated *sn*-1 S_{CD} with experimental data [11] at relatively low temperatures (c) DPPC at 50 °C and d) SDPC at 30 °C) and relatively high temperature (e) DPPC at 80 °C and f) DSPC at 80 °C). Both the simulation and experimental data are sorted from C2 to the last *sn*-1 carbon.

Temperature, head group and tail dependence As temperature increases, C4-C7 averaged S_{CD} of all lipids decreases, as the lipids behave closer to liquid and become more disordered (Figure 14a and Figure A.8g). The *sn*-1 S_{CD} of PS, PE, and PA lipids show stronger temperature dependence than the PC and PG lipids, which is also the case for the *sn*-2 chain (Table A.17 to Table A.22). Generally, PE lipids have the highest S_{CD}**,** followed by PS, PA, PC, and PG lipids, which is the inverse order of SA/lip, as they are both related to the lateral packing of the lipid bilayer; i.e., the tighter the lipid packing is, the lower the SA/lip are and the higher S_{CD} are, which is also the direct consequence of head group size and hydrogen bonding. As the chain length increases, S_{CD} of both saturated and unsaturated lipids increases (Figure 14b and Figure A.8b). Unsaturation lowers S_{CD} of the double bond and neighboring carbons and thus the C4-C7 averaged S_{CD} (Figure A.8 and Table A.19, and Table A.22).

Comparison to experimental data Simulated *sn*-1 S_{CD} of PC lipids agree reasonably well with NMR results [11] (Figure 14c-d and Figure A.9a-f). However, the deviation of S_{CD} in the *sn*-1 chain occurs for DLPC and DMPC at 50 °C (Figure A.9d-e), and DPPC and DSPC at 80 °C (Figure 14e-f). For the *sn*-2 chain of PC lipids, the first (C2) and last carbons of the lipid tails agree well with NMR S_{CD} measured by Brown et al. [11], while deviations occur for all carbons between them (NMR_2 in Figure A.9g-l). However, the simulated S_{CD} agree better with S_{CD} data (NMR_1 in Figure A.9) measured by Douliez et al. [98]. The constant experimental S_{CD} for the first few carbons (starting C3) obtained by Brown et al. [11] is caused by unavailable signal splitting for these carbons. Therefore, the inaccurate experimental S_{CD} in these carbon atoms may cause the deviation between the simulated and experimental S_{CD}. Moreover, Douliez et al.'s results are expected to be more accurate as only the *sn*-2 chain is measured, while Brown et al. measures both *sn*-1 and *sn*-2 chains simultaneously, which involves additional derivation and uncertainty in NMR data analyses.

4.1.2 Further In-depth Study of Lipid Bilayer

4.1.2a Hydrogen Bonding

Inter-lipid The number of inter-molecular hydrogen bonds per lipid ($N_{\mathrm{HB}}^{\mathrm{inter}}$) of varied lipid types is summarized in Table A.23-A.25. As temperature increases, $N_{\mathrm{HB}}^{\mathrm{inter}}$ decreases for all lipid types, and PS lipids show stronger temperature dependence than the others (Figure 15a). PS and PE lipids have higher $N_{\mathrm{HB}}^{\mathrm{inter}}$ than PG and PA lipids since $\mathrm{NH_3^+}$ in PS and PE lipids provides three protons for hydrogen bonding with the surrounding phosphate. PS lipids have higher $N_{\mathrm{HB}}^{\mathrm{inter}}$ than PE lipids due to its additional H-acceptor in carboxylate ($\mathrm{COO^-}$) in its head group which results in formation of complex network (Figure 15c). The PG lipids have two protons, resulting in intermediate $N_{\mathrm{HB}}^{\mathrm{inter}}$, while PA lipids have only one proton, which results in the lowest $N_{\mathrm{HB}}^{\mathrm{inter}}$. For both saturated and mono-unsaturated lipids, as chain length increases, PS lipids show obvious increase in $N_{\mathrm{HB}}^{\mathrm{inter}}$, but the mono-unsaturated PE, PG, and PA lipids show very little increase (Figure 15b and Figure A.10). Comparing PO to PL, and SO to SL lipids, except for SLPS at 40 °C, $N_{\mathrm{HB}}^{\mathrm{inter}}$ decreases as unsaturation level increases (Table A.25) as more double bonds generate more kinks in the acyl chain that prevent tight lipid packing. As the tail changes from SL to SA, $N_{\mathrm{HB}}^{\mathrm{inter}}$ of both PE and PS lipids decrease, suggesting that the unsaturation level effect is dominant. As tail changes from SA to SD, the PA, PE, and PG lipids show slightly decreasing $N_{\mathrm{HB}}^{\mathrm{inter}}$ at 30 °C, while slightly increasing $N_{\mathrm{HB}}^{\mathrm{inter}}$ at 40 °C. However, PS lipids have increasing $N_{\mathrm{HB}}^{\mathrm{inter}}$ at both 30 and 40 °C, suggesting that the chain length is the dominant effect.

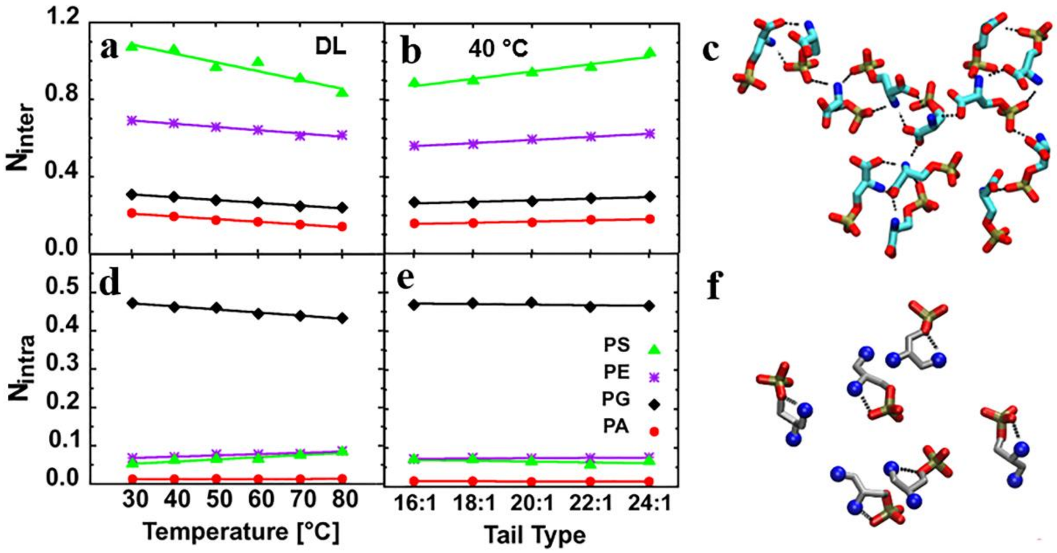

Figure 15 The inter-lipid and intra-lipid hydrogen bonding in the lipid bilayers.
The number of inter-lipid hydrogen bonds at a) different temperatures of DL lipids and b) different chain lengths of mono-unsaturated lipids. c) A snapshot (top view) of a DLPS network by inter-lipid hydrogen bonds at 30 °C. The number of intra-lipid hydrogen bonds at d) different temperatures of DL lipids and e) different chain lengths of mono-unsaturated lipids. f) A snapshot (top view) of a DLPG head group forming a ring-like structure by intra-lipid hydrogen bonds at 30 °C.

Intra-lipid The number of intra-molecular hydrogen bonds per lipid ($N_{\mathrm{HB}}^{\mathrm{intra}}$) is summarized in Table A.26 to Table A.28. As temperature increases, $N_{\mathrm{HB}}^{\mathrm{intra}}$ decreases for PG lipids, increases for PS and PE, and is constant for PA lipids (Figure 15d). PG lipids have the highest $N_{\mathrm{HB}}^{\mathrm{intra}}$ and prefer intra-lipid over inter-lipid hydrogen bonding, because PG lipids have two hydroxyl groups as proton donors (Figure 2) and form stable ring-like structures through intra-lipid hydrogen bonds (Figure 15f). PE and PS lipids have much smaller $N_{\mathrm{HB}}^{\mathrm{intra}}$ as their proton donors are mostly involved in stable inter-molecular hydrogen bonds. PA lipids have the lowest $N_{\mathrm{HB}}^{\mathrm{intra}}$ and their $N_{\mathrm{HB}}^{\mathrm{intra}}$ are from the intra-lipid hydrogen bonds between the proton and the two carbonyl groups, while the phosphate oxygens make more contributions in $N_{\mathrm{HB}}^{\mathrm{intra}}$ of the other head group types. Moreover, the $N_{\mathrm{HB}}^{\mathrm{intra}}$ do not show any dependence on tail types for all head groups (Figure 15e, Figure A.10c-d, and Table A.26 to Table A.28).

4.1.2b Lipid Clustering

The SA/lip describes the ensemble averaged lipid packing (based on the entire bilayer system), while the cluster size can characterize the local packing behavior of lipid head groups (since we are using the center of mass of the head group atoms for clustering analysis; see Section 3.1.2f). As shown in Figure 16a, for DLPA at the end of the simulation at 30 °C, five clusters are formed with average six lipids per cluster. Both hydrogen bonding and lipid-K^+ interactions contribute to the formation of clusters in PA lipids. The corresponding simplified schematic plot based on the center of mass of PA head groups is shown in Figure 16b. With the minimum cluster size of three lipids, the cluster size strongly depends on a cutoff distance (D_{cut}) used for clustering analysis (Figure 16c). As D_{cut} increases, the cluster size increases, and for each head group type, there is a transition D_{cut} over which the cluster size is not sensitive to D_{cut} variation. The transition D_{cut} is about 5.5 Å for PS, PE, and PA, and 6.5 Å for PC and PG lipids. Despite different D_{cut}, the cluster size is about four lipids per cluster for all head group types. The increasing rate of the cluster size of PA lipids shows a local maximum at $D_{\mathrm{cut}} = 4.6$ Å (a), at which the probability of both hydrogen bonds and lipid-K^+ interactions (considering the K^+ within 4 Å of phosphate O^-) reach a maximum (Figure A.11b). At $D_{\mathrm{cut}} = 5.2$ Å, the increasing rate of cluster size reaches its minimum, and the cluster size curve approaches the transition point, which is correlated with the dramatic decrease in the probability of both hydrogen bonds and lipid-K^+ interactions. For head groups larger than PA, the head group distance involving hydrogen bonds and lipid-K^+ interactions span a larger distance range, and there may be more types of interactions that make it difficult to directly relate individual interactions to the cluster size. Therefore, to capture all possible interactions that contribute to lipid clusters (especially hydrogen bonding for PE, PG, and PS lipids), a D_{cut} of 6.5 Å was used in our analysis.

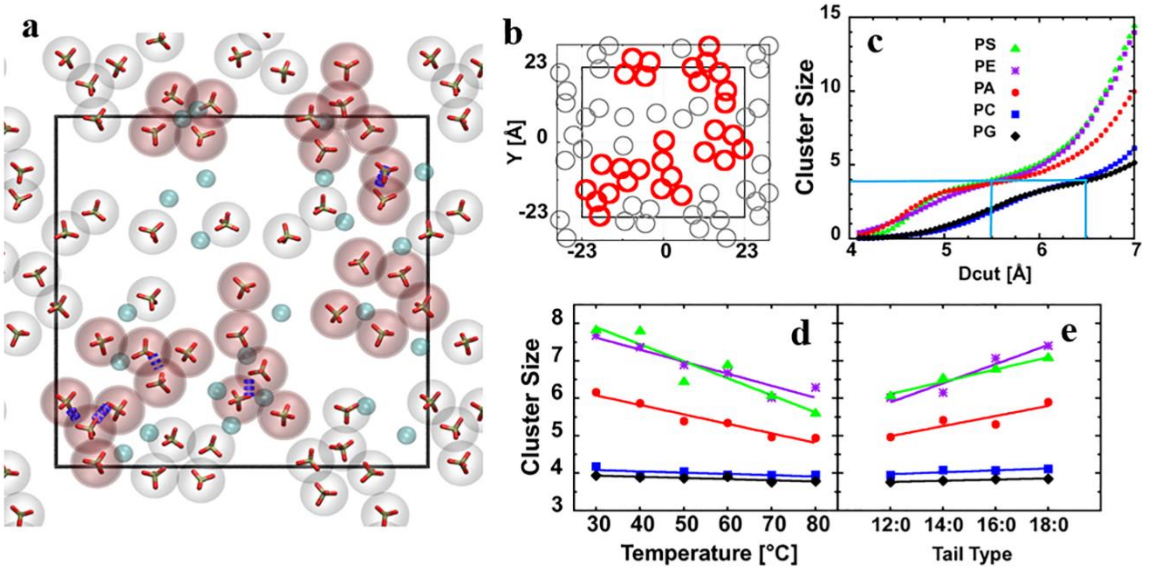

Figure 16 Snapshot of DLPA and the clustering scheme plots and the analysis results: a) Snapshot of the top leaflet of DLPA at the end of the simulation at 30 °C. The cyan spheres are K^+ ions. The orange and red in licorice represent the phosphate, the red spheres indicate the head groups that form clusters, and clear spheres indicate the ones that do not form a cluster. The blue dashed lines represent the hydrogen bonds. b) Schematic plot of DLPA clusters based on the center of mass of PA head groups. c) The number of lipids per cluster of saturated lipids as a function of distance cutoff used for clustering analysis. d) The number of lipids per cluster of saturated lipids at different temperatures. e) The number of lipid per cluster of saturated lipids with different chain lengths.

Temperature, head group, and tail dependence The average number of lipids per cluster of saturated lipids at varied temperatures is summarized in Table A.29. As temperature increases, the cluster size of PS, PE, and PA lipids decrease, while it does not change for PC and PG lipids (Figure 16d) since the applied D_{cut} of 6.5 Å is their transition cutoff distance. PS and PE have the highest cluster size due to high $N_{\mathrm{HB}}^{\mathrm{inter}}$. PS lipids have a larger head group and higher $N_{\mathrm{HB}}^{\mathrm{inter}}$ than PE, and these two counter effects result in a similar cluster size as in PE lipids. PA lipids have the smallest head group size and the lowest

$N_{\mathrm{HB}}^{\mathrm{inter}}$, which result in the intermediate cluster size. The cluster size of PC is small as it has no hydrogen bonding, but the results show that electrostatic interactions are strong enough in PC lipids to contribute to a small clustering of lipids. For example, as shown in Figure A.11c-e, the N^+ of choline group and two O^- of phosphate group point toward each other in the clustered DLPC to form connections due to electrostatic interactions. The PG lipids have the smallest cluster size due to the intra-lipid hydrogen bonding that enables formation of the bulky ring-like structure and thus prevents the tight packing of PG lipids.

4.1.2c Lipid Motions

Temperature and head group dependence The slow overall lipid relaxation time, i.e., the correlation time τ_2 of the C_{22}-C_{32} vector obtained by fitting $C_2(t)$ to Eq. (9), is a measure of the mobility of lipid, which is commonly known as lipid wobbling. A larger τ_2 indicates lower mobility. For all head groups, as temperature increases, τ_2 decreases due to more thermal motion (Figure 17a, Table A.30 to Table A.32). In particular, the additional inter-lipid hydrogen bonds of the PS lipids due to its carboxyl group in the head group allow formation of a stable multiple lipid complex network (Figure 15c), which significantly reduces mobility and thus results in a much higher value of τ_2 (i.e., nearly twice larger than those in the other head group types). In addition, as temperature increase, the τ_2 of the PS lipids show stronger temperature dependence than the other head groups lipids, which is well correlated with the faster decrease in both $N_{\mathrm{HB}}^{\mathrm{inter}}$ of PS (Figure 15a) and the stability of the PS lipid complex network (Figure 16d with a decrease in cluster size from 8 to 5.5 going from 30 to 80 °C). At 80 °C, the τ_2 of PS lipids is very close to the other lipids. Despite the paucity of the inter-lipid hydrogen bonding ability, the PC lipids do not have the slowest τ_2, indicating that the electrostatic interactions and head group size may also affect τ_2 in addition to the hydrogen bonds. In fact, PA lipids show the lowest τ_2 due to its smallest size as well as the deficiency in the inter-lipid hydrogen bonding.

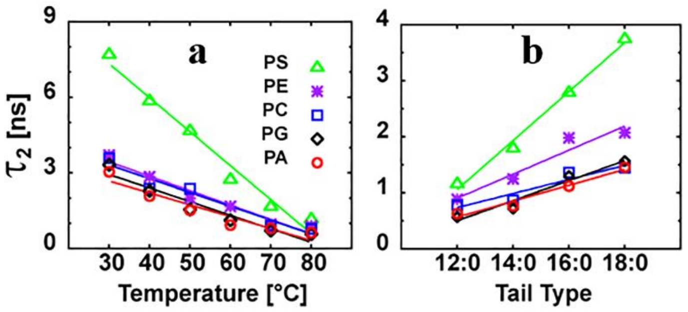

Figure 17 The of second rank reorientational correlation time τ_2:
a) DL lipids at different temperatures of various head group types and b) different chain lengths at 80 °C of various head group types.

Tail type dependence As shown in Figure 17b, τ_2 increases as the chain length increases. Compared to other lipid types, the faster increase in the τ_2 of PS lipids again correlates with the faster increase in $N_{\text{HB}}^{\text{inter}}$. Since $N_{\text{HB}}^{\text{inter}}$ of PA, PE, and PG lipids show negligible increase with chain length (Figure A.12a), the increase in the τ_2 of these lipids is due to the increasing van der Waals interactions of the acyl tails. PE lipids generally have higher τ_2 than PA, PC, and PG lipids. Moreover, the τ_2 of the unsaturated PS lipids are not always higher than the other head group types (Figure A.12, Table A.31, and Table A.32). The τ_2 of PO lipids are smaller than PL at 30 °C, while are greater at 40 °C (Table A.32). The τ_2 of SO is higher than SL for PE lipids probably due to decreasing $N_{\text{HB}}^{\text{inter}}$, while it is lower for PC and PS lipids. Based on these, we propose that there may be a threshold of the interaction strength that affects lipid mobility. When the combined inter-lipid interaction reaches a certain level, the lipid wobble slows down significantly.

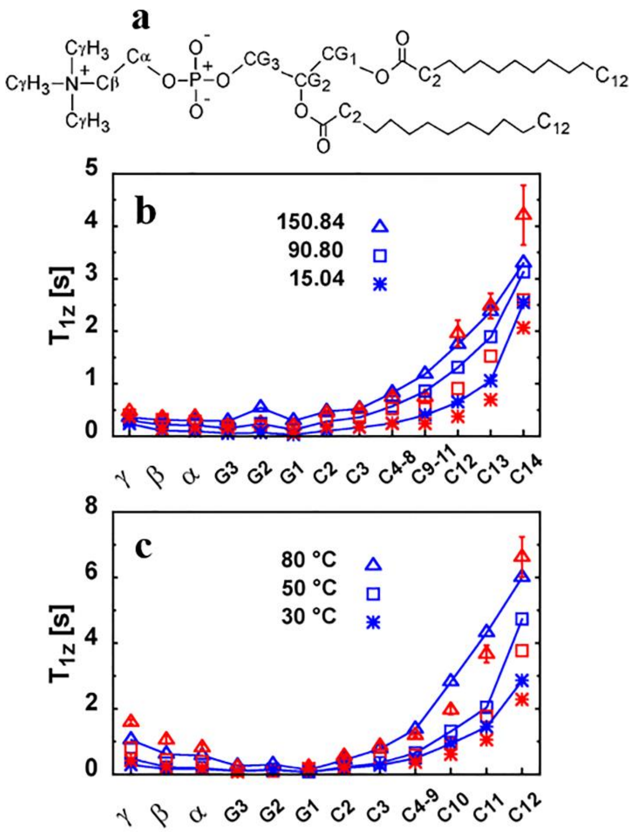

Figure 18 DLPC structures and Spin-lattice relaxation time T_1 of PC lipids:
a) The chemical structure of DLPC shown with some carbon atom names. Comparison of simulated relaxation time T_1 to experimental data [11]. b) DMPC with varied carbon frequency at 30 °C and c) DLPC with ω_C = 45.29 MHz at different temperatures. The blue represents the simulation data, and the red represents the experimental data. The standard errors of MD-based T_1 and the experimental uncertainty at low frequencies and temperatures are less than the symbol size.

Comparison of T_1 to experimental data Our final validation of the C36FF is to examine dynamical motions of lipid carbons, as quantified by NMR T_1. The carbon atom names for which T_1 was calculated are shown in Figure 18a. As the carbon Larmor frequency (ω_C) increases, T_1 of DMPC increases (Figure 18b). The simulated T_1 agree well with the experiment from Cγ to C$_8$, except for carbon G$_2$ with high ω_C (150.84 MHz), in which the MD-based T_1 is higher than experimental data. This deviation is also observed in the T_1 of DPPC (201.2 MHz) reported in [113]. Moreover, MD-based T_1 agree better with experiment at relatively high frequency (except C_{14}). The frequency dependence is also observed in DLPC (Figure A.13a-c), and the frequency dependence gets weaker as temperature increases. Figure 18c shows that T_1 increases as temperatures increases. Best agreement of DLPC occurs at an intermediate temperature (Figure 18c and Figure A.13). With the same frequency, T_1 of the equivalent positions have very weak chain length dependence (Figure A.13f). Overall, the T_1 obtained from the simulation show quite good agreement with experimental values.

4.2 United-atom CHARMM36 Force Field

4.2.1 Development

The united-atom CHARMM36 force field (C36FF-UA) of common lipids is developed to improve the simulation efficiency. The united atom topology files are generated by grouping together the hydrogen atoms and the carbon atom in methyl (CH3), methylene (CH2) and methine (CH) based on the all-atom CHARMM 36 force field (C36FF-AA).

The approach and the C36FF-UA of a few lipids have been developed and has been published previously in [125]. The goal of my study is to develop extended types of lipids, such as more saturated glycerol phosphate lipids of various head groups, saturated PC lipids and monounsaturated PE lipids with varied chain lengths, and also a few types of sphingomyelin and ceramides. A good portion of C36FF-UA results has been done by undergraduates which will be submitted for publication in the future. What presented here is only small section of this larger study which describes my work focus on this study.

The same force field parameters as C36FF-AA are applied to the non-grouped atoms of this lipid model in the hydrophilic head group region. The dihedral parameters involved the united carbon atom (CH1E, CH2E, and CH3E) may require obtaining new parameters by fitting the molecular mechanical energy (MME) to the quantum mechanical energy (QME). For instance, in sphingomyelins and ceramides, the fitting result for the dihedral parameter of CH2-CTL2-C-NHL is shown in Figure 19a. The dihedral involved double bond CEL1- CEL1-CTL2-CH2E require to fitting the two angles CEL1-CEL1-CTL2-CH2E and CEL1-CTL2- CH2E-CH3E simultaneously as shown in Figure 19b.

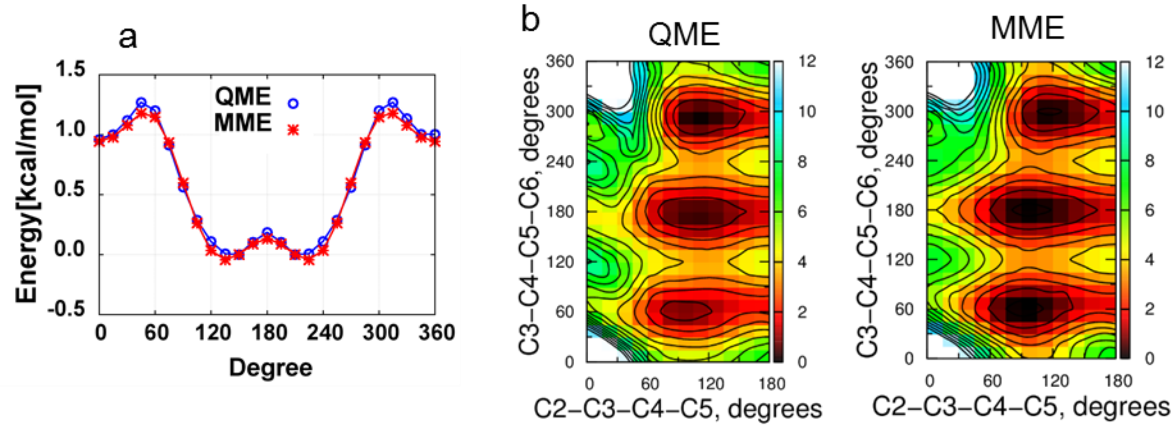

Figure 19 Comparison plots of the results of fitting MME on QME
The results of fitting MME on QME to obtain the dihedral parameters for angles such as a) CH2-CTL2-C-NHL and b) the dihedral involved double bond CEL1- CEL1-CTL2-CH2E

4.2.2 Test the accuracy

The accuracy of the C36FF-UA is analyzed by the bilayer properties comparisons between the simulated data with C36FF-UA and C36FF-AA lipid models, and also the comparisons between the simulated data obtained by C36FF-UA and the experimental data.

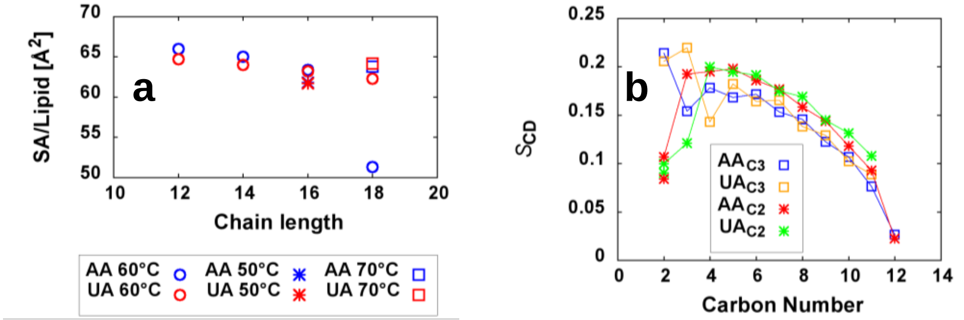

Figure 20 The SA/lip and S_{CD} for C36FF-AA and C36FF-UA.
The comparison between simulated all-atom (AA) and united-atom (UA) models of C36 force field for saturated PC lipids with the a) the surface area per lipid for saturated lipids with varied chain lengths at varied temperatures and with the b) the deuterium order parameter of DLPC at 60 ºC.

Figure 20a shows that the MD simulated SA/lipid from C36FF-UA and C36FF-AA tends to agree well for saturated PC lipid with longer chain length, and for shorter chains, the C36FF-UA-based SA/lipid is slightly smaller than the SA/lipid from C36FF-AA. The most deviation occurs for the lipid with the carbon chain length 18 (DSPC) at 60 ºC, at which the C36FF-AA suggests a gel-phase transition that cause a dramatic decrease in SA/lipid, while the C36FF-UA does not. The experimental SA/lipid of DSPC at 57 ºC is 67.5 Å^2 and its gel transition temperature is 54 ºC [126]. So in this case, the prediction from C36FF-AA is off while the C36FF-UA is more accurate. Figure 20b shows that except for major deviation in C3 and some deviation in C4 in chain *sn*-1 (chain 3), the S_{CD} from C36FF-UA and C36FF-AA agree well. The C36FF-UA models of other varied lipid types are also developed and the detail comparison will be presented in the future publication co-authored by two other undergraduate students.

Chapter 5 Results on Multiple-component Soybean Membrane Models

This chapter covers three topics, the pure ω-6 linoleoyl phosphate lipid bilayer, followed by the multiple-component soybean membrane models, and ends with the comparison of DLiPC in the pure and mixture hypocotyl membrane (Figure B.1). As shown in Table 3 and Table 4, besides two sterols, the main components of soybean hypocotyl and root membranes are the ω-6 linoleoyl phosphate lipids, which are a class of lipids that has not been well studied using computational techniques. Therefore, the pure membrane of two ω-6 lipids, 1-stearoyl-2-linoleoyl-sn-glycero-3-phosphocholine (SLPC, 18:0/18:2) and dilinoleylphosphatidylcholine (DLiPC, di-18:2) as shown in Figure 3, are studied and the results are presented in Section 5.1. As we mentioned, soybean is interesting to study mainly because it has plenty of ω-6 essential that is very important for human health. The MD simulations are performed on soybean hypocotyl and root membrane and results are presented in Section 5.2. DLiPC is one component in hypocotyl membrane, the comparison of DLiPC properties are demonstrated in Section 5.3.

5.1 Single-Component Membranes with Linoleic Chain(s)

The C36 lipid force field has been parameterized for monounsaturated and polyunsaturated (>4 double bonds) lipids [14, 127], but has not been tested against experimental data on dual-unsaturated tails such as linoleic acid. Therefore, MD simulations of SLPC and DLiPC were performed to compare with experimental data and provide a baseline to compare with our soybean membrane models. The equilibration of SA/lip was extremely fast (less than 20ns) as shown in Figure B.2. The average SA/lip for DLiPC was 4 Å^2 higher than SLPC with a single chain with dual-unsaturation (Table 5).

The K_A is slightly higher for SLPC compared to DLiPC (Table 5) suggesting that the presence of the 18:0 chain slightly rigidifies the membrane.

Table 5 System size and overall properties of lipid bilayers

Membrane	Lipid	T (K)	# Lipid	# Water	Time (ns)	SA/lip (Å^2)	K_A (N/m)
One-lipid	SLPC	303.15	80	2800	100	66.6 ± 0.2	0.30 ± 0.02
	DLiPC	303.15	80	2800	124	70.7 ± 0.2	0.26 ± 0.01
Hypocotyl		298.15	200	6000	150	52.7 ± 0.2	0.51 ± 0.07
Root		298.15	200	6000	150	51.9 ± 0.1	0.57 ± 0.06

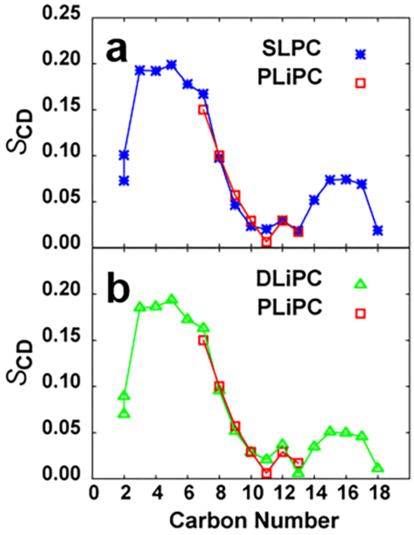

Figure 21 MD-based S_{CD} of SLPC and DLiPC compared to experimental values.
The comparison of the averaged deuterium order parameter (S_{CD}) of *sn*-2 chain for a) SLPC and b) DLiPC to experimental data PLiPC[128]. The double bonds at C6-7 and C9-10 are shifted to C9-10, C12-13 to match SLPC

Although there lack direct S_{CD} NMR measurements for DLiPC and SLPC, there does exist S_{CD} data on the isolinoleoyl (18:2$^{\Delta 6,9}$) with its double bonds shifted up from linoleoyl (18:2$^{\Delta 9,12}$)[128]. Considering this shifted the placement of the double bonds, the order parameters are in excellent agreement between the MD simulations and experiment (Figure 21). The S_{CD}s for the *sn*-2 chain for DLiPC are nearly identical to the *sn*-1 chain except for Carbon-2 (Figure B.2) and the expected trends are observed for the *sn*-1 chain of SLPC (Figure B.3).

There is published data on the ^{13}C NMR spin-lattice relaxation times for various carbons for DLiPC [129]. Since carbon assignment is not straightforward and estimated with NMR, Figure 22 shows the ranked order of the NMR T_1 values from low to high. The agreement between experiment and MD is excellent. Table B.1 contains the raw data for the NMR T_1s and also compares the assigned values for the T_1 from an experiment in comparison with the directly calculated values from MD simulations (Table B.2). Previous experimental values for C14/C15 were suggested to be 0.59 and 0.51, respectively but roughly have the value we obtained from our MD simulations for the C4-6/C7-8 position (Table B.2). Although MD simulations show a non-monotonic increase in T_1 from C2 to C18 it is less severe than that suggested previously by experiment [129].

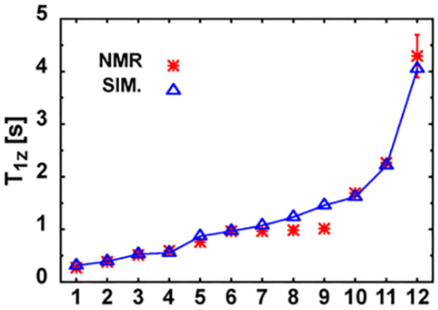

Figure 22 The relaxation time T_1 (s) of tail carbons of DLiPC at 30 °C. The carbon Larmor frequency ω_C is 90.80 MHz, and experimental data[129] is monotonically sorted from low to high due to assignment uncertainty.

The density profiles show slight changes between the SLPC and DLiPC bilayers (Figure 23). The main effect is in the thickness of these bilayers (Table 6) which is inversely related to the SA/lip in Table 5. The presence of two double bonds for each chain influences all measures of bilayer thickness (head group, water, and hydrocarbon densities). DLiPC bilayers have more density of unsaturated carbons at the center of the bilayer and the maximal peaks shifted toward the center as compared to SLPC. However, this has a minimal effect on the overall density in this hydrocarbon region (Figure 23a).

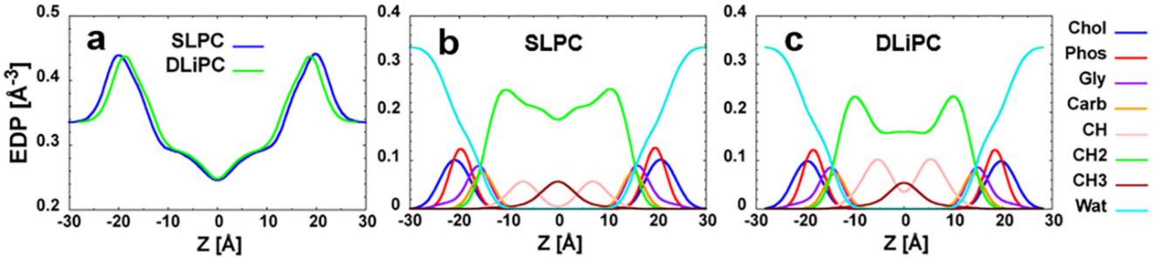

Figure 23 The electron density profile (EDP) of SLPC and DLiPC:
a) The comparison of the total EDP of SLPC and DLiPC b) The group EDP of SLPC and c) DLiPC, which includes choline (Chol), phosphate (Phos), glycerol (Gly), carbonyl (Carb), methine (CH), methylene (CH2), methyl (CH3), and water (Wat).

Table 6 The average thicknesses of bilayers (Å).

Thickness (Å)	SLPC	DLiPC	Hypo	Root
D_{HH}	39.9 ± 0.3	37.3 ± 0.2	42.7 ± 0.3	42.7 ± 0.1
D_{B}	38.9 ± 0.3	35.5 ± 0.03	40.1 ± 0.1	39.3 ± 0.1
$2D_{\mathrm{C}}$	29.1 ± 0.2	26.9 ± 0.03	32.0 ± 0.1	32.2 ± 0.04

5.2 Soy Membrane Models

Based on C36 MD simulations with SLPC and DLiPC, the lipid force field for polyunsaturated lipid tails [127] is in excellent agreement with limited experimental data. Therefore, our next step is to develop model membranes for the soy membranes of the hypocotyl and root to structurally quantify how these differ within the soybean plant and compare to the simple single-component membranes with two double bonds per chain.

5.2.1. Surface Areas and Area Compressibility

The SA/lip as a function of time (Figure B.2) demonstrates that the replicas reached equilibrium for this metric after 50ns. The average SA/lip was the smallest for the root membranes with 0.8 Å^2 lower than the hypocotyl (Table 5). This is the result of having slightly more sterols in the root membrane and containing PA lipids with their small head group and thus lower component SA/lip, which is evident from the component SA/lip (Table 7) with PLPA being several Å^2 lower than similar PLPC or PLPE lipids. Differences in the component SA/lip between the hypocotyl and root membranes for the same lipid show statistically identical values between these two model membranes, further demonstrating that sterols and PA lipids are influencing the change in the overall SA/lip. The K_A nearly doubled compared to the single-lipid membranes (Table 5) due to the elevated concentration of sterol.

Table 7 Component Surface Area per Lipid (SA/lip) for Soybean Membranes

Model	Lipid	SA/lip (Å^2)
Hypo	Sterols	30.6 ± 0.1
	PLPC	61.1 ± 0.4
	PLPE	60.6 ± 0.4
	PLPI	62.0 ± 0.5
	LLPE	60.2 ± 0.5
	DLiPC	62.3 ± 0.4
	DLiPE	60.3 ± 0.7
	DLiPI	60.7 ± 1.2
Root	Sterols	30.4 ± 0.2
	PLPA	58.9 ± 0.9
	PLPC	61.4 ± 0.2
	PLPE	60.6 ± 1.0
	PLPI	61.0 ± 0.6
	LLPA	60.4 ± 0.6
	LLPC	61.9 ± 0.1
	LLPE	60.7 ± 0.8
	LLPS	60.0 ± 0.7

5.2.2. Chain Order Parameters

The S_{CD} provides a way to compare with available NMR data (as above) and also to provide information on the acyl chain order in the lipid bilayer. The order parameters for most of the lipids were approximately the same between the hypocotyl and root membranes. However, for lipid LLPE (Figure 24), some of the order parameters on the C2 chain of the hypocotyl version were slightly higher than the equivalent values in the root LLPE. This was especially noticeable for carbons 3-7. The opposite was true for the PLPC lipids (Figure B.5a-b). Although the shape of the order parameters was more consistent between the hypocotyl and root, the root carbons tended to be slightly higher for both chains, indicating that the root membrane lipids are more ordered, which is consistent with the decrease in the overall SA/lip (Table 5).

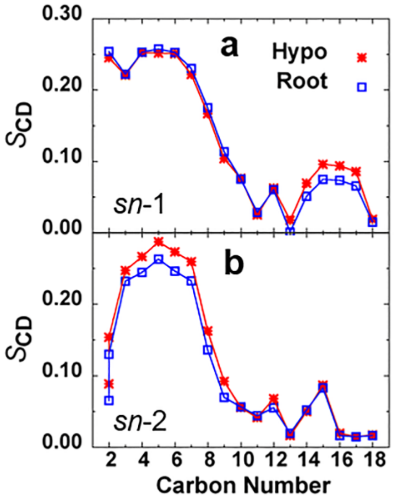

Figure 24 The comparison of S_{CD} of LLPE lipid in the soybean membranes: a) *sn*-1 chain and b) *sn*-2 chain of soybean hypocotyl and root membranes.

5.2.3. Tilt Angle Distributions

As shown in Table 8, comparing two sterols in each membrane model, the average tilt angle for sitosterol is slighter greater than stigmasterol (but within standard error) in hypocotyl membrane, whereas the root membrane models show sitosterol with a noticeable higher tilt angle than stigmasterol. The distribution of the tilt angles shows limited variation for each sterol between different soy membranes (Figure 25). The tilt angles of sitosterol are very similar in hypocotyl and root membranes. However, as shown in Figure 25, the stigmasterol has a slightly larger probability having tilts greater than 17.5^{o} in hypocotyl (0.45) than root (0.41) membrane model. This indicates that the stigmasterol is more upright in the root membrane compared to the hypocotyl model which agrees with slightly lower SA/lip of sterols in the root membranes. The difference of title angle distribution of stigmasterol between the same sterols in two membrane models may be the consequence of the hydrogen bonding or clustering effects, the analysis results of which will be presented in the later sections to understand the possible reason.

Table 8 Average Tilt Angle for Sterols in Soy Membranes

Type	Hypocotyl	Root
Sitosterol	17.9 ± 0.1	18.1 ± 0.3
Stigmasterol	17.6 ± 0.4	16.9 ± 0.4

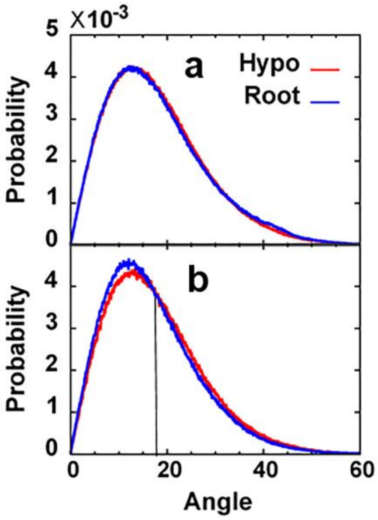

Figure 25 Tilt angle distributions of sterols in soy membrane.
a) sitosterol and b) stigmasterol with the vertical lines indicate the angle at 17.5°.

5.2.4. Density Profiles

The EDPs show the minor difference between the hypocotyl and root membrane models (Figure 26a). The general position and shape of the two models are nearly identical with between the models. The D_{HH} and D_C are statistically identical between the two soy membrane models (Table 6). The main overall difference is in the water measure for the bilayer thickness (D_B) with the root model having a thinner membrane by 0.8 Å. The replacement of PC/PE lipid with PA allows for more water penetration into the bilayer for the root membranes, consistent with single-component bilayers having the lower D_B compared membranes with other lipid head groups in our soy membrane models [109].

Secondary peaks at roughly ±10Å are the result of increased density due to the sterols (Figure 26c). This increased density is the maximum in the ring density for the sterols (Figure 26d and Figure B.6b). The EDP for sitosterol has the same shape in between the two soy membranes, whereas stigmasterol shows some variation between these membranes. The hydroxyl and ring carbons for stigmasterol show a slight positive shift from the center of the bilayer (Figure B.6b). This results in the tail density for the root membrane to have additional peaks that are shifted from the center of the bilayer. The more upright orientation of stigmasterol suggests this ordering influence the position of this sterol in the root membrane.

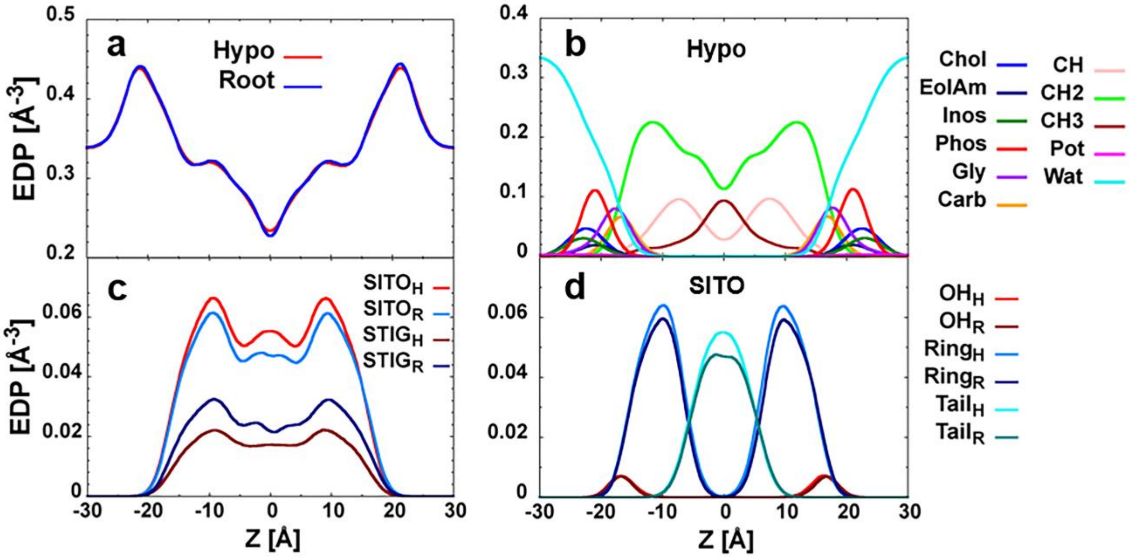

Figure 26 Total and group EDP of hypocotyl and root membranes:
a) The comparison of the total electron density profile of hypocotyl and root membranes b) The component electron density profiles of hypocotyl membrane, which includes choline (Chol), ethanolamine (EolAm), inositol (Inos), phosphate (Phos), glycerol (Gly), carbonyl (Carb), methine (CH), methylene (CH2), methyl (CH3), potassium ion (Pot), and water (Wat). c) Total EDP of sterols in hypocotyl and root membrane (SITO-sitosterol and STIG-stigmasterol) d) Group EDP of sitosterol, which includes OH, ring, and tail.

5.2.5. Lipid Hydrogen Bonding

The lateral ordering of lipids is promoted through either hydrogen bonding or electrostatic interactions. The sterols on average are hydrogen bonded to other lipids 24% of the time (Table B.3). Although PC lipid cannot form self-hydrogen bonds, these can act as acceptors to other lipids and for the hypocotyl and root membranes, there are 0.40 and 0.33 hydrogen bonds per lipid (Figure 27a and Table B.3). As expected the PE, PI and PS lipids have the highest occurrence of hydrogen bonding and PA intermediate between these and the PC lipids, which are similar in trends to what has been observed in single-component membranes [109]. Although several lipids can form intra-lipid hydrogen bonds, only the PI lipids have the significant formation of this class of hydrogen bonding at roughly 0.25 intra-lipid hydrogen bonds per lipid (Table B.3).

There are some statistical differences in hydrogen bonding between the hypocotyl and root membranes (Figure 27a and Table B.3). The common PE lipids (PLPE and LLPE) show a reduction in inter-lipid hydrogen bonds for the hypocotyl membrane. There is a reduction in the number of PE lipids in the root membrane that appears to slightly enhance the hydrogen bonding with PE lipids. The number of inter-lipid hydrogen bonds per lipid of sterols does not depend on membrane type as the glycerol phosphate lipids. Other common lipids and sterols show no statistical difference in hydrogen bonding between soy membranes. Moreover, comparing the sitosterol and stigmasterol in hypocotyl and root membranes, the lowest number of hydrogen bonds per lipid of stigmasterol in the root membrane (0.230± 0.0002) may explain the lowest titling angle among four (Table B.3). This suggests that hydrogen bonding is one of the causes of the titling angle difference of stigmasterol in the two membrane models.

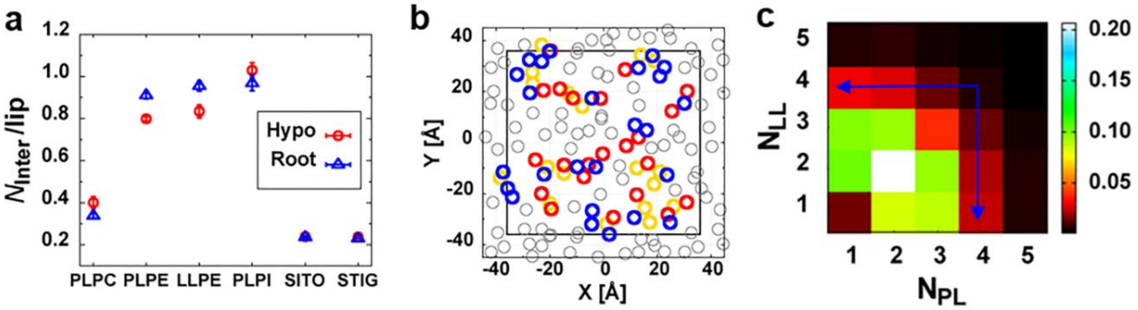

Figure 27 Hydrogen bonding and clustering in the soybean membranes:
a) Comparison of the number of inter-lipid hydrogen bonds of the common lipids in hypocotyl and root membrane. b) Schematic plot of clustering of root membrane. PL lipids are shown in red, LL are shown in blue, and sterols are shown in yellow c) Probability of the cluster composition of PL and LL lipids in root membrane. The blue arrows indicate increasing probability as composition difference of PL and LL lipids increases.

Table 9 Number of hydrogen bonds of PL and LL lipids
Number of hydrogen bonds per PL/LL lipid or per sterol of the root membrane, (comparing the equal number PA, PC, and PE head groups only).

Pairs	N_{HB}/lip
PL-PL	0.256 ± 0.019
LL-LL	0.264 ± 0.015
PL-LL	0.215 ± 0.023
SITO-PL	0.235 ± 0.025
SITO-LL	0.186 ± 0.016
STIG-PL	0.211 ± 0.034
STIG-LL	0.198 ± 0.023

As shown in Table 9, based on lipids with the equal number of the PA, PC and PE head groups in the root membrane, the overall number of hydrogen bonds per lipid between PL-PL (16:0/18:2) and LL-LL (18:2/18:3) are higher than PL-LL, suggesting that the lipids with the same unsaturation level tails prefer to form hydrogen bonds. The sterols prefer to form hydrogen bonds with PL than LL lipids. Among the two types of sterols, the more saturated sitosterol shows stronger preference with PL lipids.

5.2.6 Lipid Clustering

The number of hydrogen bonding describes the strength of the interaction of varied lipid components in the membrane which affects the normal functioning of biomembrane. The lipid clustering is also an important membrane structural property, especially for *E. coli* because it reflects the biological activities of cell-penetrating peptides (CPPs) and antimicrobial peptides (AMPs) [130, 131], and other cationic antimicrobial agents [132]. The MD simulation result also indicates that clustering is strongly correlated to the curvature of the complex lipid bilayer surface [133].

As we know, the interaction between the sterols and glycerol phosphate lipids are stronger than the interactions between glycerol phosphate lipids. As shown in Figure 27b, the existence of sterols in soybean membrane induces the formation of clusters that are “glued” by the sterols. The fraction of sterols forming a cluster (Y_C) is higher than glycerol phosphate lipids for both hypocotyl and root membranes. Sitosterol has a slightly lower tendency to form cluster than stigmasterol in hypocotyl membrane, while slightly higher in root membrane. The average fraction of lipid in the cluster (X_c) of PL lipids (0.286) higher than LL (0.269) in the root membrane suggests PL has a higher tendency to form a cluster, and this is also the case for hypocotyl membrane (Table B.4). Stigmasterol in the root membrane has lower Y_c than in hypocotyl membrane, meaning that slightly less stigmasterol in the root membrane form the clusters, which also may explain the low tilting angle (Table 8).

In the root membrane, PL and LL lipids have a same number of lipids with the same head group types (PA, PC, and PE), therefore, it is used for the further lipid type dependence comparison. The LL lipids have both tails unsaturated while PL lipids have only one tail unsaturated (Figure B.1). As shown in Figure 27c, the unsaturation level difference in the tail shows no effect in small clusters (N_{PL}<=2), in which the probability of (2, 2) cluster composition is higher than (1, 2) and (2, 1) implying that PL and LL lipids interactions have no preference. However, for the large cluster (N_{PL}>= 3), the equal lipid (diagonal) composition demonstrate the lowest probability, suggesting that PL and LL lipids prefer to form clustering with the same lipid tails due to stronger van der Waals interactions, which agrees with the above hydrogen bonding results. The cluster compositions of sitosterol with PL and LL lipids are slightly different while there is much less difference for stigmasterol (Figure B.7). The clusters with the number of PL or LL lipids greater than 3 is present in sitosterol, but absent in stigmasterol, implying that sitosterol interact with glycerol phosphate lipids stronger than stigmasterol.

5.3 Comparison of DLiPC in the Pure and Hypocotyl Membrane

The SA/lip of DLiPC in the hypocotyl membrane is 62.3 ± 0.4 Å^2 (Table 7), which is much lower than 70.7 ± 0.2 Å^2 (Table 5) in the pure membrane. The $2D_C$ considering DLiPC only in the hypocotyl membrane is 32.7 ± 0.2 Å, which is larger than 26.9 ± 0.03 Å (Table 6) in the pure membrane. These results suggest the tighter packing of hypocotyl mixture membrane than single-component DLiPC membrane. As shown in Figure 28a, the EDP of the hydrophilic head group region, interface, and tail region of the DLiPC in the hypocotyl membrane shift outward which agrees with the elongation of DLiPC in the mixture due to tighter packing.

In comparison, the S_{CD} of a single component membrane such as the DLiPC bilayer seems to resemble that of the hypocotyl membrane (Figure 28b-c). Moreover, the S_{CD} is categorically higher for the hypocotyl soybean membrane lipids than for the single component membrane lipids. This suggests that the presence of multiple lipids, including sterols, increased chain order in the soy membrane systems.

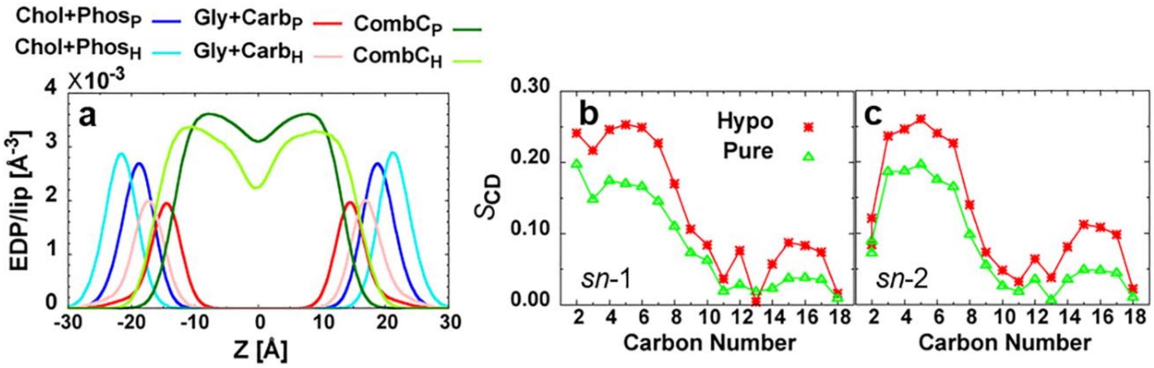

Figure 28 EDP per lipid of DLiPC in the pure and hypocotyl membrane:
a) Group EDP of the hydrophilic head group (Chol+Phos), interface region (Gly+Carb), and hydrophobic tail groups (CombC). b) The S_{CD} of DLiPC of hypocotyl and pure membranes in c) *sn*-1 and d) *sn*-2 chain.

Chapter 6 Results on Periplasmic-Open LacY Models

This Chapter covers two topics. Section 6.1 demonstrates the comparison of currently available periplasmic-open LacY models to verify the feasibility of the models. Section 6.2 present the simulation results from SGLD simulation with the method described in Section 2.3, which is the simulation method applied to obtain more unknown periplasmic-open LacY structures.

6.1 Test the Current Available Periplasmic-Open LacY Models

6.1.1 Three Periplasmic-Open LacY Models

Structural models haven been developed to represent the possible structure of periplasmic-open LacY, such as LacY$^{\text{IM-EX}}$ model [48], repeats-swapped (SW) and FucP-based (FP) LacY models [51]. The accuracy of periplasmic-open LacY$^{\text{IM-EX}}$ model obtained by Pendse et al. [48], and LacY$^{\text{SW}}$ and LacY$^{\text{FP}}$ models both obtained by Radestock et al. [51] are tested using MD simulations (with detail method protocol on page 20) in explicit membrane with sugar (ββ-(Gal*p*)$_2$) bound to the protein.

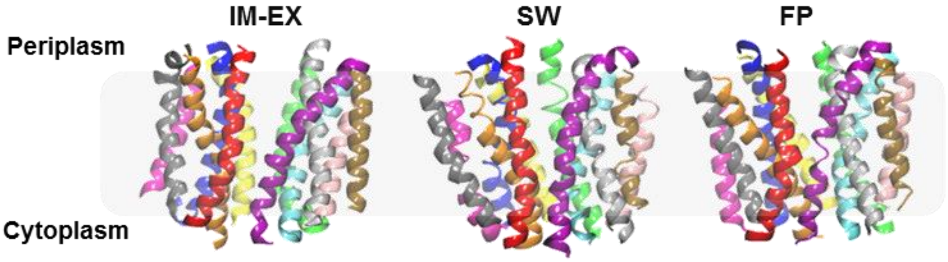

Figure 29 Side views of outward-facing LacY$^{\text{IM-EX}}$, LacY$^{\text{SW}}$, and LacY$^{\text{FP}}$ models.
The following color scheme is used for each helix: blue=H1, red=H2, gray=H3, orange=H4, yellow=H5, tan=H6, silver=H7, green=H8, pink=H9, cyan=H10, purple=H11, mauve=H12.

LacY$^{\text{IM-EX}}$ model (Figure 29, left) was obtained through a two-step hybrid implicit-explicit simulation method, in which SGLD simulations [73, 91] were performed with LacY in an implicit membrane, from which the partial periplasmic-open structures obtained were used in MD simulations in an explicit lipid bilayer to obtain the LacY structure that is fully open on the periplasmic side [48].

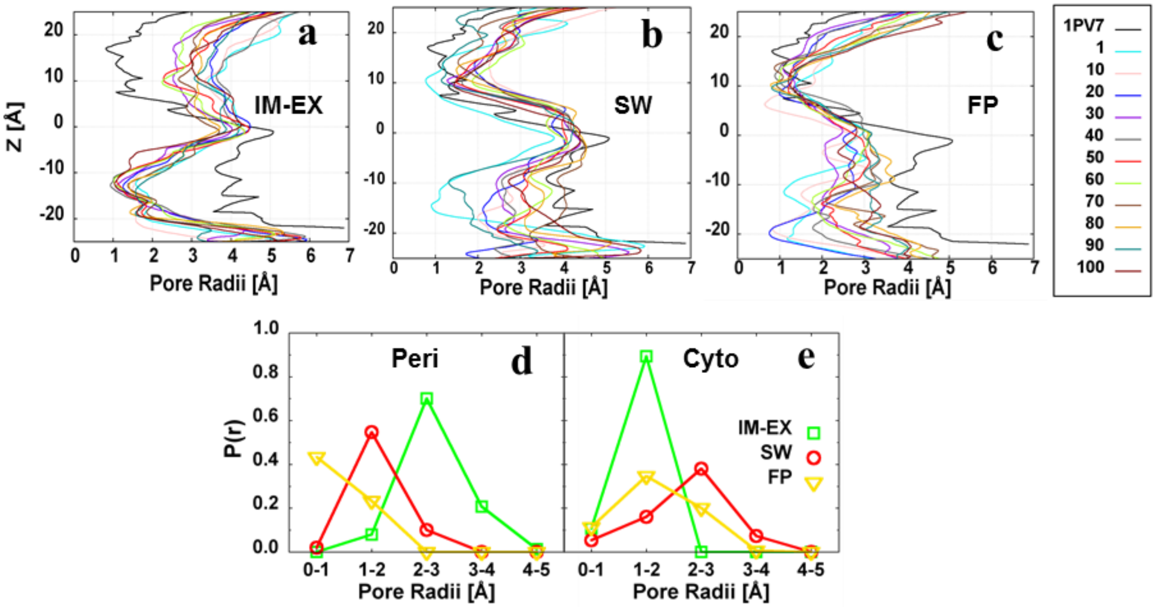

Figure 30 Pore radii of three periplasmic-open LacY models:
a-c) Pore radii of three periplasmic-open LacY models at varied simulation time. The black curve is the pore radii of the 1PV7 crystal structure. And the curves in difference colors are the pore radii of LacY$^{\text{IM-EX}}$ at time 1 ns to 100 ns with step 10 ns. d-e) The histogram of the probability of minimum pore radii on the periplasmic and cytoplasmic sides.

LacY$^{\text{SW}}$ model (Figure 29, middle) was obtained through swapping the inverted-topology repeats method, in which the helices was divided into A, B, C and D units with three helices in each unit. Since it is known that unit A and unit B are 2-fold rotational symmetric by axis that runs through the center and parallel to the membrane plane, so the cytoplasmic and periplasmic side of topology repeats units A and D were firstly inverted, then the conformations of the repeats were swapped by modeling the conformation of A using the structure of B as a template and vice versa, similarly for unit C and D [51].

LacY^{FP} model (Figure 29, right) was built by apply LacY sequence on the available outward-facing crystal structure of another *E. coli* MFS transporter, L-fucose permease, which is L-fucose/proton symporter that shares about 10% sequence identity to LacY only [51]. LacY^{IM-EX} structure is obtained from simulation predicted the natural movement of LacY, while LacY^{SW} and LacY^{FP} models are two built theoretical models. Therefore, the accuracy of these models needs to be verified through MD simulations.

6.1.2 Comparison of Models with Regular Equilibrium

The pore radii comparison between the simulation results of the three periplasmic-open LacY models (Figure 30a-c) shows that LacY^{IM-EX} model remains stable in the periplasmic-open state, while the LacY^{SW} and LacY^{FP} close on the periplasmic side and become unstable on the cytoplasmic side. The probability histogram demonstrates that on the periplasmic side, LacY^{IM-EX} shows the highest probability of opening, while the LacY^{FP} shows the lowest probability of opening (Figure 30d), while on the cytoplasmic sides, LacY^{IM-EX} shows the much higher probability of closing than the other two models. Therefore, the pore radii results suggest that the LacY^{IM-EX} is a better LacY model than the other two.

To visualize the structure change, the structure at the end of simulation compared to its initial model structure is shown in Figure 31a, which demonstrates that LacY^{IM-EX} does not have a broad conformational change and remains in the outward-facing state, while the LacY^{SW} and LacY^{FP} models show closing on the periplasmic side. The pore radii results at Figure 31b shows that the closure of pore at the periplasmic side of LacY^{SW} and LacY^{FP} occur at the end of equilibrium.

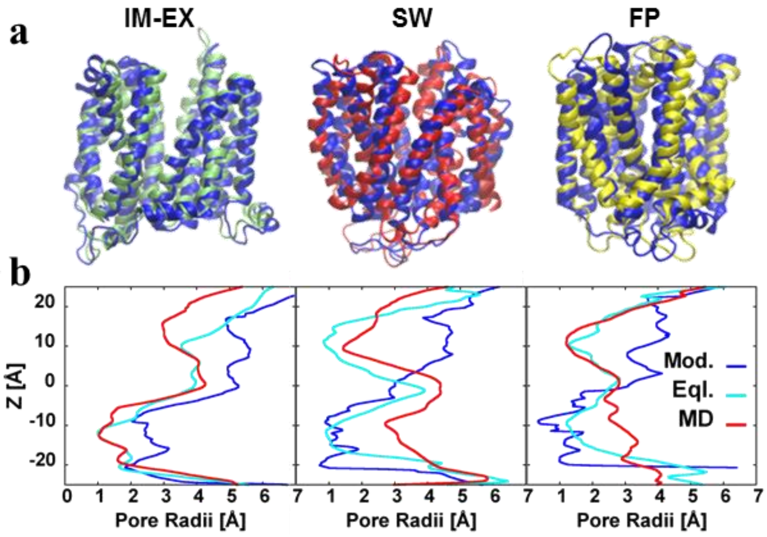

Figure 31 Snap shots and pore radii of three outward-facing LacY models:
a)Snap shots of LacY$^{IM\text{-}EX}$ (green), LacY^{SW} (red), and LacY^{FP} (yellow) LacY models at the end of simulation compared to its initial model structure b) Pore radii of the three LacY models at the initial model, the end of equilibrium and the end of simulation runs.

6.1.3 Comparison of Models with Extended 10-ns Equilibrium

Table 10 Steps for two types of equilibriums.
Equilibrium steps for (regular) standard CHARMM-GUI six step equilibrium and extended extra 10-ns equilibrium. FC_{BB} , FC_{SC} are Harmonic restraint force constants for backbone and side chain, respectively, in unit kcal·mol^{-1}·Å^{-2}

Type	Steps	Time (ns)	FC_{SC}	FC_{BB}
Regular	6.1	0.035	5	10
	6.2	0.025	2.5	10
	6.3	0.025	1	10
	6.4	0.2	0.5	10
	6.5	0.2	0.1	10
	6.6	0.2	0	10
Extra	7	5	0	10
	8.1	1	0	5
	8.2	1	0	2.5
	8.3	1	0	1
	8.4	1	0	0.5
	8.5	1	0	0.1

To make sure the closure of the LacY^{SW} and LacY^{FP} are not merely due to short equilibrium process (0.685 ns standard six-step CHARMM-GUI equilibrium), additional 10-ns equilibrium is applied. As shown in Table 10, after 0.685 ns regular equilibration, the additional 5-ns equilibration with backbone Harmonic restraint force constant of 10 kcal·mol^{-1}·Å^{-2} is applied, and then the restraint is gradually reduced from 10 to 0.1 kcal·mol^{-1}·Å^{-2} during another 5-ns equilibration.

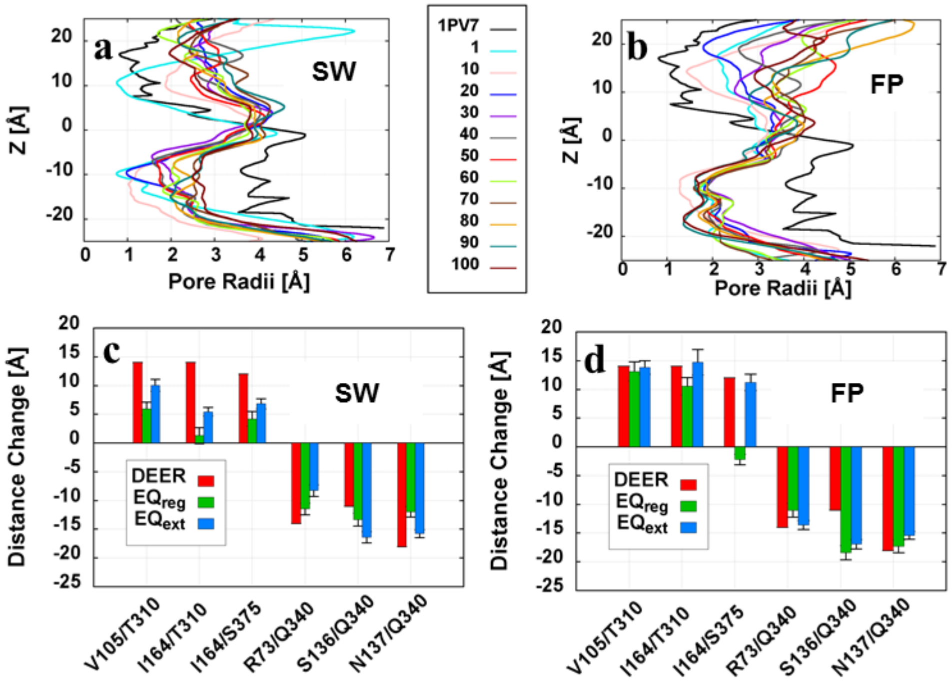

Figure 32 The extended 10 ns simulation results:
a-b) The extended 10 ns equilibrium resulted pore radii of a) LacY^{SW} and b) LacY^{FP} models, and residue distance change comparison between the regular and extended equilibrium of c) LacY^{SW} and d) LacY^{FP} models.

Figure 32a-b shows that when extra 10 ns restrained equilibrations are applied, LacY^{SW} become slightly more stable and LacY^{FP} remains stable in the closed cytoplasmic side and partially periplasmic-open state. The simulated residue distance change of LacY^{FP} structure agrees well with DEER (Figure 32c-d). However, LacY^{FP} model is merely the LacY homology structure model which is not closely related to LacY. Also, it still has higher RMSD relative to the initial LacY^{FP} model compared to LacY$^{IM\text{-}EX}$ for all protein backbone and also all helices (Figure 33).

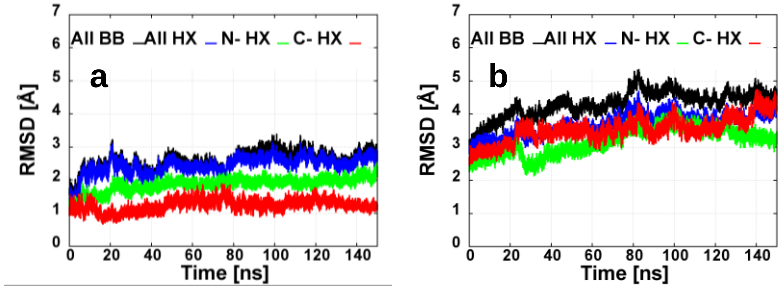

Figure 33 The backbone RMSD relative to their initial model structures:
a) LacY$^{IM\text{-}EX}$ and b) LacY^{FP} relative to their initial model structures during MD simulations (with extra 10ns restrained equilibration) of all protein (ALL BB), all helices (ALL HX), N-domain helices (N-HX) and C-domain helices (C-HX).

6.2 Search More Periplasmic-Open LacY Structures

Though the current analysis results indicate that LacY$^{IM\text{-}EX}$ may be a reasonably good model, its conformation does not fully agree with the DEER results [62]. SGLD simulation method is applied to search more periplasmic-open LacY structures, which hopefully can lead to a structure that can agree better with DEER.

SGLD simulations (with detail method protocol provided on page 25) of inward-facing LacY are performed to obtain LacY in different conformational states. The SGLD simulation can accelerate the conformational change while preserves the conformational distribution well when the guiding parameters are within the suggested range (guiding factor less than 1 is recommended for avoiding much deviation from canonical ensemble) [91]. Therefore, it is applied to enhance searching efficiency of LacY in different conformational states without the loss of the conformational accuracy [73, 91, 92].

The SGLD simulations can be run with the implicit membrane and recently with an explicit membrane in the NVT ensemble. The wild-type inward-facing LacY crystal structure 2V8N [52] (Figure 5) and pre-equilibrated C154G mutated 1PV7 [53] is used in the SGLD simulation in the implicit membrane. The pre-equilibrated inward-facing 1PV7 [53] is also used in the new SGLD simulation in the explicit membrane. The SGLD parameter sets with 2V8N [52] and pre-equilibrated 1PV7 [53] with the full residue sequence 417 are shown in Table 11 and Table 12. Together with the cut sequence (402 or 399) of 1PV7 and the occluded structure, there is total 200 SGLD simulation runs in the implicit membrane and 8 runs in the explicit membrane have been performed. The detail of each parameter will be described in Section 6.2.1.

Two implicit MD simulations are performed on inward-facing LacY 1PV7 [53] to test the effect of the implicit membrane in the SGLD simulations. As shown in Figure 34, pore radii show that the implicit membrane has no effect on the conformational change of LacY. LacY remains inward-facing states in both replicas, and there is no residue distance change relative to the initial structure at the end of the simulation.

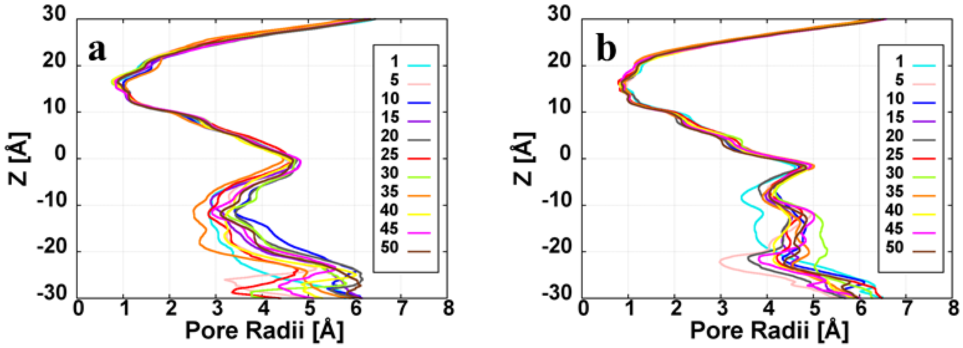

Figure 34 Pore radii of MD simulations in the implicit membrane

Table 11 Parameter setup SGLD simulation with 2V8N with 417 residues.
The beta- galactoside (bgal) is sugar ββ-(Galp)$_2$, which tends to induce periplasmic-open LacY. The residues E269, H322, and E325 are three residues involved in proton translocation of LacY.

Binding	Protonation	TSGAVG	SGFT	FBETA	# Replicas
apo	unp	0.25	1.0	1.00	1
apo	269	0.25	1.0	1.00	1
apo	322	0.25	1.0	1.00	1
apo	325	0.25	1.0	1.00	1
bgal	unp	0.25	0.5	0.10	1
bgal	unp	0.25	0.5	0.25	1
bgal	unp	0.25	0.5	0.50	1
bgal	unp	0.25	0.5	1.00	5
bgal	unp	0.25	0.5	2.00	1
bgal	unp	0.25	0.5	5.00	1
bgal	unp	0.25	1.0	1.00	5
bgal	269	0.25	0.5	0.10	1
bgal	269	0.25	0.5	0.25	1
bgal	269	0.25	0.5	0.50	1
bgal	269	0.25	0.5	0.50	1
bgal	269	0.25	0.5	1.00	4
bgal	269	0.25	0.5	2.00	1
bgal	269	0.25	0.5	5.00	1
bgal	269	0.25	1.0	1.00	5
bgal	322	0.25	0.5	0.10	1
bgal	322	0.25	0.5	0.25	1
bgal	322	0.25	0.5	0.50	1
bgal	322	0.25	0.5	0.50	1
bgal	322	0.25	0.5	1.00	4
bgal	322	0.25	0.5	2.00	1
bgal	322	0.25	0.5	5.00	1
bgal	322	0.25	1.0	1.00	5
bgal	325	0.25	0.5	0.10	1
bgal	325	0.25	0.5	0.25	1
bgal	325	0.25	0.5	0.50	1
bgal	325	0.25	0.5	0.50	1
bgal	325	0.25	0.5	1.00	4
bgal	325	0.25	0.5	2.00	1
bgal	325	0.25	0.5	5.00	1
bgal	325	0.25	1.0	1.00	5
bgal	unp	0.25	0.5	1.00	1
bgal	269	0.25	0.5	1.00	1
bgal	322	0.25	0.5	1.00	1
bgal	325	0.25	0.5	1.00	1

Table 12 Parameter setup SGLD simulation with 1PV7 with 417 residues.
The beta- galactoside (bgal) is sugar ββ-(Galp)$_2$, which tends to induce periplasmic-open LacY. The residues E269, H322, and E325 are three residues involved in proton translocation of LacY.

Binding	Protonation	TSGAVG	SGFT	FBETA	# Replicas
apo	unp	0.20	0.75	4.00	1
apo	unp	0.20	0.75	5.00	6
apo	unp	0.20	0.75	6.00	1
apo	unp	0.20	0.80	5.00	1
apo	269	0.20	0.75	5.00	1
apo	322	0.20	0.75	5.00	1
apo	325	0.17	0.75	5.00	1
apo	325	0.18	0.75	5.00	1
apo	325	0.19	0.75	5.00	1
apo	325	0.20	0.70	5.00	1
apo	325	0.20	0.75	0.50	1
apo	325	0.20	0.75	0.75	1
apo	325	0.20	0.75	1.00	1
apo	325	0.20	0.75	2.50	1
apo	325	0.20	0.75	4.00	1
apo	325	0.20	0.75	5.00	6
apo	325	0.20	0.75	6.00	1
apo	325	0.20	0.75	7.00	1
apo	325	0.21	0.75	5.00	1
apo	325	0.22	0.75	5.00	1
apo	325	0.23	0.75	5.00	1
apo	325	0.24	0.75	5.00	1
bgal	269	0.20	0.75	5.00	1
bgal	269	0.25	0.50	0.50	1
bgal	269	0.25	0.50	1.00	1
bgal	269	0.25	0.75	1.00	3
bgal	269	0.25	1.00	1.00	5
bgal	269	0.25	1.04	1.00	1
bgal	269	0.25	1.06	1.00	1
bgal	322	0.17	0.75	5.00	1
bgal	322	0.18	0.75	5.00	1
bgal	322	0.19	0.75	5.00	1
bgal	322	0.20	0.75	5.00	5
bgal	322	0.21	0.75	5.00	1
bgal	322	0.22	0.75	5.00	1
bgal	322	0.23	0.75	5.00	1
bgal	322	0.25	0.50	0.50	1
bgal	322	0.25	0.50	1.00	1

bgal	322	0.25	0.50	15.0	1
bgal	322	0.25	0.75	1.00	1
bgal	322	0.25	0.75	5.00	1
bgal	322	0.25	1.00	1.00	1
bgal	322	0.30	0.75	5.00	1
bgal	322	0.30	0.75	7.50	1
bgal	322	0.50	0.75	7.50	1
bgal	325	0.20	0.75	5.00	1
bgal	325	0.25	0.50	0.50	1
bgal	325	0.25	0.50	1.00	1
bgal	325	0.25	0.75	1.00	1
bgal	325	0.25	1.00	1.00	1

6.2.1 SGLD with the Implicit Membrane

6.2.1a NS-Averaged

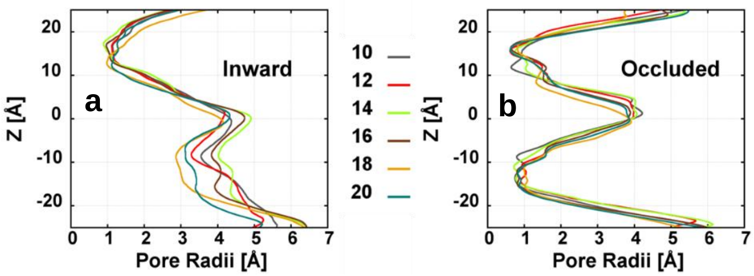

Figure 35 Exemplary pore radii resulted from implicit SGLD simulations.
LacY with Glu269 protonated, SGLD with implicit membrane from equilibrated six time frames (ns) that result in the a) inward-facing state and b) occluded state.

When we consider the ns-averaged pore radii as the target, for SGLD with the implicit membrane (1~5 replica per setting, with varied SGFT and FBETA), the pore radii suggest that LacY either remain inward-facing (Figure 35a) or changes to an occluded state (Figure 35b).

Guiding Factor The degree of guiding effect in a SGLD simulation can be adjusted by varying the guiding factor (SGFT). The stronger guiding force [73] will be introduced and the larger energy barrier will be overcome when the SGFT is larger. It was believed that the reason that the periplasmic-open LacY structure could not be obtained by experimental crystallization is due to its higher free-energy state [57]. Therefore, my work in SGLD simulations focused on varying the SGFT (with the range 0.5 to 1.25) in order to overcome the energy barrier for the outward-facing state.

Friction Constant The friction constant (FBETA, collisions per picosecond) is also called collision frequency. The higher FBETA, the higher guiding force as well as the frictional dragging force [73]. Besides the energy barrier effect, the diffusion effect is also considered when selecting the value of FBETA. At low FBETA, the larger FBETA, the higher transition rate since the energy barrier effect dominant. The higher guiding force helps the transition more than the slow down by the higher frictional force. At high FBETA, the diffusion effect dominates the transition rate. The larger FBETA, the lower transition rate due to higher friction. The transition rate reaches a maximum when the energy barrier effect is compensated by the diffusion effect [73]. Meantime, to avoid applying too high guiding force, the lower SGFT should be used when higher FBETA is applied. The varied values of FBETA in the range between 0.5 and 25 ps^{-1} have been applied.

Local average Time The parameter local average time (TSGAVG) in ps sets the boundary of the time scale of motions to enhance. All motions with time scale or periods greater than TSGAVG will be enhanced. Therefore, a greater TSGAVG will result in a slower motion to be enhanced. The TSGAVG of 0.25ps is applied in most SGLD simulations. The varied values of TSGAVG in the range between 0.19 and 0.25 ps have been applied to study its effect on the conformational change.

Initial Structure-Protein Conformational States Besides adjusting parameters, different starting structures were used. It is possible that due to time scale limitation or unable to overcome a further energy barrier, once the intermediate occluded state is formed, it is difficult to observe any further conformational change that can lead to an outward-facing state. Therefore, instead of inward-facing LacY, we used *apo*- and sugar bounded occluded and even a few unphysical LacY structures as the starting structures, which were already obtained from quite a few from previous SGLD simulations, expecting that it can lead the conformational change following transition step 1 or 4 that results in periplasmic-open LacY (Figure 4).

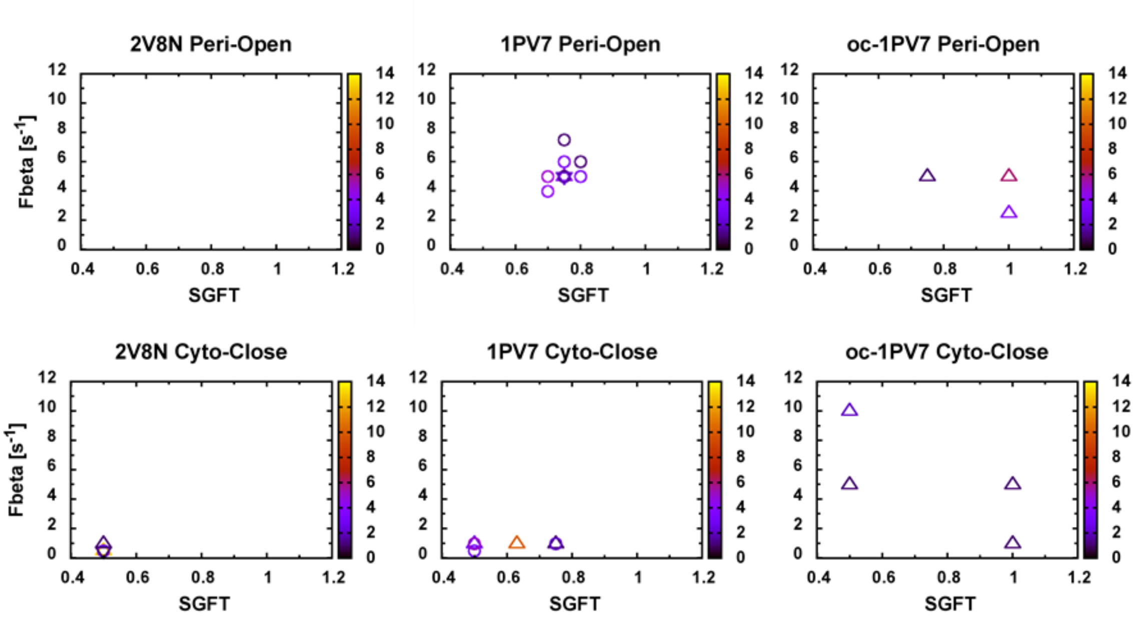

Figure 36 The frequency of minimum pore radii of LacY resulted from SGLD simulations. The varied guiding factor SGFT and friction constant FBETA resulted in periplasmic-open and cytoplasmic-close structures with initial structures inward-facing 2V8N and 1PV7, and the occluded state obtained from 1PV7 for Glu269 (up-triangle), His322 (circle), and Glu325 (down-triangle) protonated states.

Initial Structure-Protonation States Since LacY is a H/galactoside symporter. LacY conformation may depend on the proton translocation, which is presented by the protonation states of the relevant residues. Therefore, SGLD simulations are performed on the Glu269-, and H322-, and Glu325- protonated LacY.

Initial Structure-Protein Sequence Cut The C-terminus loop 400-417 is a relatively free loop that tends to expand more widely and may interact with its own protein image due to the periodic boundary. Therefore, to avoid this affect, in some conformation, LacY is cut at 399 or 402, and the tail residues are not included in the simulations.

Figure 36 shows that 2V8N does not result in any periplasmic-open state; 1PV7 results in some periplasmic-open and cytoplasmic-close structures, however, there is no common friction constant FBETA that can have both parts valid simultaneously. The occluded 1PV7 with guiding factor 0.75 and friction constant 5 ps^{-1} satisfy the outward-facing pore radii. However, there are two issues. Since the guiding force applied is quite strong, there are unphysical structures which have both the periplasmic and cytoplasmic sides open (Figure 37a). Moreover, the C_{α} distance of this structure disagrees with the DEER residue pair distance. More importantly, based on the VMD as shown in Figure 37b, these structures are incorrect as the center of the protein pull apart very far way, implying that the strong guiding force may have broken the hydrophobic force of the implicit membrane, therefore destabilizes the whole protein.

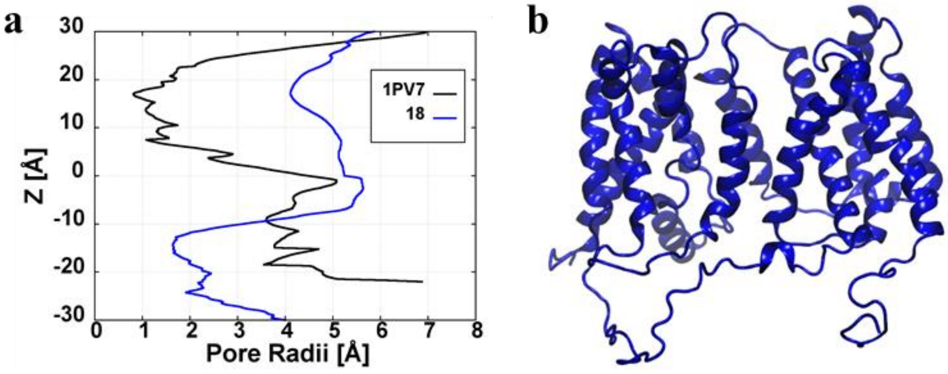

Figure 37 Unphysical LacY structure showing opening on both sides:
a) The pore radii of the periplamic-open 322-protnanted LacY structure with guiding factor 0.75 and friction constant 5 ps^{-1} b) The snap shot of LacY showing the opening on the both peri- and cyto-plasmic sides and the between center region.

Part of the Structure Applying Guiding In most SGLD simulations, the guiding force (i.e. the friction constant FBETA) is applied to the entire LacY structure. To try to obtain the periplamic-open LacY structure in a more robust method, we also tried to apply the guiding force on only the part of the LacY, such as N-domain, C-domain, or periplasmic side, cytoplasmic side. Figure 38 shows that with entire protein guided, as the SGFT increases and FBETA decreases, LacY open on the periplasmic side as well as cytoplasmic side (which may be caused by hydrophobic interaction on the protein-water interface), while with partially guided structure, LacY closes on the cytoplasmic side, and slightly or nearly no opening on periplasmic side. Even though no periplamic-open structure is observed, with the partially guiding, the overall LacY structure seems more stable. A similar result is shown in the N-domain guided simulation. When guiding periplasmic side on the occluded structure, and guiding cytoplasmic side on the both-open unphysical structure, the periplasmic-open conformation is not obtained, as the guided half does not show much conformational change. From these, we conclude that there are strong interior correlations between the N- and C-domain, and peri- and cyto-plasmic sides of LacY, so that we can not only change part of the conformation of LacY. In order to obtain the periplamic-open LacY, the movements of different parts of LacY communicate or correlate with each other, and the conformational change of the whole LacY is required.

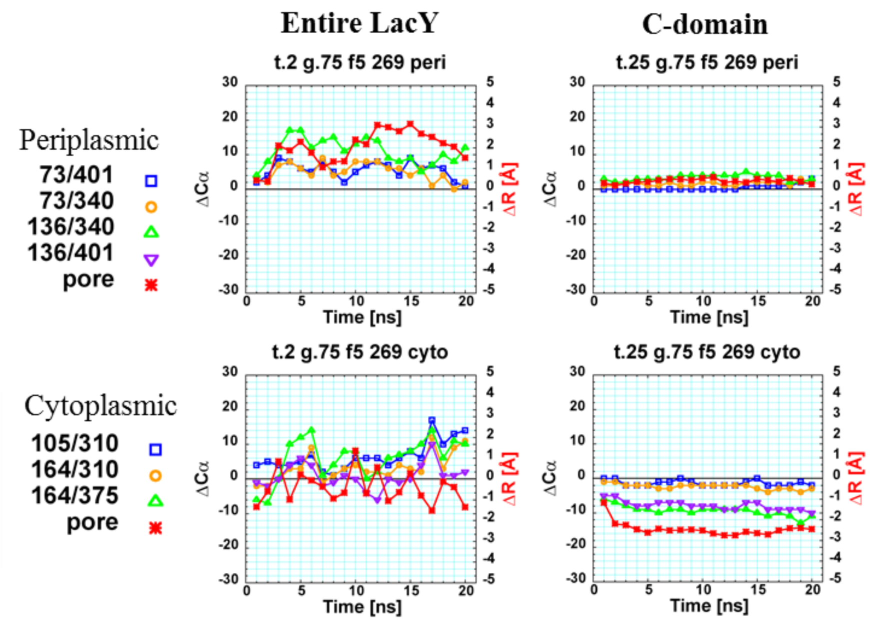

Figure 38 The residue distance and minimum pore radii change
At lower SGFT and higher FBETA of entire protein guided (which is the case for all unless specified) and partially protein guided.

6.2.2b Individual Frame

Based on more statistical ns-averaged results, which is the average of 1000 frames, there is no LacY structure resulted from the SGLD that can result in reasonable pore radii and residue distance change that show opening on the periplasmic side, and closing on the cytoplasmic side. However, the pore radii fluctuate significantly in this timeframe. So the low occurrence of the outward-facing structure may not be captured from the 1000-frame average. However, we can observe its existence if we look at each individual frames. Specifically, we see some reasonably good outward-facing LacY structures.

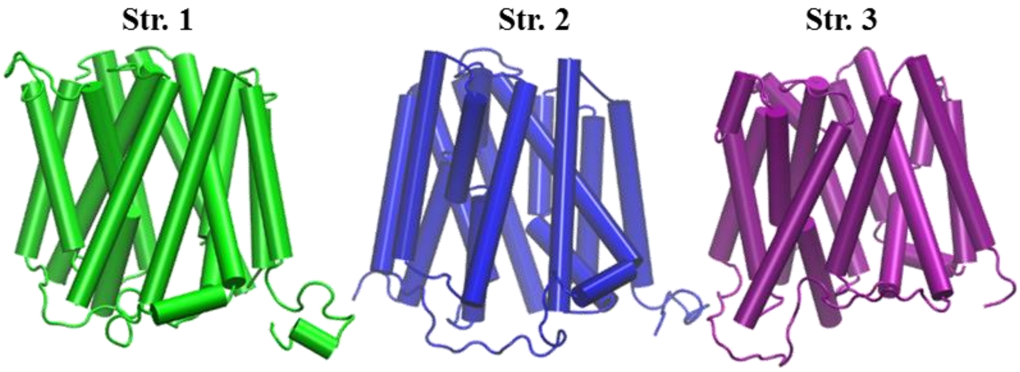

Figure 39 Snapshots of the outward-facing LacY from implicit SGLD simulations. Structure 1 is obtained from Glu269-protonated LacY with guiding factor 0.75, friction constant 1 ps^{-1}; Structure 2 is from Glu269-protonated LacY with guiding factor 1.04, friction constant 1 ps^{-1}; Structure 3 is from His322-protonated with guiding factor 0.75, friction constant 5 ps^{-1}.

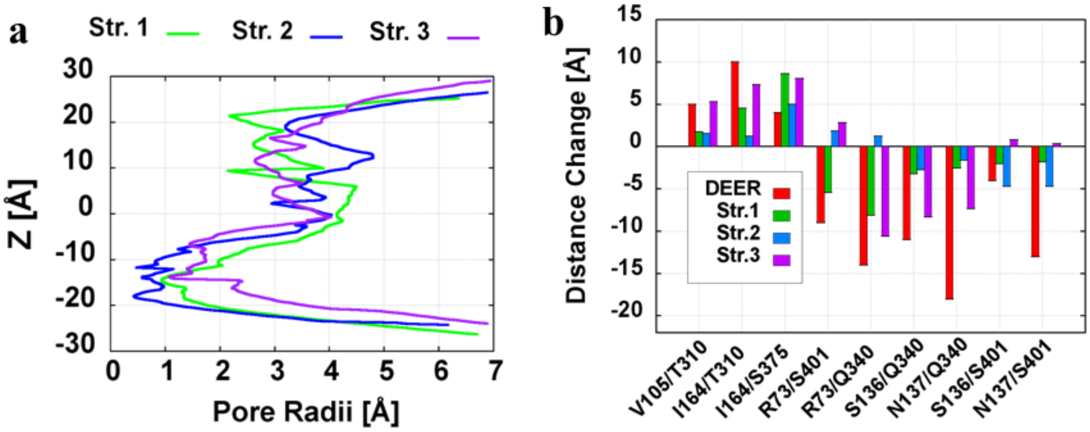

Figure 40 Structure analyses of outward-facing LacY from implicit SGLD simulations They are starting from pre-equilibrated 1PV7 with reasonable a) pore radii and b) Residue pair distance changes based on C$_\alpha$ compared to DEER data [62].

The snapshots of the obtained periplasmic-open LacY structures are shown in Figure 39. The pore radii show that Structure 1 and 3 are not fully closed on the cytoplasmic side (Figure 40a). And the C$_\alpha$ of residue pair distance change demonstrates that all the three structures show reasonable opening on the periplasmic side, but have less closing on the cytoplasmic side compared to DEER data [62] (Figure 40b). Based on the C$_\alpha$ of residue pair distance Structure 3 has the most cytoplasmic closure with Structure 2 having little closure compared to the cytoplasmic-open state.

In order to obtain fully equilibrated outward-facing LacY structure that closes more on the cytoplasmic sides, MD simulations are performed on these structures in the explicit membrane for 100 ns with sugar ββ-(Galp)$_2$. As show in Figure 41, on the periplasmic side, the pore radii of Structure 1 and 2 LacY close, and the Structure 3 partially close. On the cytoplasmic side, Structure 1 and 3 close, and Structure 2 partially close. The residue pair distance results in Figure 42 demonstrate that neither of the LacY structure agrees with the DEER experiment [62]. These results suggest that MD simulation with the explicit membrane cannot cause LacY close on the cytoplasmic side.

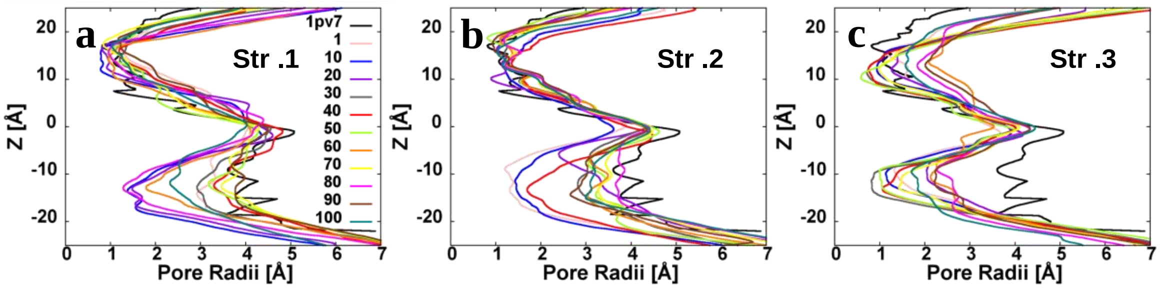

Figure 41 Explicit MD simulation resulted pore radii.
Three outward-facing LacY structures initially obtained from the implicit SGLD simulations.

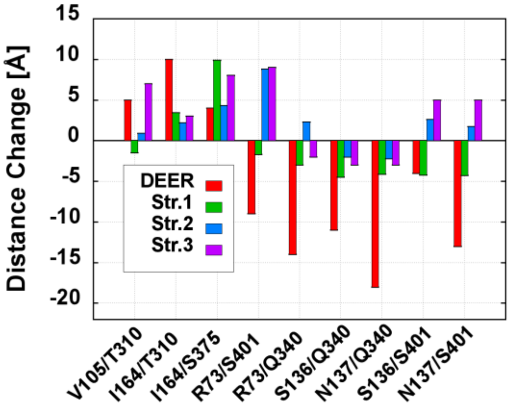

Figure 42 Explicit MD simulation resulted residue pair distances.
Three outward-facing LacY structures initially obtained from the implicit SGLD simulations.

6.2.2 SGLD with the Explicit Membrane

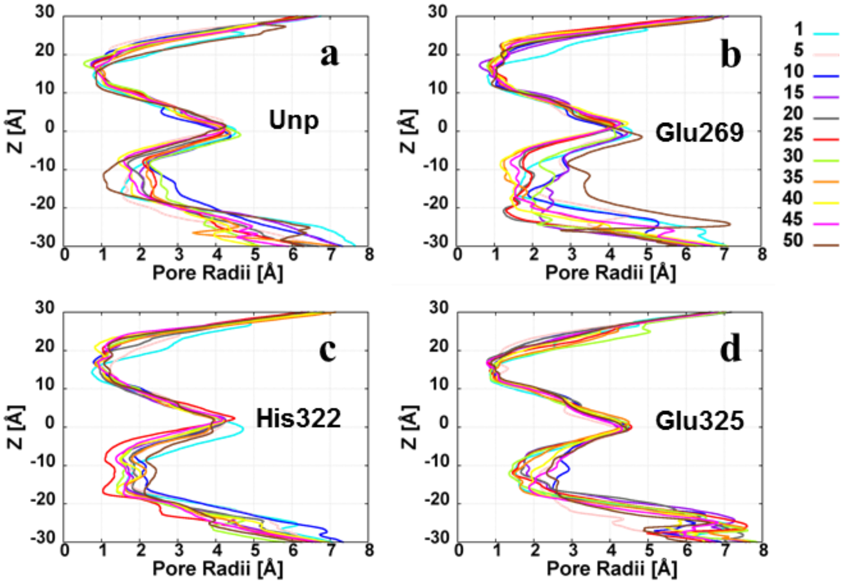

Figure 43 Protonation dependence of explicit SGLD simulations.
Pore radii obtained from the SGLD simulations in the explicit membrane on pre-equilibrated 1PV7 LacY with the guiding factor 0.75, and friction constant 5 ps^{-1} in the varied protonation states of a) un-protonated b) Glu269-protonated c) His322-protonated and d) Glu325-protonated.

SGLD has been made available for use with *NPT* dynamics and allows for the use of this method with an explicit membrane, but this is much more computationally expensive. Therefore, compared to regular SGLD with the implicit membrane, fewer simulations were performed. Only seven SGLD in the explicit membrane with SGFT and FBETA sets of (1, 1), (0.75, 5), and (0.75, 10) with three protonation states have been performed.

As shown in Figure 44, with the guiding factor 0.75, and friction constant 5 ps^{-1}, the ns-averaged pore radii data demonstrates that there is no protonation state dependence as LacY of all protonation states close on the cytoplasmic sides but LacY on the periplamic side show no motion. And the guiding factor and friction constant dependence are not clearly shown either for SGLD simulation with the explicit membrane (Figure 44). Overall, the results show that LacY has a dynamic conversion from inward-facing to occluded state on the cytoplasmic side, while with the presence of membrane, there is no motion of LacY on the periplasmic side.

To consider the short time scale effect, the LacY simulations with guiding factor 1, and friction constant 10 ps^{-1} is run for 100 ns instead of 50 ns as others (Figure 44c). There is still no motion observed on the periplasmic side. Moreover, when we look at the individual pore radii data, the transition between inward-facing to outward-facing was not observed in any run. These results suggest that the existence of the membrane may inhibit the opening on the periplasmic side or timescales are too great with this approach.

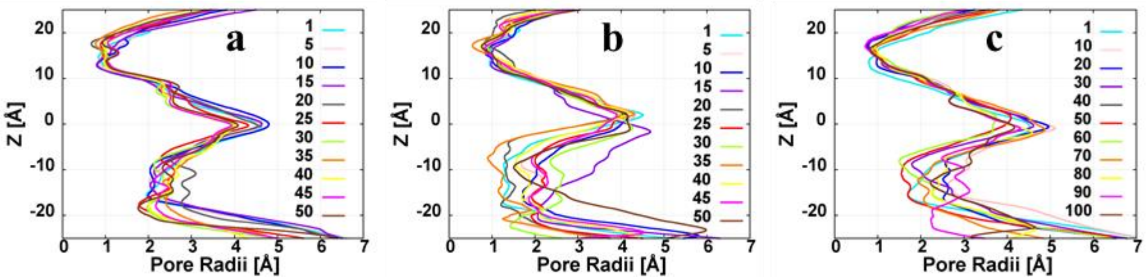

Figure 44 Exemplary pore radii from SGLD with an explicit membrane.
LacY with Glu269 protonated, SGLD with an explicit membrane on pre-equilibrated 1PV7 LacY with varied parameter sets of a) guiding factor 1, and friction constant 1 ps^{-1} b) guiding factor 0.75, and friction constant 10 ps^{-1} and c) guiding factor 1, and friction constant 10 ps^{-1} (which is run for 100 ns instead of 50 ns).

Chapter 7 Results on LacY$^{IM\text{-}EX}$ Model

This Chapter covers two topics on LacY$^{IM\text{-}EX}$ Model. Firstly, further analyses results on LacY$^{IM\text{-}EX}$ structure is presented in Section 7.1 to better understand the conformational changes and stability of this periplasmic-open model. Secondly, since DEER is only experimental method up-to-date that provides the detail information on the conformation of outward-facing LacY, the method and results of the verification of DEER data are provided in Section 7.2. The further detail related to LacY$^{IM\text{-}EX}$ and MDDS simulation is published in [21].

7.1 Further Study on the LacY$^{IM\text{-}EX}$ Structure

7.1.1 Pore Radii

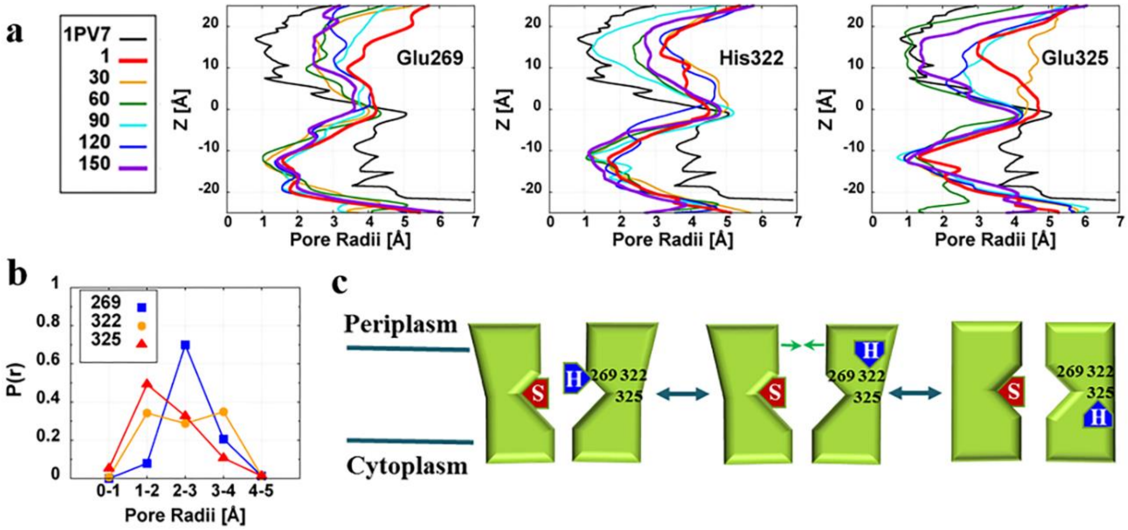

Figure 45 Protonation effect on the pore radii of LacY$^{IM\text{-}EX}$ and a proposed mechanism: a) Pore radii of LacY$^{IM\text{-}EX}$ -sugar in the explicit membrane with each of three protonation states at varied simulation time. The black curve is the pore radii of 1PV7 crystal structure, and the curves in difference colors are the pore radii of LacY$^{IM\text{-}EX}$ at time 1 to 150 ns with step 30 ns. Z in the plot represents the z-coordinate of the axis along the bilayer normal. Positive z corresponds to the periplasmic side of LacY, and negative z corresponds to the cytoplasmic side of LacY. b) Probability distribution with bin size 1Å of 1-ns block averaged minimum pore radii on the periplasmic side for Glu269 (blue square), His322 (orange circle), and Glu325 (red triangle) protonated states. c) Schematic model of conformational change of LacY during proton transport.

The MD simulation is performed on LacY$^{IM\text{-}EX}$ with sugar in the explicit membrane with three protonation states. The system information is shown in Table C.1. The pore radii of LacY$^{IM\text{-}EX}$ in the three protonation states compared to 1PV7 crystal structure are shown in Figure 45a. It demonstrates that LacY$^{IM\text{-}EX}$ remains closed on the cytoplasmic side in all three protonation states. However, the periplasmic side depends strongly on the protonation state of the three residues. For the periplasmic side, LacY$^{IM\text{-}EX}$ with E269-protonated remains open, LacY with H322-protonated is partially open, and with E325-protonated has periplasmic side tends to close. The probability distribution (histogram with bin size 1Å) of the 150-ns simulation time for E269-, H322-, and E325-protonated LacY is shown in Figure 45b. The results demonstrate that the pore radii for E269-protonated LacY are mostly in the range of 2-3 Å, and it has a higher probability in range 3-4 Å than E325-protonated LacY. E325-protonated LacY shifts the highest probability to 1-2 Å, and the probability decreases as pore radius increases from 1 to 4 Å. In the H322-protonated LacY, the pore radii are evenly distributed in 1 to 4 Å which indicates the transition of decreasing the pore radii. One main step of LacY transport cycle is the outward-facing LacY picks up a proton and a disaccharide from the periplasmic side (external side of the cell). Therefore, the pore radii suggest the process that E269 in helix VIII picks up the proton, and then the proton shuttle H322 in helix X passes the proton to E325 which is also in helix X but closer to the cytoplasmic side (Figure 45c, Figure C.1). As a proton is transported, the periplasmic side of LacY closes gradually in order to prevent the substrate leak back to the external of the cell, which agrees with the results in Pendse et al. [48].

By comparing the structures of E269-, H322-, and E325-protonated LacY at the end of simulations, it is found that the residues that make the major contribution to the closing on the periplasmic side are K42, D240, and Q241 as shown in Figure 46, which agree with the recent experimental result in [55] that the interaction between K42 and the nano-body is the key to maintaining LacY in the periplasmic-open state. In the E269-protonated LacY, D240 interacts with E325 through potassium salt bridge, while in the 325-protonated state, D240 interacts with E322 through two water bridges. In the E269-protonated LacY, K42 is in free space and does not interact with another residue; in the H322-protonated LacY, K42 interacts with D36 through a water bridge (not shown in Figure 46 for image clarity); in the E325-protonated LacY, K42 form hydrogen bonds with H35 and Q241. As protonation state changes from E269 to E325, the histidine residues H35 and H322 change rotationally rather than translationally to allow the formation of new hydrogen bonds of H35 with K42, and D240 with H322.

Besides pore radii change, the protonation effect on both the interaction of residues involved in sugar and proton transport is also observed. The experimental mutagenesis studies have found that the residues involved in the sugar recognition and binding are E126 and R144 [134]. The simulation results show that in the E269-protonated structure, Y26, F27, N119 and K358 interact with sugar ββ-(Gal*p*)$_2$ through hydrogen bonding, Y350 interacts with sugar through water bridge; in the H322-protonated structure, only F27 forms hydrogen bond with sugar ββ-(Gal*p*)$_2$; in the E325-protonated state, E126 forms hydrogen bond with sugar ββ-(Gal*p*)$_2$ which agree with the results in [55, 134] that E126 involved in TDG or NPG binding. According to recent experimental results that F27 determines the smallest pore radii on the periplasmic side [55]. The interaction difference of F27 in three conformational states confirms the importance of F27 in the conformational change. In all three (E269, H322 and E325) protonated states, R144 has electrostatic interaction with

E126 and E130. In the H322-protonated state, R144 forms an additional hydrogen bond with Y350, which suggests that, besides R144 and E126, the residues E130 and Y350 may also involve in sugar transport. Our simulation results also agree with the experimental conclusion that the residues involved in proton transport are D240, E269, R302, K319, H322 and E325 [134]. In the E269-protonated LacY, neutral E269 forms hydrogen bonds with S156. In the E325-protonated structure, negative E269 form electrostatic interaction with positive K319. In E269 and H322-protonated states, E325 form a hydrogen bond with residue Y236. Y236 may be a key residue as it forms hydrogen bonds with R302, and R302 forms hydrogen bonds with D240 in all three protonation states. This suggests that Y236 may also be involved in proton transport from H322 to E325. Identify the key residues involved in the proton/sugar transport through simulation facilitates the understanding of the correct transport process of LacY. The agreement of the key residues between the simulation and the experiment validate the outward-facing LacY^{IM-EX} model, which will help to experimentally identify and search correct outward-facing LacY crystal structure.

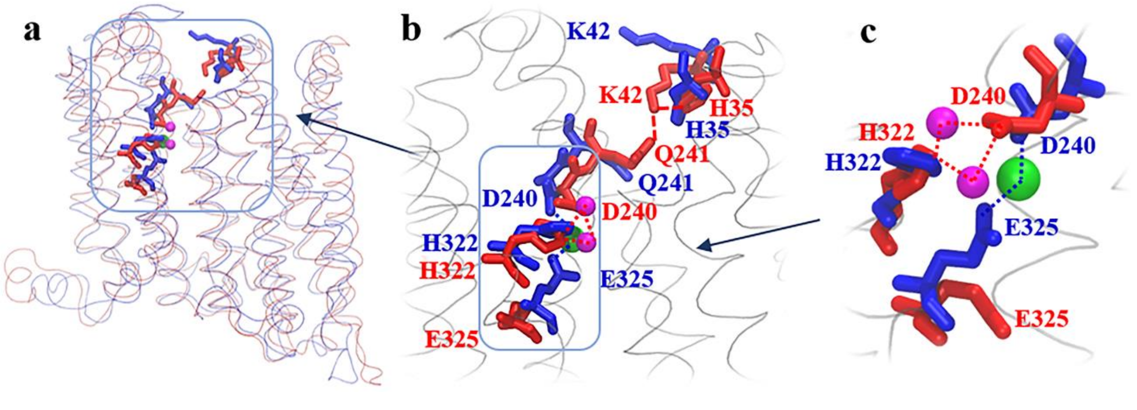

Figure 46 Residues involved pore radii change of LacY.

Residues E269 (blue) and E325-protonated (red) (snapshots at 150ns) in a) the overall LacY structure b) the region for minimum pore radii c) the region involved D240, H322, and H325 interactions. The dash lines indicate the H-bond, and the dotted lines indicate the salt/water bridge. The green sphere is the potassium involved in a salt bridge for the E269-protonated system (blue), and the magenta spheres are the water molecules involved in a two water bridge for the 325-protonated system (red). The E269-protonated background structure is not shown in the zoom-in image for clarity.

7.1.2 Residue Pair Distances

The MD simulations on previously developed outward-facing LacY$^{\text{IM-EX}}$ model [48] are performed with three (E269, H322, and E325) protonation states. The (60-150 ns averaged) residues pair C_α distance changes relative to crystal structure are calculated for each state and are compared to the distance changes obtained from DEER [62] as shown in Figure 47. The first three pairs (V105/T310, I164/T310, and I164/S375) are located on the periplasmic side, and the remaining six pairs (R73/S401, R73/Q340, S136/Q340, N137/Q340, S136/S401, and N137/S401) are on the cytoplasmic side of LacY. Compared to the initial structure (IM-EX$_\text{i}$ [48]), the residue pair distances obtained from MD simulations agrees less with DEER data in the periplasmic side. As shown in Figure 47, the residue distances for the first three pairs are generally smaller than for initial structure for all three pairs on the periplasmic side. The C_α distance increment of V105/T310 for both LacY$^{\text{IM-EX}}$ model and three MD simulations are smaller than DEER data [62], and compared to the initial LacY$^{\text{IM-EX}}$ structure, the C_α distances show decrement after MD simulations (Figure 47). For I164/S375, the MD simulations for H322 and E325 show reasonable agreement with IM-EX$_\text{i}$ and DEER data[62], but the residue distance change is smaller for E269. For I164/T310, H322 and E325 have smaller distance change compared to IM-EX$_\text{i}$, while E269 show very negligible opening. All the structures show less distance change than suggested by DEER. The deviation of the distance changes in the periplasmic side suggests that the periplasmic side of the LacY closes slightly during the MD simulation, which is probably a natural stochastic change in paired distances due to the closure of pore radii (Figure 45a). The change of CHARMM force field (FF) is another possible reason that may cause the closure of LacY on the periplasmic side. The LacY^{IM-EX} model is generated with CHARMM22/CMAP protein FF [70] and CHARMM27 lipid FF [135, 136], while CARHMM36 FF for protein [137] and lipid [138] are applied in MD simulation, which may result in a minor difference in LacY and lipid interactions as well as LacY structure.

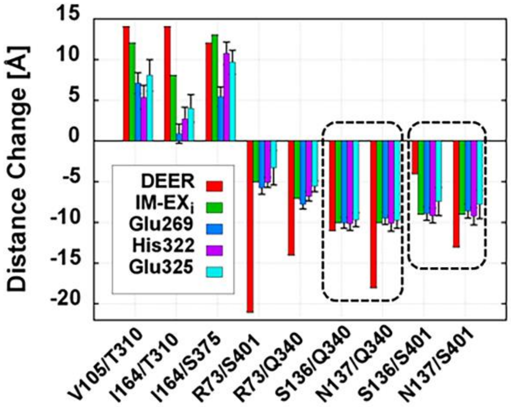

Figure 47 Residues pair C$_\alpha$-C$_\alpha$ distance changes relative to crystal structure.
The explicit MD simulated (60-150 ns averaged) residues pair C$_\alpha$-C$_\alpha$ distance changes relative to crystal structure compared to the ones obtained from DEER [62] and from the initial outward-facing LacY^{IM-EX} model [48]. The dashed rectangular marked two residue pairs that involve two neighbor residues S136 and N137.

The residue distance changes data of LacY^{IM-EXi} [48] are very similar to simulated LacY^{IM-EX} on the cytoplasmic side, and this agrees with the closed pore radii for all three protonation states on the cytoplasmic side (Figure 45a). The simulation results on LacY^{IM-EX} for the distance change of residue pairs S136/Q340 and N137/S401 agree with DEER. The major difference is that the simulated cytoplasmic distance changes R73/S401, R73/Q340, and N137/Q340 are smaller than DEER data [62]. Out of nine residues pairs provided by DEER, four pairs involve neighbor residues (inside dash rounded rectangle in Figure 47). Similar to the residue distance data of LacY^{IM-EXi} [48], the neighbor spin-labeled residue distance change based on C$_\alpha$ between S136/Q340 and N137/Q340, S136/S401 and N137/S401 are very similar in the simulated results for all three protonation states, but they are significantly different in DEER experimental data [62]. However, the C_α distance is not experimentally measured, instead, it is the distance between the coupling spins of unpaired electrons in nitroxide radical of MTSSL. Therefore, the differences may be the result of conformational changes in the spin label that are influencing the distance changes and thus gave us the motivation to investigate this further with MDDS simulations.

7.2 Verify DEER data by MDDS Simulation

7.2.1 Comparisons of the MDDS Simulation Results to DEER

The length from the protein C_α to the nitroxide O of DEER spin label MTSSL is about 8Å [90], which is large enough to affect the distance measurement. Moreover, there are five dihedral angles between C_α to the nitroxide O, each with multiple rotameric states [88, 139] (due to bond rotations). Therefore, MTSSL has its internal flexibility/dynamics, and the conformational distribution of the MTSSL will contribute to the DEER spin-pair distance distribution. This introduces additional broadening of the protein distance distribution [140], and may also shift the overall mean distance away from the real C_α protein mean distance. As shown in Figure C.2, the dummy ON spin labels sample a relatively wide range in paired distances for a given fixed structure. The raw DSL distances of pair V105/T310 obtained from MDDS on inward-facing, occluded, and outward-facing LacY are compared to DEER distance distribution (black curve line) and mean distance peaks (black vertical line) [62] (Figure C.4). As the residues V105 and T310 are on the periplasmic side, we expect the inward-facing state has the lowest residue distances, occluded has the intermediate (if there is one), and outward-facing has the maximum ones. The raw data of each ns structure for residue pair V105/T310 shows that the inward-facing and outward-facing structures contribute to the occluded expected peak at 45 Å (Figure

C.4a, c), while the occluded structure contributes to the inward-facing expected peak at 39 Å (Figure C.4b). This suggests that DEER mean residue distances may not provide the real distance of the corresponded distinct structural states of LacY, but potentially have distances that span all the states.

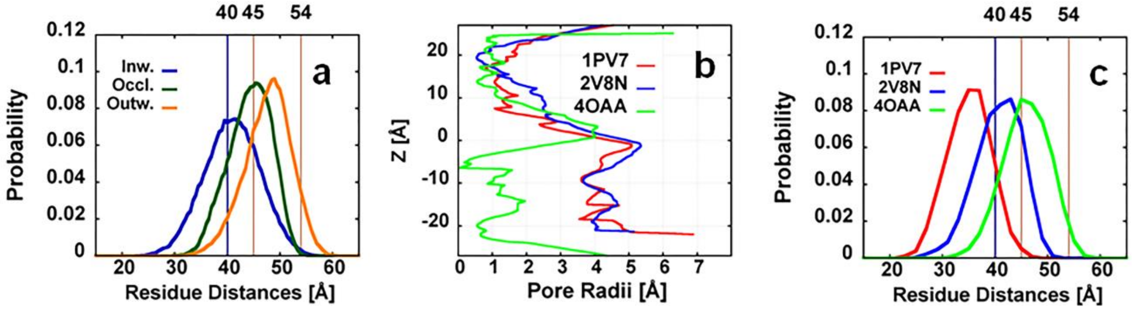

Figure 48 Spin label distance distribution obtained from MDDS simulations:
a) MDDS resulted periplasmic residue pair V105/T310 ($z_{C\alpha}$=16~20 Å) spin labels distance distribution of overall structurally averaged inward-facing, occluded and outward-facing LacY compared to DEER distance distribution [62]. The solid vertical lines are the distances that are interpreted to be in the three states for the protein. Three spin labels distances (40, 45, and 54 Å) obtained from DEER are interpreted as DEER residue distances of LacY in inward (blue), occluded to outward (orange) states, respectively. b) The pore radii of the crystal structures. c) MDDS resulted pair distance distribution of V105/T310 of the crystal structures compared to DEER data [62].

Table 13 Comparison of spin label distance on V105/T310.
The distance between C_α ($d_{C\alpha}$) and dummy ON spin labels (d_{ON}) for crystal structures (PDB code provided) and structures obtained from simulations of residue pair V105/T310 on the periplasmic side. The different states of structures from the simulation are based on pore radii. The distances from simulations are structurally averaged data and the standard errors are calculated based on six or ten structures of each state.

	Crystals			Simulations		
Distance (Å)	1pv7	2v8n	4oaa	Inward	Occluded	Outward
$d_{C\alpha}$	34	36	43	40±0.8	38±0.4	43±0.6
d_{ON}	36	42	47	41±1.4	47±0.8	50±0.7

The structurally averaged spin label distance distribution of V105/T310 for inward-facing, occluded, and outward-facing LacY are compared to DEER [62] as shown in Figure 48a. The distance distributions for the other pairs are shown in Figure C.5. The mean distance of the six structurally averaged residue pair distance of occluded LacY structure agree with DEER well (at 45 Å), but the distribution of inward-facing LacY is shifted about 3 Å higher than the suggested peak of 39 Å for this state. For the outward-facing state, our MDDS distribution maximum is shifted about 5 Å lower from the experimentally-suggested value of 54 Å. However, based on the pore radii data in Figure C.3a, the inward-facing structures have a very closed periplasmic side, from which we expect good agreement in the mean distance 39 Å. Also, if we assume the outward-facing structure should be near to the inverted inward-facing crystal structures in terms of pore radius, then the pore radii on the periplasmic side should be around 4.5 Å (Figure 48b). The average periplasmic pore radii in Figure C.3c is 3.5 Å, from which we expect the deviation between the outward-facing MDDS data and DEER will be within 2 Å. The conflicts between the residue distance distribution and pore radii analysis suggest that DEER mean distance peaks may not actually correspond to different opening/closing structural states of LacY.

A direct comparison between the pore radii and spin-label distance analysis is performed by running the MDDS simulations on the known crystal structures. The pore radii of three LacY crystal structures (2V8N [52], 1PV7 [53] and 4OAA [54]) are calculated and shown in Figure 48b, which shows that the pore radii of 2V8N and 1PV7 are very similar, and this is supported by very similar residue pair C_α distances as shown in Table 13. However, the spin label distance of periplasmic side residue pair V105/T310 in Figure 48c and Table 13 show that the spin label distances between 1PV7 and 2V8N are different by about 6 Å even though the difference in the pore radii and $d_{C\alpha}$ are very small. This clearly shows that the spin label distance is not equivalent to the pore radii nor $d_{C\alpha}$ analysis. The deviation is probably due to the orientation and movement of the spin labels.

The orientation effect is also observed by comparing the spin label distances of 4OAA and the structurally averaged occluded state, which shows that there is a 5 Å difference in $d_{C\alpha}$, while the MDDS simulation result in the same d_{ON} as shown in Table 13. The detail of orientation effect is demonstrated in the following section.

7.2.2 Effect of Orientations on Spin Label Distances

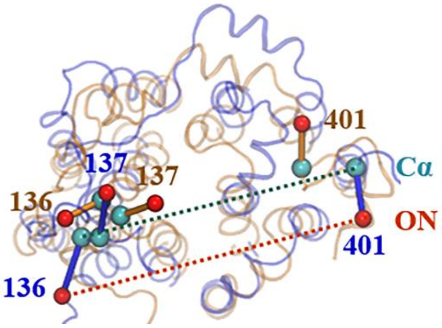

Figure 49 The top view on the cytoplasmic side of LacY with dummy spin label.
It shows the orientation of spin labels in the averaged mean distance involving the neighboring residues (S136, N137, and S401) of inward-facing (blue) and outward-facing (orange) LacY. The cyan sphere represents C_α, cyan dot line represents $d_{C\alpha}$, red sphere represents ON spin label and red dot line represents d_{ON}.

DEER measures the spin label distance between residue pairs, while the MDDS simulation allows accessing both the spin label distance (d_{ON}) and C_α distance ($d_{C\alpha}$) between residue pairs as shown in Figure 49. Through the results from MDDS simulation, we aim to answer the questions such as how much spin label distance would deviate from C_α distance, or how accurate it is to estimate protein structure directly from spin label distance. Intuitively, we know that two spin labels are closest when they point toward each other, and farthest apart when they point away from each other, which both result in greatest deviation of the spin label distance from C_α distance. But these extremes do not occur often (Figure C.2). MDDS simulated resulted d_{ON} and $d_{C\alpha}$ are provided to describe the orientation of spin label quantitatively.

Table 14 Structural averages of spin label distances of neighboring residue pairs. The two pairs are on the cytoplasmic side. The different states of structures from the simulation are based on pore radii. The distances from simulations are structurally averaged data and the standard errors are calculated based on six or ten structures of each state.

	Inward (Å)		Occluded (Å)		Outward (Å)		Outward – Inward (Å)		
Pairs	$d_{C\alpha}$	d_{ON}	$d_{C\alpha}$	d_{ON}	$d_{C\alpha}$	d_{ON}	$\Delta d_{C\alpha}$	Δd_{ON}	DEER[62]
S136/S401	43±0.7	45±1.3	36±0.5	39±1.0	31±0.2	38±0.5	-12	-7	-4
N137/S401	40±0.6	39±1.4	36±0.5	35±1.1	29±0.3	27±0.4	-11	-12	-13

To illustrate the effect of orientation on d_{ON}, we will compare the MDDS and experiments that focus on the initial motivation for this work, i.e., to understand the cause of the large differences in the spin label distances between neighboring residue the pairs (Figure 47). Table 2 shows that the inward-facing and outward-facing LacY have similar C_α distance $d_{C\alpha}$ for neighboring residue pairs S136/S401 and N137/S401, however, there are large differences in spin label distances d_{ON}. The d_{ON} of residue pair S136/S401 is much greater than d_{ON} of N137/S401 in all states, which is the direct consequence of the orientation of spin labels. As shown in Figure 49, the mean distance of dummy ON spin labels of S136 in both inward-facing (blue) and outward-facing states (orange) points away from ON spin label of S401, while ON of N137 points toward ON of S401. Moreover, the d_{ON} agrees better with $d_{C\alpha}$ for N137/S401 than S136/S401 in all three states. As demonstrated in Table 14, the deviation between d_{ON} and $d_{C\alpha}$ are within 2 Å for N137/S401, while it reaches 7 Å for S136/S401. This is also the case for the other neighboring residue pairs S136/Q340 and N137/Q340 (Table C.2). The mean spin label distances and their corresponding orientations of the remaining residues pairs are also provided in Table C.2 and Figure C.7, respectively. The fundamental reason for the difference in the deviation and the ON spin label distributions is the local environment of the spin labels. As shown in Figure 50a-c, the spin label of S136 is spatially hindered in all three states. The residues in the proximity of 10 Å of C_α and 8 Å of ON spin labels are considered to cause hindrance to the movement of spin labels. The spin label of S136 in inward-facing LacY is hindered by residues N8 to F9, and S193 to V197 (shown in blue), and the occluded state is hindered by residues K5 to W10, and A191 to P192 (shown in green), and the outward-facing sate is hindered by residues N8 to F9 and T189 to S193 (shown in orange). The hindrance in these states results in significantly different spin label distributions in three states due to volume exclusion. The spin label of N137 for the inward-facing state is hindered by residues S193 to V197 in inward-facing LacY (Figure 50d, blue), however, the spin labels for occluded and outward-facing LacY are not hindered. Unlike S136, the spin label distributions of N137 for three states show very similar hemisphere-like shape (Figure 50d-f), which explains the more accurate of d_{ON} for N137 than S136. To quantitatively analyze this, the angle between C-ON vector to the mean vector (φ) and the angle between the projected vectors of C-ON vector to the unit x-axis vector on the plane normal to mean vector (θ) are calculated, which are the equivalent spherical coordinates by rotating z-axis to the mean vector. The bottom panel of each subplot of Figure 50 shows that the distribution of φ and θ varied very different for S136 (Figure 50a-c) but is very similar for N137 (Figure 50d-f), which agrees with the visual spin label distribution on the top panel. The distribution of the C_α-ON distances, C_β-C_α-ON bond angles, and N-C_β-C_α-ON dihedral angles obtained from the MDDS simulations in inward-facing(blue), occluded(green), and outward-facing states are also calculated and shown in Figure C.7. The results demonstrate that the C_α-ON distances are very similar for all states in both S136 and N137, while distinct multiple peaks occur in the C_β-C_α-ON bond angles and the N-C_β-C_α-ON dihedral angles for S136. Among the three parameters, the dihedral angles best reflect the difference of local environment in structures in different states.

Table 14 also shows that Δd_{ON} is closer to $\Delta d_{C\alpha}$ for N137/S401 than S136/Q340. Also, both $\Delta d_{C\alpha}$ and Δd_{ON} for N137/S401 (-12 Å) agree well with DEER (-13 Å), while both $\Delta d_{C\alpha}$ and Δd_{ON} for S136/S401 deviated from DEER (-4 Å), which provides the evidence to disqualify the DEER value for S136/Q340 as the local environment alters the flexibility of the spin label in DEER. The S136 spin-labeling site is tight (Figure 50), the MTSSL may be distorted [141] and results in a different orientation in different states that affect the DEER distance distributions. Therefore, it is difficult to fully interpret the DEER d_{ON} as a metric to describe the residue distance change of LacY. The incorporation of simulation is necessary to fully probe protein structural changes. One should compare the simulated d_{ON} with DEER d_{ON}, or compare simulated $d_{C\alpha}$ to DEER derived $d_{C\alpha}$ after analyzing the orientation effect of spin labels. From the MDDS simulation results in Table 14, we estimate that the overall spin label distances change Δd_{ON} are about 3 Å greater than C_α distances changes $\Delta d_{C\alpha}$. Ideally, The $d_{C\alpha}$ should be derived in each inward-facing and outward-facing state from DEER data by incorporating MDDS simulated d_{ON} and $d_{C\alpha}$ deviation results, and then obtain the difference due to conformational change.

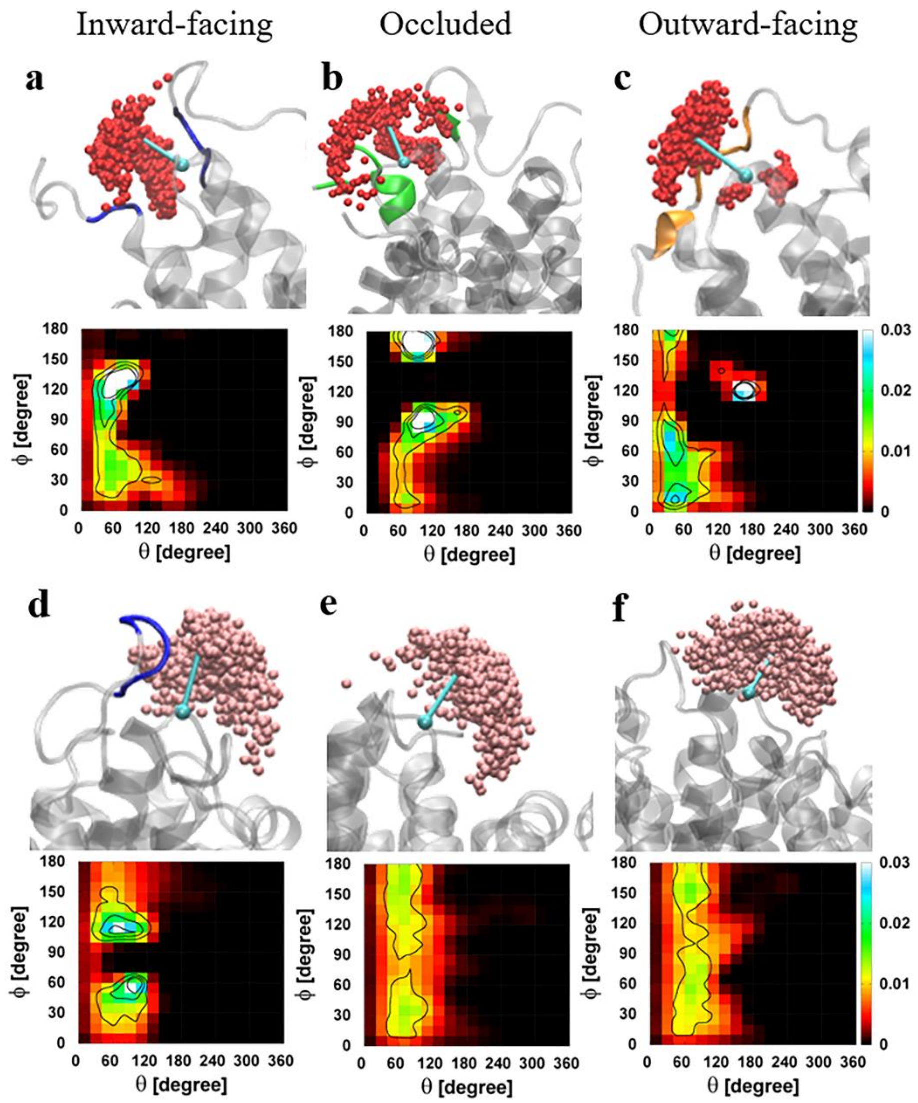

Figure 50 Spin label and angle distribution in three conformational states of LacY. Dummy ON spin label distribution in three conformational states of LacY for a-c) residue S136 (ON are shown in red), d-f) N137 (ON are shown in pink), C_α and the mean distance orientation is shown in cyan, and the possible residues that hindered the ON movement are shown in blue for inward-facing, green for occluded, and orange for the outward-facing state of LacY. Top panel of each subplot shows the distribution visualized in VMD, and the bottom panel shows angle distributions, with φ defined as the angle between C-ON vector and the mean vector, and θ as the angle between C-ON vector and the unit x-axis projection vector on the plane normal to mean vector (θ).

The pore radii and $d_{C\alpha}$ are two methods that have being used to determine the degree of opening/closing of LacY [48, 142, 143] but there is disagreement between pore radii and $d_{C\alpha}$. The pore radii of 1PV7 and 4OAA (Figure 48b) are very similar on the periplasmic side, while there is 9Å difference in $d_{C\alpha}$ of residue pair V105/T310 (Table 13). The disagreement between pore radii and $d_{C\alpha}$ are also observed during structural sampling. Figure C.3 shows that all the structures sampled in each state have very similar pore radii, however, the deviation between structures of $d_{C\alpha}$ for V105/T310 is 5Å for both inward-facing and outward-facing LacY. Let us now look at the difference between these two methods. The pore radii measure the minimum pore size which can be considered the upper bond of the cross sessional area that allows the sugar to pass. It mainly depends on the sidechain behaviors of protein lumen near the center of inner LacY structure (Figure 5). The $d_{C\alpha}$ of the spin labeled residues show the distance of the edge ends of LacY, which are mainly caused by the overall helical or bundle movement. The $d_{C\alpha}$ supports that 4OAA is in the outward-facing state as suggested in [144], while the pore radii in Figure 48b shows that it is in an occluded state. A disaccharide is not able to go through the periplasmic side with the pore size less than 2 Å. Therefore, the expected outward-facing LacY structure should meet both the criteria that having a pore size large enough to allow sugar transport on the periplasmic side and small enough to prevent sugar transport on cytoplasmic side, and also having residue distance in reasonable agreement with DEER (the derived $d_{C\alpha}$). However, we suggest that it is more appropriate to consider DEER data as a guide for helix tilt in a membrane, and pore radii should be used to classify open/closed of LacY structures.

7.2.3 Multiple Peaks in MDDS and DEER

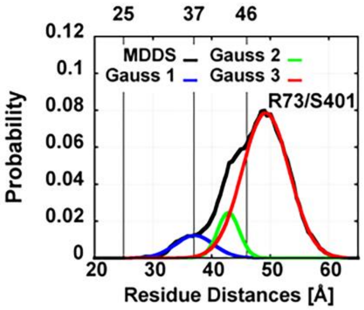

Figure 51 Spin label distance distribution showing multiple peaks from Gaussian fittings**.** MDDS resulted multiple peaks structurally averaged spin label distances of residue pair R73/S401 on the cytoplasmic side of inward-facing LacY. The black curve is the structurally averaged spin label distance distribution obtained from MDDS simulations. The blue, green and red curves are the three Gaussian functions used to fit the MDDS data. The solid vertical lines are the distances that are interpreted to be in the three states for the protein. Three spin labels distances (46, 37, and 25 Å) obtained from DEER [62] are interpreted as DEER residue distances of R73/S401 in inward-facing, occluded, and outward-facing states, respectively.

For each fixed protein structure during MDDS simulations, multiple mean distance peaks can occur. Figure 51 shows for the six structural averages of the inward-facing LacY, the residue pair R73/S401 distribution of spin label distance could be fit to three Gaussians with each having a unique mean spin label distance. The multiple peaks occur in most of the pairs as shown in Table C.3, which list all the means peaks from Gaussian fit for each pair. With the protein backbone fixed, these multiple peaks are merely due to movement of dummy spin labels. Similarly, even though multiple structural states can be captured in DEER, each peak of Gaussian need not represent the unique structure, some of the peaks may also be the consequence of the varying spin label conformations. As shown in Figure 51, the Gaussian fits DEER experimental data for R73/S401 of wild-LacY results in three distance peaks [62]. To interpret these DEER peaks, it is expected that the peak at 46 Å corresponds to inward-facing of LacY, the peak at 25 Å corresponds to an outward-facing state, and the intermediate distance 37 Å represents transition state between inward-facing and outward-facing [62]. The DEER data of C154G LacY mutant which is restricted to be inward-facing conformation shows two peaks at 38 Å and 45 Å [62] even though only one peak is expected. This unexpected additional peak can be explained by the spin label distances from MDDS simulations on the inward-facing LacY in Figure 51 which shows that the peak at 37 Å is not true intermediate states, instead, it is an indication of spin label movement in MDDS simulations and varied spin labels conformations in DEER experiment.

Additionally, the C_α distances of LacY structures obtained from simulations show three levels of distances (long, intermediate and short) for most of the residue pairs on cyto- and periplasmic sides. The pore radii of our SGLD simulation results on inward-facing LacY structure suggest that the LacY has three levels of pore size (open, loosely closed, and tightly closed) on the cytoplasmic side. However, the intermediate pore size is observed in only a few simulations, so we cannot determine whether there are two or three distinct levels of pore size on the periplasmic side. Moreover, the structures of intermediate pore radii do not correspond to intermediate residue distances. For example, the occluded structure in Figure C.3b has the tightly closed pore radii while having intermediate residue pairs C_α distances. This implies that there may be different mechanisms that induce the opening of the pore (movement of residue sidechain near pore) and the opening of helical bundles. Further studies are required to understand the correlation between the opening of the pore radii and helical bundles and their possible mechanisms.

7.2.4 Approach to Use MDDS to Interpret DEER Measurement in LacY

The MDDS results show that spin label distances strongly depend on the orientation of spin labels. Therefore, accurate analysis of DEER data for LacY also needs to consider the flexibility of the spin label [145, 146]. As shown in Figure 52, the main peak structurally averaged C_α distances, MDDS spin label distances, and DEER spin label distances all show strong linear dependence. The correlation coefficient between spin label distance obtained from MDDS simulation and DEER experiment of inward-facing and outward-facing LacY is about 0.70±0.18. The correlation coefficients of each state are provided in Table C.4. This correlation is quite strong with a small population of each state in simulation. In the MDDS simulation, the structural averages are taken from five inward-facing and ten outward-facing structures are performed, while as mentioned earlier, DEER catches (almost) all conformations which may include some transition state in sugar transport cycle. From the MDDS simulations, we estimate that the actual C_α distance changes for LacY are about 3 Å lower than provided DEER data [62]. Moreover, the MDDS simulations can be performed on the outward-facing LacY that satisfies pore-radii criteria with varied $d_{C\alpha}$ from the expected range. When the d_{ON} from MDDS simulation agrees with the DEER, the corresponded $d_{C\alpha}$ is likely to be the correct structure. By this method, we can verify whether the existing outward-facing model is correct, and also estimate the correct $d_{C\alpha}$ that facilitates to determine the structure. The MDDS simulations can also be performed on structures (having occluded pore radii) with varied $d_{C\alpha}$, in order to determine whether the intermediate DEER spin labels distance peaks are real intermediate $d_{C\alpha}$ changes or they are the consequence of the flexibility of spin labels.

The application of MDDS simulation to analyze DEER data to determine LacY structure requires accurate DEER data. However, there can be some issues in DEER data. Firstly, DEER results are strongly affected by the selection of peaks, while unfortunately, the standard of selection is vague. The selection of peaks that corresponds to inward-facing and outward-facing state is mostly based on the maximum percentage of area under the peak of each state, while some peaks that have the maximum inward-facing to outward-facing distance difference are also considered to be meaningful peaks. Secondly, DEER measures one residue pair at each experiment, so the residue distances of different pairs are not necessary from same structures. The FRET experiment results [61] shows that two domains of LacY do not behave as rigid bundles. Based on which, we expect that with three levels of residue distances, LacY may have structures beyond the inward-facing, occluded, and outward-facing state. These cases cause difficulty to propose a single model structure based on DEER measurements.

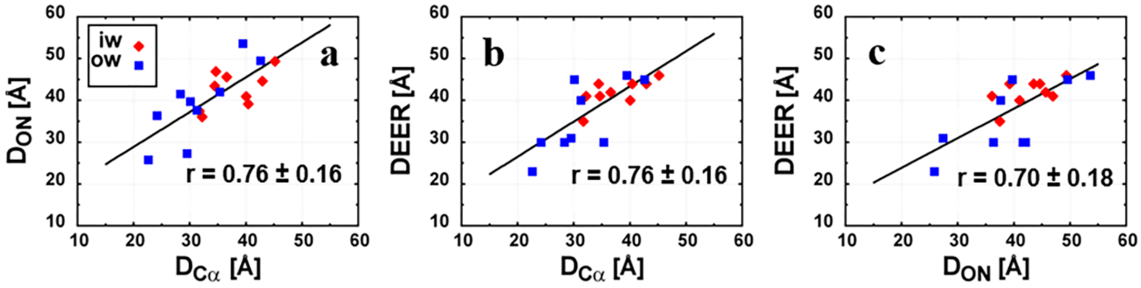

Figure 52 Correlations among the C_α-C_α distance of the residue pairs, MDDS, and DEER. The main peaks the structurally averaged distance values for nine residue pairs of LacY a) C_α-C_α distance vs. MDDS spin label distances, b) C_α-C_α distance vs. DEER spin label distances c) MDDS vs. DEER spin label distances. Inward-facing LacY (red diamond) and outward-facing (blue square) are shown with the fitting line of both states. The correlation coefficients are shown with the standard deviations in the plots.

Chapter 8 Future Direction

8.1 Studies with Lipid Membranes

My dissertation study confirmed that C36FF-AA is an accurate model for wide range of glycerol phosphate lipid types. Based on the accurate pure linoleoyl membrane, the resulted multiple-component membrane models are also reasonable. Having accurate models for membranes with linoleoyl allow for further studies on membrane-associated proteins in plants. Overall, the accurate C36FF-AA provides excellent grounds to study model development of other complex plant and even mammal membranes. Moreover, the accurate membrane based C36FF-AA model provides accurate bilayer environment for protein study. It greatly facilitates the accurate MD simulation of LacY and other transmembrane proteins in the explicit membrane, which is important for the accurate prediction of the structure and mechanism of the transmembrane proteins.

The C36FF-UA of a few saturated glycerol phosphate lipids have been developed and tested which has been published previously in [125]. Through our new study, we developed extended types of lipids, such as DM lipids of various head groups, saturated PC and monounsaturated PE lipids with varied chain lengths, and also four types of sphingomyelin and ceramides. The accuracy of these C36FF-UA will be tested, and the results of which will be helpful to look at various large-scale or long time behaviors.

8.2 Study the Proton/Sugar Transport Mechanism of LacY

The MD simulation on LacY$^{IM\text{-}EX}$ model shows that LacY closes its periplasmic side as proton transport from Glu269 to 325 (toward the cytoplasmic side). And it also demonstrates that D240, R302 involved in proton transport, and R144 involved in sugar transport, which agree with the experiment. In the future, the study on the other residues that may be either directly or indirectly involved in the proton/sugar transport will be of interest. For example, it is found that S53, Q60, and F354 are involved in sugar transport [48]. Some other residues such as K319, etc. may not directly bind to neither proton nor sugar, but they facilitate the proton/sugar transport.

Also, the helix-helix interactions will be studied by calculating and observing the helix packing changes during LacY conformational change or proton/sugar transport. Besides the helix-helix distance change, the breaking/formation of hydrogen bonding during the helix packing change will be probed. Umbrella sampling simulations to calculate the Potential of Mean Force (PMF) at the varied conformational states and varied positions of sugar inside LacY to estimate the free energies may also provide more thermodynamic insight. Moreover, PMFs along multiple paths between the inward-facing to the outward-facing state will be calculated and compared. To improve efficiency, weighted histogram analysis (WHAM) [147] will be combined with PMF calculation.

8.3 Further Study the Method for Effective Protein Structure Searching

Currently, we have obtained the inward-facing, occluded structures, and also some reasonable outward-facing LacY structures from SGLD simulations in the implicit membrane. We are currently testing the stability of the outward-facing LacY structures obtained from SGLD simulation by the MD simulations with the explicit membrane, from which the potentially better structures may be obtained. If the results show that the structures obtained are stable, then the similar method may be applied to other SAT transporters.

The guiding factor and friction constant set from the previous SGLD simulation on LacY, such as (0.75, 1), (1, 1), and (0.75, 5), may be applied to obtain the structures of other transmembrane proteins that are not experimentally available. After the implicit SGLD simulations, then explicit membrane MD simulation will be performed to test the stability or even obtain a better structures. If successful, this could be an excellent tool to probe conformational changes in SAT proteins.

Computation Resources

The high performance computational resources are used. Deepthought maintained by the Division of Information Technology at the University of Maryland will be used to setup and equilibrate the systems and the post-production analysis. The production run simulations will be performed either in Deepthought2 at the University of Maryland or Extreme Science and Engineering Discovery Environment (XSEDE) allocations on Stampede at Texas Advanced Computer Center (TACC).

Appendices

Appendix A

Table A.1 Temperature (ranges imply 10 **°C** increments) and simulation time of each lipid system.

Tail	Head	T (°C)	Time (ns)	Tail	Head	T (°C)	Time (ns)
DL	PA	30-80	100	DX	PA	40	100
	PC	30-80	100		PC	30-40	100
	PE	30-80	100		PE	30-40	100
	PG	30-80	100		PG	40	100
	PS	30-80	100		PS	30-40	100
DM	PA	40-80	100	DO	PA	30	150
						40	100
	PC	30-80	100		PC	30-40	100
	PE	40-80	100		PE	30-40	100
	PG	30-80	100		PG	30-40	100
	PS	40	150		PS	30-40	100
		50-80	100				
DP	PA	50-80	100	DG	PA	40	100
	PC	50-80	100		PC	30-40	100
	PE	60-80	100		PE	30-40	100
	PG	50-80	100		PG	40	100
	PS	50	150		PS	30	150
		60-80	100			40	100
DS	PA	70-80	100	DE	PA	40	
	PC	60-80	100		PC	30-40	100
	PE	70-80	100		PE	30-40	100
	PG	60-80	100		PG	40	100
	PS	70-80	100		PS	30-40	100
PO	PA	30-40	100	DN	PA	40	100
	PC	30-40	100		PC	30	150
						40	100
	PE	30-40	100		PE	30-40	100
	PG	30-40	100		PG	40	100
	PS	30	100		PS	30-40	100
		40	150				
SO	PA	30-40	100	SA	PA	30-40	100
	PC	30-40	100		PC	30-40	100
	PE	30-40	100		PE	30-40	100
	PG	30	150		PG	30-40	100
		40	100				

	PS	30	150		PS	30-40	100
		40	100				
PL	PA			SD	PA	30-40	100
	PC	30-40	100		PC	30-40	100
	PE	30-40	100		PE	30-40	100
	PG				PG	30-40	100
	PS	30	150		PS	30	150
		40	100			40	100
SL	PA			DA	PA	30-40	100
	PC	30-40	100		PC	30-40	100
	PE	30-40	100		PE	30-40	100
	PG				PG	30-40	100
	PS	30	150		PS	30-40	100
		40	100				

Table A.2 The comparison of structural properties with bilayer system size of 72 and 128 lipids. The results in Table R1 show that the finite size of the 72-lipids system has no or small effect on DOPA and DOPG for almost all structural properties included in our analyses. Specifically, there is no finite size effect on the surface area per lipid (SA/lip) in Å^2, overall bilayer thickness(D_B), hydrophobic thickness (D_C), and headgroup-to-headgroup thickness (D_{HH}) in Å; there is a minor size effect on the order parameters (S_{CD}) in the *sn*-1 and *sn*-2 chains, and intra-lipid and inter-lipid hydrogen bonds (N_{HB}^{intra} and N_{HB}^{inter}), and average cluster size, i.e. number of lipids per cluster ($S_{cluster}$). The lipid wobble C_{22}-C_{32} correlation time (τ_2) can be affected by the system size due to the change of N_{HB}^{inter}.

Lipid	DOPA		DOPG		DOPS	
N_{lipid}	72	128	72	128	72	128
SA/lip	65.1±0.4	64.8±0.2	72.6±0.2	72.6±0.2	64.1±0.4	63.5± 0.1
D_B	36.1±0.3	36.2±0.3	35.1±0.3	34.9±0.3	38.8±0.3	39.7±0.3
D_C	29.7±0.1	29.8±0.1	26.8±0.1	26.8±0.1	30.1±0.1	30.4±0.1
D_{HH}	38.2±0.1	38.4±0.1	37.6±0.1	37.8±0.1	42.8±0.1	44.0±0.1
S_{CD} *sn*-1	0.191±0.004	0.184±0.005	0.157±0.005	0.148±0.005	0.199±0.005	0.195±0.006
S_{CD} *sn*-2	0.202±0.006	0.196±0.003	0.171±0.005	0.164±0.004	0.206±0.006	0.200±0.003
N_{HB}^{intra}	0.014±0.0005	0.015±0.0004	0.481±0.003	0.475±0.002	0.072±0.005	0.056±0.002
N_{HB}^{inter}	0.172±0.003	0.147±0.001	0.288±0.004	0.287±0.003	0.922±0.010	0.964±0.007
τ_2 (ns)	8.75	5.03	4.17	6.10	7.10	10.43
$S_{cluster}$	4.41	4.69	3.56	3.74	5.45	5.53

Table A.3 χ^2 of MD-based and scaled x-ray form factor of saturated lipids with unilamellar vesicle sample [6-8]. (The temperatures for DLPE experimental x-ray form factors are 35, 45, 55 °C, respectively.)

	Tail	DL	DM	DP	DS
Head	T/°C	12:0	14:0	16:0	18:0
PC	30	4.3	2.3		
	50	3.4	7.7	3.4	
	60	3.1	5.4	5.8	6.7
PE	40	1.8			
	50	3.0			
	60	5.5			
PG	30	3.4	6.7		
	50	2.7	1.8	4.3	
	60	4.2	2.5	1.7	4.5

Table A.4 χ^2 of MD-based and x-ray form factor of unsaturated lipids with unilamellar vesicle sample [6-10]. (The temperature for POPS experimental x-ray form factors is 25 °C.)

	Tail	DX	DO	DG	DE	DN	PO	SO
Head	T(°C)	16:1	18:1	20:1	22:1	24:1	16:0/18:1	18:0/18:1
PC	30	40.2	67.5	28.2	39.4	50.5	2.8	5.4
	40						24.6	
PE	30							
	40						14.2	9.8
PG	30		6.2				5.7	3.7
	40							
PS	30						561.0	

Table A.5 MD-based surface area per lipid (Å^2) with standard errors of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	56.3±0.3			
	40	57.5±0.2	54.7±0.4		
	50	60.2±0.2	57.4±0.3	53.6±0.4	
	60	61.5±0.2	58.9±0.3	58.0±0.3	
	70	63.1±0.2	61.5±0.2	60.6±0.3	58.2±0.3
	80	65.1±0.1	63.6±0.2	63.0±0.2	61.8±0.2
PC	30	63.1±0.3	60.2±0.6		
	40	63.3±0.4	62.1±0.3		
	50	64.3±0.3	63.1±0.3	61.8±0.3	
	60	66.0±0.2	65.0±0.4	63.4±0.4	61.8±0.3
	70	66.8±0.2	66.0±0.2	65.1±0.3	63.4±0.2
	80	68.2±0.3	67.3±0.2	65.7±0.3	66.1±0.2
PE	30	53.8±0.2			
	40	55.0±0.4	52.5±0.3		
	50	56.8±0.2	53.6±0.5		
	60	58.4±0.2	56.9±0.2	54.8±0.3	
	70	59.9±0.2	58.3±0.2	57.1±0.2	55.9±0.3
	80	61.0±0.2	59.8±0.2	58.9±0.2	58.6±0.2
PG	30	66.9±0.3	64.8±0.4		
	40	68.1±0.2	67.2±0.4		
	50	70.0±0.2	69.0±0.4	66.4±0.4	
	60	70.9±0.3	70.7±0.3	70.1±0.4	68.1±0.4
	70	72.7±0.2	71.7±0.3	71.1±0.3	70.6±0.2
	80	73.5±0.2	73.7±0.3	72.6±0.3	72.4±0.3
PS	30	55.8±0.3			
	40	57.0±0.2	54.3±0.2		
	50	58.4±0.2	57.0±0.4	53.1±0.3	
	60	59.5±0.2	58.0±0.3	56.7±0.2	
	70	60.9±0.2	60.3±0.2	59.4±0.3	57.7±0.2
	80	63.2±0.2	62.1±0.3	60.9±0.2	59.8±0.3

Table A.6 MD-based surface area per lipid (Å^2) with standard errors of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		65.1±0.4			
	40	66.5±0.3	66.7±0.3	65.2±0.2	62.5±0.3	60.7±0.3
PC	30	68.3±0.3	69.3±0.3	67.6±0.2	65.3±0.5	62.0±0.3
	40	69.7±0.2	69.4±0.2	68.4±0.2	66.6±0.3	65.0±0.3
PE	30	62.1±0.3	61.4±0.2	60.9±0.3	58.7±0.2	50.9±0.1
	40	63.7±0.2	63.2±0.3	61.4±0.3	60.6±0.2	57.9±0.3
PG	30		72.6±0.2			
	40	73.9±0.3	73.1±0.4	72.0±0.3	70.2±0.3	69.7±0.2
PS	30	64.8±0.2	64.1±0.4	61.6±0.3	58.9±0.3	53.5±0.3
	40	65.2±0.2	64.8±0.4	63.7±0.3	61.5±0.3	58.0±0.3

Table A.7 MD-based surface area per lipid (Å^2) with standard errors of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	59.8±0.4	58.7±0.3			66.9±0.3	65.7±0.2	74.0±0.3
	40	61.6±0.4	61.2±0.2			68.5±0.2	67.8±0.2	74.2±0.2
PC	30	65.5±0.3	65.2±0.3	66.8±0.3	66.7±0.2	71.0±0.3	71.2±0.2	76.4±0.2
	40	67.6±0.3	66.3±0.4	67.4±0.4	67.4±0.3	71.5±0.2	70.8±0.3	76.4±0.2
PE	30	56.3±0.4	55.4±0.2	59.6±0.3	58.2±0.3	64.3±0.3	63.1±0.3	70.8±0.3
	40	57.5±0.2	57.7±0.2	61.0±0.4	60.6±0.2	65.4±0.2	63.9±0.3	71.2±0.3
PG	30	68.0±0.5	68.8±0.3			73.9±0.4	74.4±0.3	79.9±0.2
	40	69.5±0.3	69.6±0.4			76.3±0.3	76.0±0.3	80.7±0.2
PS	30	58.5±0.3	57.5±0.3	60.1±0.3	61.1±0.3	65.1±0.3	64.8±0.4	72.3±0.3
	40	59.4±0.2	59.2±0.4	62.6±0.2	62.4±0.3	66.4±0.3	63.8±0.3	73.1±0.3

Table A.8 MD-based overall bilayer thickness D_B (Å) of saturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.3 Å.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	30.8			
	40	30.4	36.0		
	50	29.3	34.6	40.6	
	60	29.0	34.1	38.3	
	70	28.3	33.1	37.2	42.4
	80	28.0	32.1	36.0	40.5
PC	30	31.6	36.0		
	40	31.1	35.9		
	50	31.1	35.3	39.6	
	60	30.4	34.9	38.9	44.1
	70	30.6	34.6	38.8	43.3
	80	30.3	34.1	38.4	41.7
PE	30	35.1			
	40	34.7	40.4		
	50	33.8	39.6		
	60	33.1	37.8	43.5	
	70	32.5	37.1	42.1	46.8
	80	32.2	36.7	41.1	45.0
PG	30	28.8	33.3		
	40	28.4	32.4		
	50	28.0	31.5	36.4	
	60	27.9	31.1	34.9	39.1
	70	27.4	31.0	34.4	37.9
	80	27.3	30.4	34.0	37.2
PS	30	36.1			
	40	35.7	41.7		
	50	34.6	39.6	46.8	
	60	34.3	39.2	43.9	
	70	33.3	37.8	42.0	47.3
	80	32.1	36.7	41.2	45.8

Table A.9 MD-based overall bilayer thickness D_B (Å) of mono-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.3 Å.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		36.1			
	40	32.4	35.6	39.6	45.2	49.5
PC	30	34.6	37.1	41.6	46.6	52.7
	40	34.1	37.5	41.5	46.1	50.8
PE	30	36.6	40.7	44.6	50.2	61.0
	40	36.0	40.3	44.7	48.9	54.8
PG	30		35.1			
	40	31.8	35.1	38.8	43	46.8
PS	30	34.2	38.8	45.6	51.3	60.7
	40	35.1	39.2	43.6	48.9	56.5

Table A.10 MD-based overall bilayer thickness D_B (Å) of mixed and poly-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.3 Å.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	37.8	40.3			36.1	37.9	33.0
	40	36.9	39.3			35.5	37.1	33.2
PC	30	37.4	39.6	37.4	38.8	36.9	37.9	34.3
	40	37.7	39.8	37.1	38.6	37.0	38.4	34.7
PE	30	43.0	45.6	40.2	43.0	39.8	41.7	36.4
	40	41.8	44.1	39.5	41.4	39.4	41.6	36.6
PG	30	36.2	37.3			35.2	36.1	33.1
	40	34.7	37.3			34.3	35.5	33.0
PS	30	42.7	45.4	41.0	42.2	39.2	41.1	34.9
	40	41.7	43.5	39.4	41.5	39.4	42.3	34.9

Table A.11 MD-based hydrocarbon region (i.e. twice the hydrophobic) thickness $2D_C$ (Å) of saturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	23.3			
	40	23.1	28.1		
	50	22.4	27.2	32.5	
	60	22.2	26.8	30.9	
	70	21.8	26.2	30.1	34.9
	80	21.6	25.5	29.2	33.5
PC	30	21.2	25.3		
	40	21.1	25.1		
	50	21.0	24.9	28.9	
	60	20.7	24.6	28.5	33.1
	70	20.8	24.5	28.2	32.4
	80	20.6	24.2	28.2	31.4
PE	30	24.2			
	40	24.0	29.1		
	50	23.5	28.5		
	60	23.2	27.5	32.5	
	70	22.9	27.2	31.8	36.2
	80	22.8	26.8	31.1	35.1
PG	30	20.0	24.0		
	40	19.9	23.6		
	50	19.6	23.0	27.2	
	60	19.6	22.8	26.2	30.0
	70	19.3	22.7	25.9	29.3
	80	19.4	22.3	25.7	28.8
PS	30	23.5			
	40	23.2	28.2		
	50	23.0	27.3	33.3	
	60	22.8	27.0	31.6	
	70	22.5	26.4	30.6	35.2
	80	22.0	26.0	30.1	34.4

Table A.12 MD-based hydrocarbon region thickness $2D_C$ (Å) of mono-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		29.7			
	40	26.2	29.4	33.2	38.3	42.5
PC	30	25.2	28	31.9	36.3	41.5
	40	24.9	28.2	31.7	36	40.3
PE	30	27.4	31.4	35.3	40.2	48.9
	40	27.1	31.1	35.2	39.3	44.6
PG	30		26.8			
	40	23.8	26.8	30.2	34.2	37.8
PS	30	26.3	30.1	34.8	39.9	47.0
	40	26.4	30.3	34.0	38.7	44.2

Table A.13 MD-based hydrocarbon region thickness $2D_C$ (Å) of mixed and poly-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	30.9	33.1			30.0	31.7	27.8
	40	30.1	32.3			29.6	31.0	27.9
PC	30	28.1	29.9	27.5	29.0	28.5	29.3	27.0
	40	27.8	29.8	27.4	29.1	28.5	29.7	27.1
PE	30	32.6	35.0	30.6	33.0	31.2	32.9	28.9
	40	31.8	34.0	30.1	32.0	30.9	32.8	29.1
PG	30	27.2	28.6			27.4	28.1	25.9
	40	26.2	28.4			26.6	27.8	25.8
PS	30	31.3	33.9	30.1	31.5	30.8	32.1	28.3
	40	30.9	32.3	29.4	31.2	30.7	32.7	28.3

Table A.14 MD-based headgroup-to-headgroup distance D_{HH} (Å) of saturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	32.4			
	40	32.4	37.0		
	50	31.4	36.6	41.8	
	60	30.8	36.2	39.8	
	70	30.6	35.0	39.2	43.8
	80	30.6	34.4	37.8	42.6
PC	30	30.8	37.0		
	40	31.0	36.4		
	50	31.2	35.6	39.2	
	60	30.4	34.6	38.2	44.2
	70	31.4	34.6	38.6	43.4
	80	30.8	34.0	38.8	42.2
PE	30	34.8			
	40	34.6	40.0		
	50	34.2	39.6		
	60	33.4	38.2	43.2	
	70	33.4	37.2	42.2	46.4
	80	32.2	36.8	41.6	45.6
PG	30	31.4	36.4		
	40	31.2	36.0		
	50	30.2	35.0	39.0	
	60	30.6	33.4	38.0	41.4
	70	30.2	33.4	36.6	40.6
	80	30.2	32.8	36.6	40.4
PS	30	36.8			
	40	36.8	41.4		
	50	36.6	40.6	47.8	
	60	35.8	40.4	44.8	
	70	35.6	39.4	43.2	48.6
	80	34.4	38.8	43.0	47.6

Table A.15 MD-based headgroup-to-headgroup distance D_{HH} (Å) of mono-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		38.2			
	40	34.8	38.2	41.8	47.6	51.8
PC	30	35.6	38.6	42.6	47.8	52.6
	40	35.2	38.4	42.6	46.8	51.6
PE	30	37.8	41.8	45.2	50.4	60.2
	40	36.8	40.8	45.4	49.8	55.0
PG	30		37.6			
	40	34.4	37.4	41.6	45.6	48.8
PS	30	39.6	42.8	47.8	53.4	
	40	39.2	43.0	46.8	52.0	58.0

Table A.16 MD-based headgroup-to-headgroup distance D_{HH} (Å) of mixed and poly-unsaturated lipids at varied temperatures and chain lengths. The standard error of each is less than 0.1 Å.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	39.8	42.0			38.2	40.6	36.0
	40	39.2	41.2			38.2	39.2	36.0
PC	30	38.0	39.8	38.4	39.8	38.2	40.0	37.2
	40	38.4	40.4	37.8	40.2	38.6	40.2	36.8
PE	30	43.2	45.6	40.8	43.8	40.2	42.4	38.2
	40	42.0	44.2	40.2	42.0	40.8	42.4	38.4
PG	30	38.4	39.6			38.4	39.0	36.0
	40	37.6	39.8			36.8	37.8	35.8
PS	30	44.8	47.2	43.8	44.2	43.2	44.0	40.6
	40	44.4	46.0	42.4	43.8	43.0	45.2	40.6

Table A.17 MD-based *sn*-1 C4 to C7 averaged deuterium order parameter S_{CD} of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	0.24			
	40	0.23	0.26		
	50	0.21	0.24	0.27	
	60	0.20	0.23	0.24	
	70	0.18	0.21	0.22	0.24
	80	0.18	0.20	0.21	0.22
PC	30	0.19	0.21		
	40	0.19	0.20		
	50	0.18	0.21	0.21	
	60	0.17	0.19	0.20	0.22
	70	0.17	0.18	0.20	0.21
	80	0.16	0.17	0.19	0.20
PE	30	0.26			
	40	0.25	0.29		
	50	0.24	0.27		
	60	0.22	0.24	0.27	
	70	0.21	0.23	0.25	0.26
	80	0.20	0.22	0.24	0.24
PG	30	0.17	0.19		
	40	0.16	0.18		
	50	0.15	0.16	0.19	
	60	0.15	0.16	0.17	0.18
	70	0.14	0.15	0.16	0.17
	80	0.13	0.14	0.16	0.17
PS	30	0.24			
	40	0.23	0.27		
	50	0.23	0.25	0.28	
	60	0.21	0.24	0.25	
	70	0.21	0.22	0.23	0.25
	80	0.19	0.21	0.22	0.23

Table A.18 MD-based *sn*-1 C4 to C7 averaged deuterium order parameter S_{CD} of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		0.19			
	40	0.18	0.18	0.22	0.24	0.25
PC	30	0.17	0.17	0.20	0.22	0.26
	40	0.16	0.16	0.20	0.22	0.23
PE	30	0.21	0.22	0.24	0.27	0.33
	40	0.19	0.20	0.23	0.25	0.29
PG	30		0.16			
	40	0.14	0.15	0.17	0.19	0.20
PS	30	0.19	0.20	0.23	0.26	0.33
	40	0.19	0.19	0.21	0.25	0.28

Table A.19 MD-based *sn*-1 C4 to C7 averaged deuterium order parameter S_{CD} of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	0.24	0.25			0.22	0.22	0.07
PA	40	0.22	0.24			0.20	0.20	0.06
PC	30	0.20	0.21	0.20	0.20	0.20	0.20	0.07
PC	40	0.19	0.20	0.19	0.19	0.19	0.19	0.06
PE	30	0.28	0.28	0.24	0.26	0.22	0.23	0.07
PE	40	0.25	0.25	0.23	0.23	0.21	0.23	0.07
PG	30	0.19	0.22			0.18	0.18	0.06
PG	40	0.17	0.18			0.17	0.17	0.06
PS	30	0.25	0.26	0.24	0.23	0.23	0.23	0.07
PS	40	0.24	0.23	0.23	0.23	0.21	0.23	0.07

Table A.20 MD-based *sn*-2 C4 to C7 averaged deuterium order parameter S_{CD} of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	0.25			
	40	0.25	0.27		
	50	0.23	0.25	0.28	
	60	0.21	0.24	0.25	
	70	0.20	0.23	0.23	0.25
	80	0.19	0.21	0.22	0.23
PC	30	0.21	0.23		
	40	0.21	0.22		
	50	0.20	0.21	0.22	
	60	0.19	0.21	0.22	0.23
	70	0.18	0.20	0.21	0.22
	80	0.18	0.19	0.20	0.20
PE	30	0.27			
	40	0.27	0.30		
	50	0.25	0.28		
	60	0.24	0.25	0.28	
	70	0.22	0.24	0.26	0.27
	80	0.21	0.23	0.24	0.25
PG	30	0.19	0.21		
	40	0.17	0.19		
	50	0.17	0.18	0.20	
	60	0.16	0.17	0.18	0.19
	70	0.15	0.17	0.18	0.18
	80	0.15	0.16	0.17	0.17
PS	30	0.27			
	40	0.25	0.27		
	50	0.24	0.25	0.30	
	60	0.23	0.25	0.26	
	70	0.22	0.23	0.24	0.25
	80	0.21	0.22	0.23	0.24

Table A.21 MD-based *sn*-2 C4 to C7 averaged deuterium order parameter S_{CD} of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		0.20			
	40	0.19	0.20	0.22	0.24	0.25
PC	30	0.18	0.19	0.22	0.23	0.26
	40	0.18	0.18	0.21	0.23	0.22
PE	30	0.21	0.22	0.25	0.27	0.32
	40	0.20	0.21	0.24	0.25	0.29
PG	30		0.17			
	40	0.16	0.17	0.19	0.21	0.22
PS	30	0.20	0.21	0.24	0.28	0.33
	40	0.20	0.20	0.23	0.25	0.29

Table A.22 MD-based *sn*-2 C4 to C7 averaged deuterium order parameter S_{CD} of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	0.22	0.23			0.05	0.06	0.04
	40	0.21	0.22			0.05	0.05	0.05
PC	30	0.18	0.20	0.18	0.19	0.05	0.05	0.05
	40	0.18	0.19	0.18	0.18	0.05	0.05	0.04
PE	30	0.24	0.26	0.22	0.23	0.05	0.06	0.05
	40	0.23	0.24	0.21	0.22	0.06	0.06	0.05
PG	30	0.18	0.16			0.04	0.05	0.05
	40	0.17	0.17			0.04	0.05	0.04
PS	30	0.23	0.23	0.22	0.21	0.05	0.06	0.05
	40	0.22	0.22	0.21	0.21	0.05	0.06	0.05

Table A.23 Inter-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{inter}}$ of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	0.212			
	40	0.194	0.213		
	50	0.175	0.186	0.205	
	60	0.166	0.176	0.175	
	70	0.153	0.161	0.161	0.171
	80	0.141	0.147	0.146	0.151
PE	30	0.693			
	40	0.678	0.706		
	50	0.659	0.697		
	60	0.643	0.632	0.692	
	70	0.614	0.625	0.674	0.664
	80	0.617	0.618	0.630	0.615
PG	30	0.309	0.325		
	40	0.295	0.296		
	50	0.278	0.281	0.293	
	60	0.267	0.269	0.271	0.272
	70	0.247	0.252	0.257	0.254
	80	0.239	0.239	0.244	0.240
PS	30	1.071			
	40	1.057	1.066		
	50	0.967	0.994	1.082	
	60	0.993	0.970	0.984	
	70	0.909	0.914	0.919	0.969
	80	0.834	0.907	0.905	0.920

Table A.24 Inter-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{inter}}$ of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		0.172			
	40	0.159	0.161	0.165	0.178	0.182
PE	30	0.586	0.581	0.582	0.624	0.768
	40	0.565	0.572	0.598	0.611	0.626
PG	30		0.288			
	40	0.270	0.266	0.275	0.292	0.299
PS	30	0.922	0.922	0.987	1.029	1.152
	40	0.890	0.903	0.945	0.972	1.047

Table A.25 Inter-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{inter}}$ of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	0.193	0.200			0.168	0.166	0.150
	40	0.182	0.179			0.155	0.156	0.141
PE	30	0.669	0.690	0.606	0.632	0.561	0.548	0.498
	40	0.622	0.671	0.617	0.603	0.546	0.551	0.500
PG	30	0.316	0.255			0.283	0.278	0.268
	40	0.276	0.281			0.262	0.268	0.253
PS	30	1.075	1.064	1.056	1.007	0.894	0.983	0.840
	40	0.951	0.957	0.947	0.980	0.916	0.962	0.829

Table A.26 Intra-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{intra}}$ of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	0.013			
	40	0.013	0.013		
	50	0.014	0.014	0.012	
	60	0.013	0.013	0.014	
	70	0.013	0.013	0.013	0.013
	80	0.014	0.014	0.014	0.013
PE	30	0.068			
	40	0.071	0.074		
	50	0.078	0.071		
	60	0.078	0.081	0.075	
	70	0.079	0.083	0.081	0.081
	80	0.087	0.082	0.084	0.085
PG	30	0.473	0.466		
	40	0.462	0.459		
	50	0.460	0.458	0.455	
	60	0.445	0.448	0.442	0.450
	70	0.440	0.438	0.435	0.441
	80	0.434	0.430	0.432	0.425
PS	30	0.052			
	40	0.062	0.058		
	50	0.065	0.067	0.047	
	60	0.065	0.066	0.070	
	70	0.075	0.083	0.074	0.074
	80	0.084	0.078	0.077	0.081

Table A.27 Intra-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{intra}}$ of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		0.014			
	40	0.015	0.015	0.013	0.014	0.014
PE	30	0.072	0.069	0.072	0.065	0.080
	40	0.072	0.075	0.074	0.075	0.076
PG	30		0.481			
	40	0.469	0.473	0.475	0.463	0.466
PS	30	0.055	0.072	0.052	0.058	0.047
	40	0.070	0.070	0.065	0.056	0.066

Table A.28 Intra-lipid number of hydrogen bonds per lipid $N_{\mathrm{HB}}^{\mathrm{intra}}$ of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	0.015	0.014			0.015	0.015	0.016
PA	40	0.014	0.014			0.015	0.015	0.017
PE	30	0.067	0.066	0.072	0.078	0.071	0.069	0.070
PE	40	0.075	0.069	0.071	0.072	0.074	0.078	0.076
PG	30	0.467	0.530			0.490	0.480	0.489
PG	40	0.465	0.465			0.472	0.479	0.482
PS	30	0.057	0.051	0.062	0.068	0.061	0.058	0.069
PS	40	0.076	0.065	0.072	0.061	0.066	0.073	0.072

Table A.29 The average number of lipids per cluster S_{cluster} of saturated lipids at varied temperatures.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	6.17			
	40	5.87	7.14		
	50	5.39	6.07	7.20	
	60	5.35	5.65	5.80	
	70	4.97	5.42	5.31	5.90
	80	4.95	5.09	5.07	5.31
PC	30	4.18	4.30		
	40	3.94	4.28		
	50	4.04			
	60	3.96	4.05	4.25	4.36
	70	3.95	4.08	4.08	4.12
	80	3.96	4.04	4.09	4.09
PE	30	7.69			
	40	7.37	8.72		
	50	6.90	7.80		
	60	6.68	6.55	7.70	
	70	6.02	6.16	7.07	7.41
	80	6.29	6.22	6.52	6.58
PG	30	3.94	4.14		
	40	3.89	4.15		
	50	3.87	3.97	4.16	
	60	3.91	3.90	3.92	4.06
	70	3.77	3.81	3.85	3.86
	80	3.79	3.77	3.80	3.82
PS	30	7.82			
	40	7.80	8.51		
	50	6.44	7.16	8.57	
	60	6.89	7.11	7.39	
	70	6.06	6.54	6.77	7.08
	80	5.59	6.18	6.39	6.59

Table A.30 The correlation time τ_2 (ns) of lipid wobble C_{22}-C_{32} of saturated lipids at varied temperatures and chain lengths.

	Tail	DL	DM	DP	DS
Head	T(°C)	12:0	14:0	16:0	18:0
PA	30	3.04			
	40	2.08	2.51		
	50	1.52	2.18	2.77	
	60	0.93	1.55	1.95	
	70	0.79	1.09	1.31	2.03
	80	0.63	0.77	1.12	1.46
PC	30	3.59	5.14		
	40	2.38	3.97		
	50	2.38			
	60		1.45		3.04
	70	0.91	1.35	1.76	2.25
	80	0.78	0.87	1.36	1.45
PE	30	3.71			
	40	2.85	5.13		
	50	2.00	2.87		
	60	1.68	3.23	3.05	
	70	0.98	1.33	3.14	2.40
	80	0.89	1.25	1.98	2.07
PG	30	3.32	5.60		
	40	2.28	5.22		
	50	1.54	2.45	2.76	
	60	1.13	1.56	1.91	2.45
	70	0.72	1.45	1.61	1.57
	80	0.57	0.72	1.30	1.57
PS	30	7.70			
	40	5.85	11.76		
	50	4.67	5.99	13.85	
	60	2.74	3.21	3.59	
	70	1.66	2.88	3.60	3.83
	80	1.16	1.80	2.79	3.74

Table A.31 The correlation time τ_2 (ns) of lipid wobble C22-C32 of mono-unsaturated lipids at varied temperatures and chain lengths.

	Tail	DX	DO	DG	DE	DN
Head	T(°C)	16:1	18:1	20:1	22:1	24:1
PA	30		8.75			
	40	2.79	3.90	3.39	9.29	9.41
PC	30	3.76	6.36	4.26	9.19	11.90
	40	2.50	4.68	4.50	7.75	4.61
PE	30	5.43	4.03	5.73	5.47	5.26
	40	5.56	6.07	5.08	5.73	7.42
PG	30		4.17			
	40	2.67	3.93	3.32	4.56	7.55
PS	30	7.16	7.10	12.35	16.13	21.20
	40	4.85	5.26	6.89	6.00	11.23

Table A.32 The correlation time τ_2 (ns) of lipid wobble C22-C32 of mixed and poly-unsaturated lipids at varied temperatures and chain lengths.

	Tail	PO	SO	PL	SL	SA	SD	DA
Head	T(°C)	16:0 /18:1	18:0 /18:1	16:0 /18:2	18:0 /18:2	18:0 /20:4	18:0 /22:6	20:4 /20:4
PA	30	5.11	7.26			4.38	3.86	4.18
	40	3.36	3.06			5.10	3.02	1.87
PC	30		6.53	4.39	6.67	6.46	4.02	3.55
	40	4.11	3.37	3.36	5.45	3.62	3.27	3.59
PE	30	6.81	10.65	7.46	6.85	4.66	8.23	3.24
	40	4.30	6.15	3.85	3.87	3.79	3.36	3.01
PG	30	5.23	9.77			3.57	9.77	3.15
	40	4.23	4.44			3.74	3.24	2.28
PS	30	11.03	15.09	9.27	13.40	7.19	11.28	6.78
	40	14.03	4.58	8.18	5.14	8.97	6.47	4.90

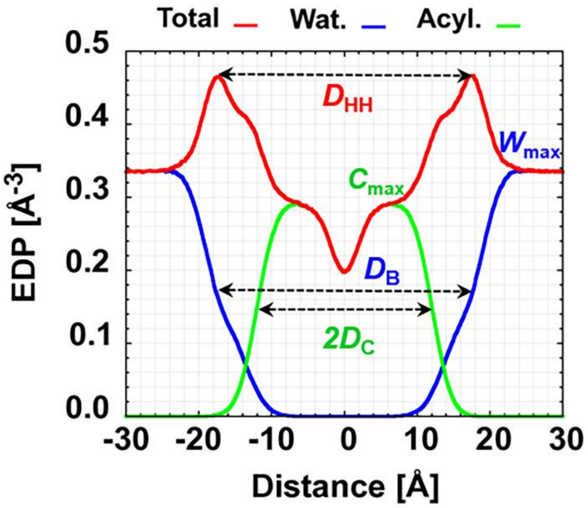

Figure A.1 Electron density profile (EDP) of DLPE at 30 °C. D_{HH} is the distance between two maximum total EDP peaks, D_B is the distance between the half of the maximum water EDP, $2D_C$ is the distance between the half of the maximum of acyl chain EDP.

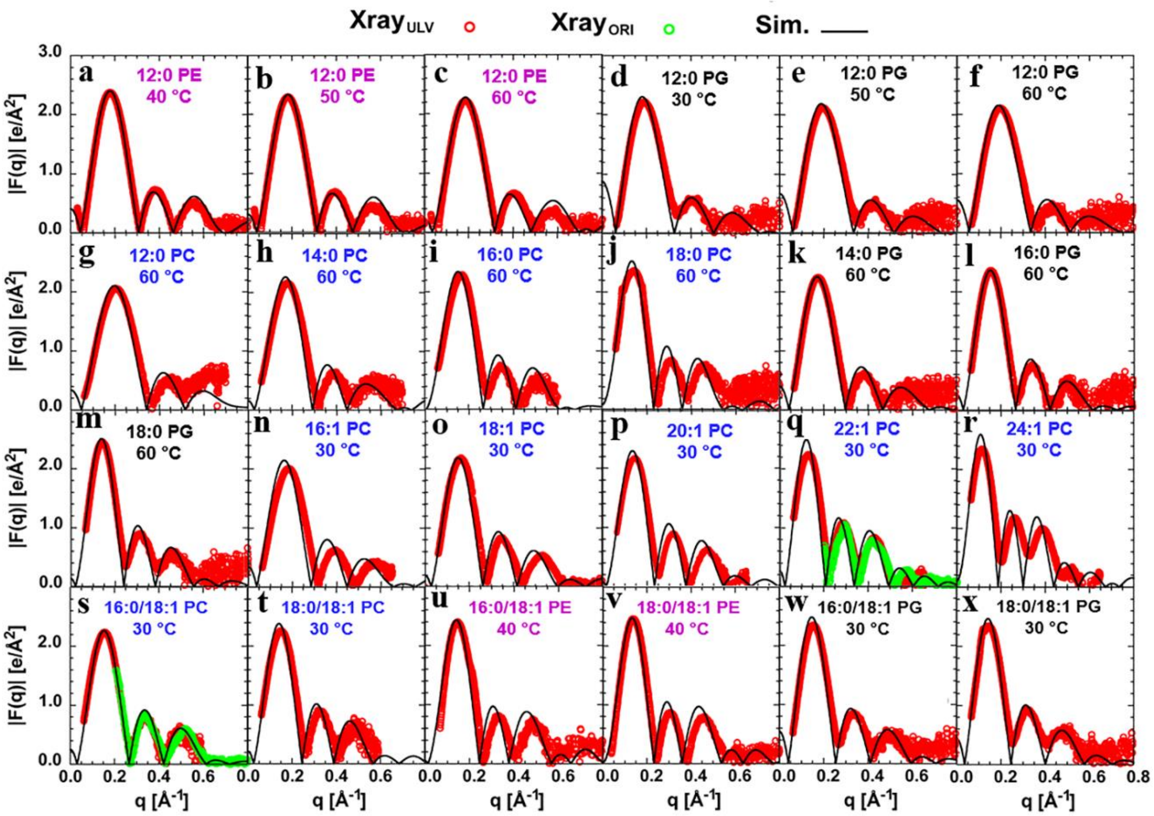

Figure A.2 Comparison of MD-based form factor data (the black curve) to experimental x-ray form factor (red circle) data with the unilamellar vesicle (red) and oriented bilayer sample (green) [6-8, 10] for a-c) DLPE at 40, 50, 60 °C. d-f) DLPG lipids at 30, 50, 60 °C. g-j) PC lipids at 60 °C with chain length 12, 14, 16, and 18. k-m) PG lipids at 60 °C with chain length 14, 16, and 18. n-r) mono-unsaturated PC lipids at 60 °C with chain length 16, 18, 20, 22 and 24. s-t) POPC and SOPC at 30 °C. u-v) POPE and SOPE at 40 °C w-x) POPG and SOPG at 30 °C.

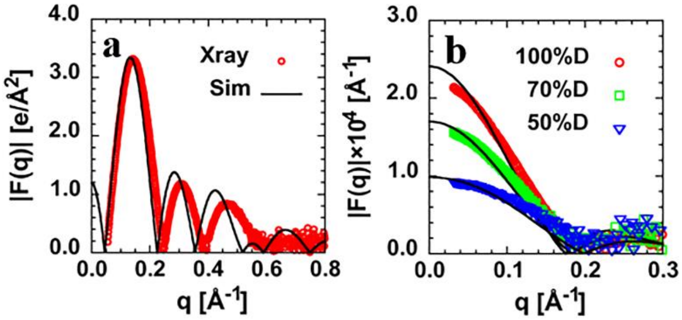

Figure A.3 Comparison of MD-based form factor data of POPS at 30 °C (the black curve) to the available experimental data at 25 °C [9] of a) x-ray form factor (red circle) b) deuterium neutron scattering data in three contrasts. %D is %D2O. The black curves are the MD-based form factor at each corresponded deuterium concentration.

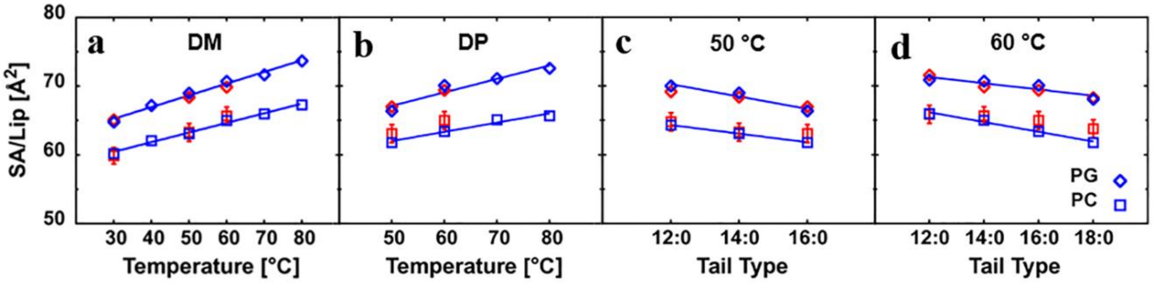

Figure A.4 Comparison of simulated SA/lip with experimental data [6, 8] at varied temperatures of a) DM lipids and b) DP lipids, and varied tails length at c) 50 °C and d) 60 °C of saturated lipids. The blue represents the simulation data, and the red represents the experimental data. Error bars show the uncertainty of experimental values of PC lipids. The error bars are not shown for PG lipids as their uncertainty is within symbol size.

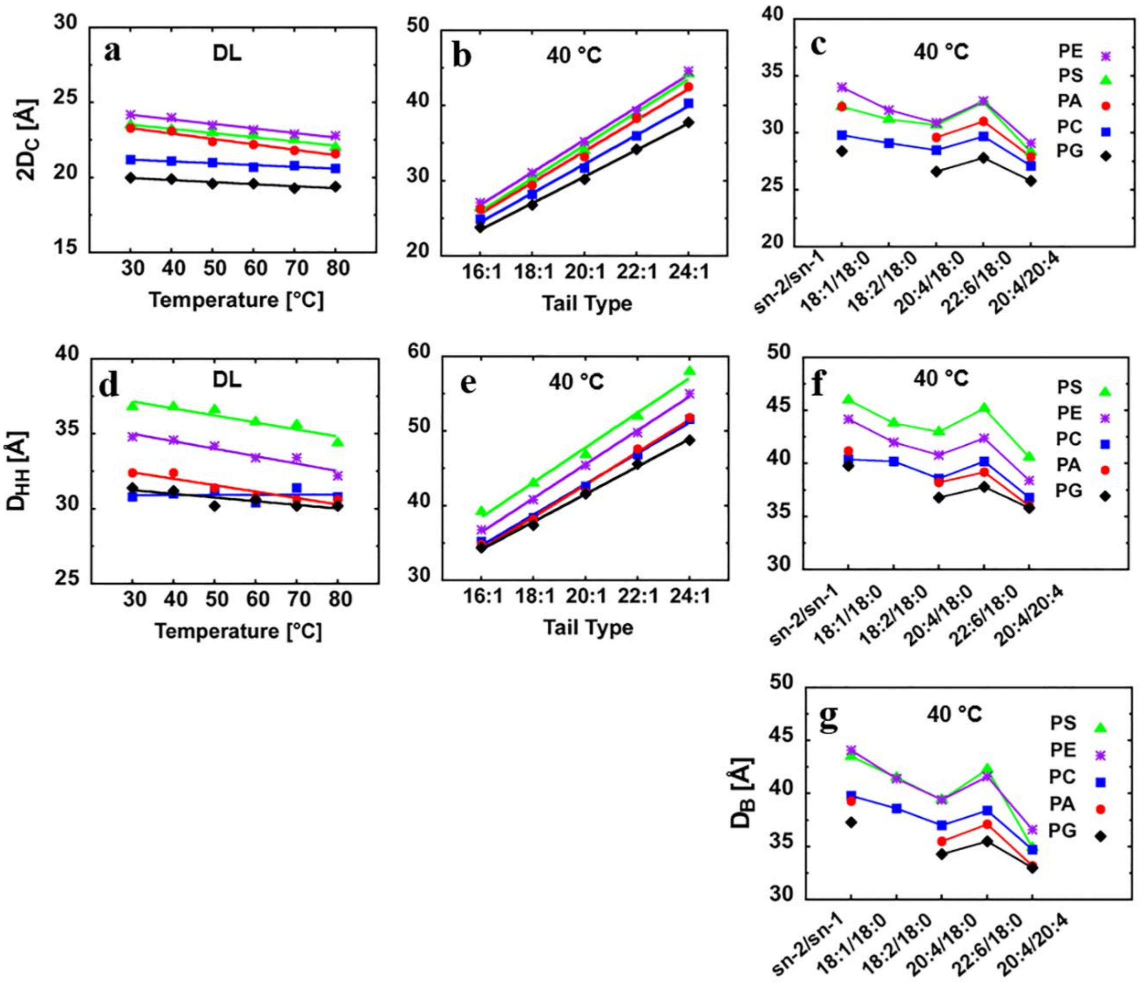

Figure A.5 The variations of 2*D*$_C$ by a) temperature b) mono-unsaturated lipid tails c) mixed and poly-unsaturated lipid tails, *D*$_{HH}$ by d) temperature e) mono-unsaturated lipid tails f) mixed and poly-unsaturated lipid tails g) *D*$_B$ by mixed and poly-unsaturated lipid tails.

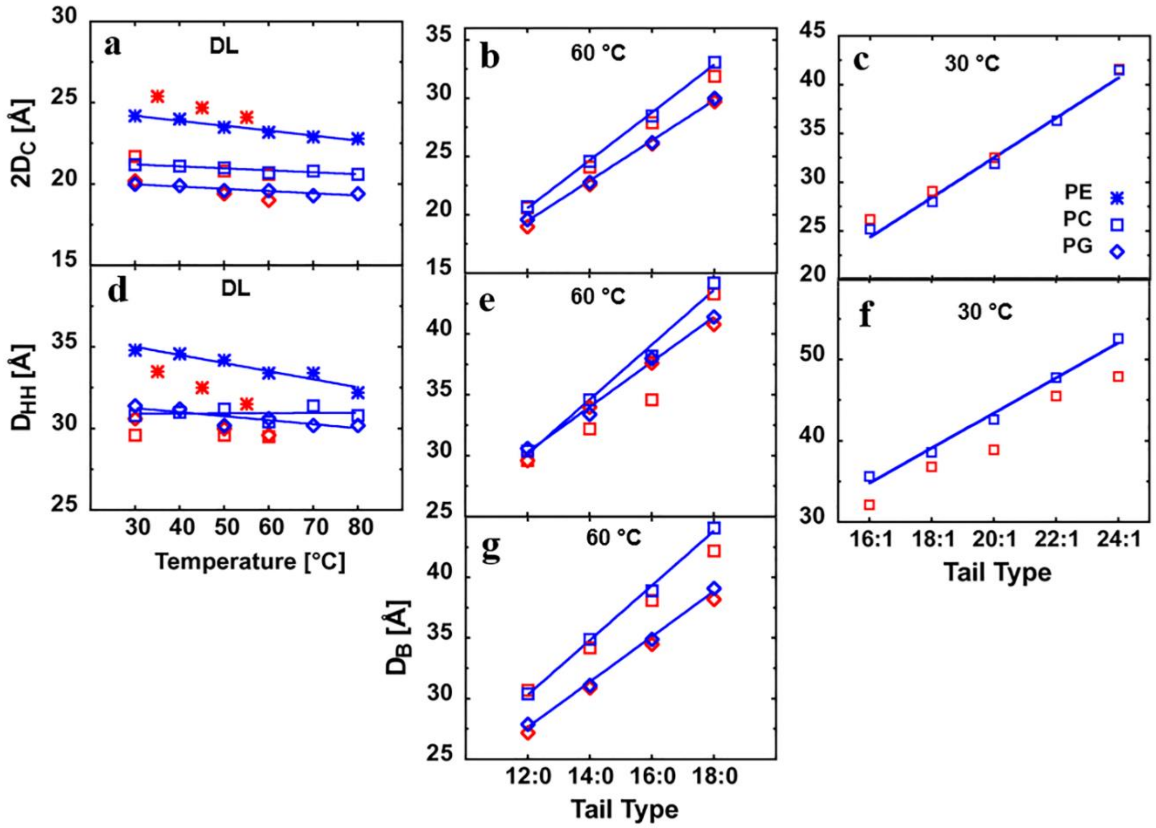

Figure A.6 Comparison of simulated thicknesses with experimental data [6-8, 10] of $2D_C$ with varied a) temperature b) saturated lipid tail length c) mono-unsaturated lipid tail length, D_{HH} with varied d) temperature e) saturated lipid tail length f) mono-unsaturated lipid tail length, and g) D_B with varied saturated lipid tail length.

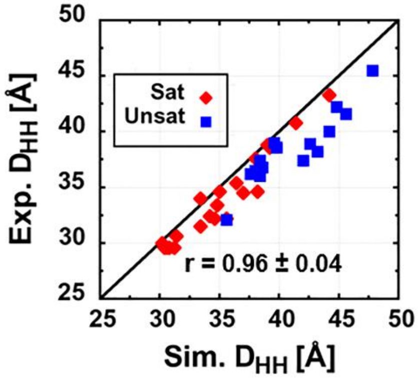

Figure A.7 Overall comparison of simulated and experimental data [6-10] of D_{HH}, shown with correlation coefficients and its standard deviation.

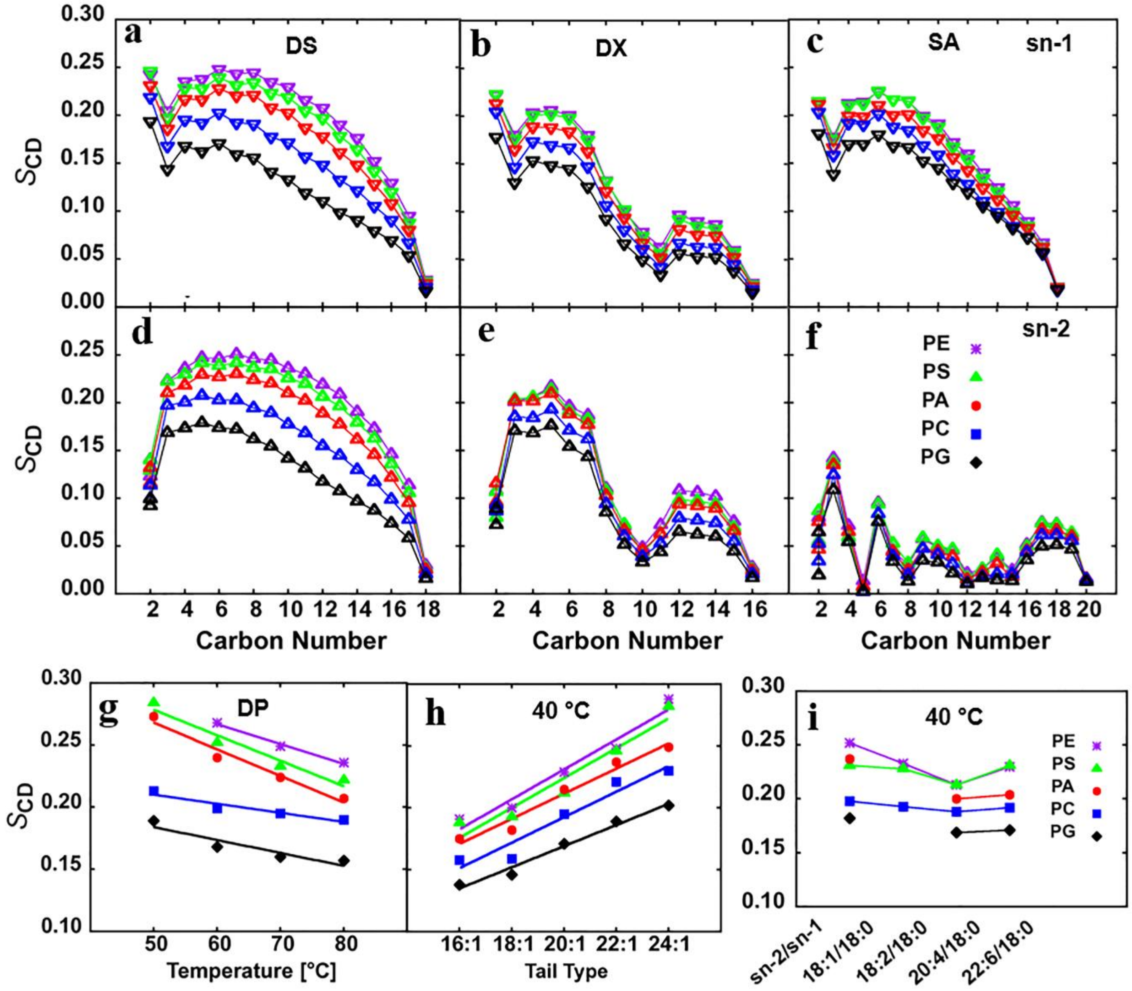

Figure A.8 The *sn*-1 S_{CD} with varied head groups of a) DS lipids b) DX lipids c) SA lipids, and *sn*-2 S_{CD} with varied head groups of d) DS lipids e) DX lipids f) SA lipids. The simulated C4-C7 averaged S_{CD} with varied saturated lipid tails length g) *sn*-1 with varied temperature h) *sn*-1 with varied mono-unsaturated lipid tail length. i) *sn*-1 with mixed and poly-unsaturated lipid tails.

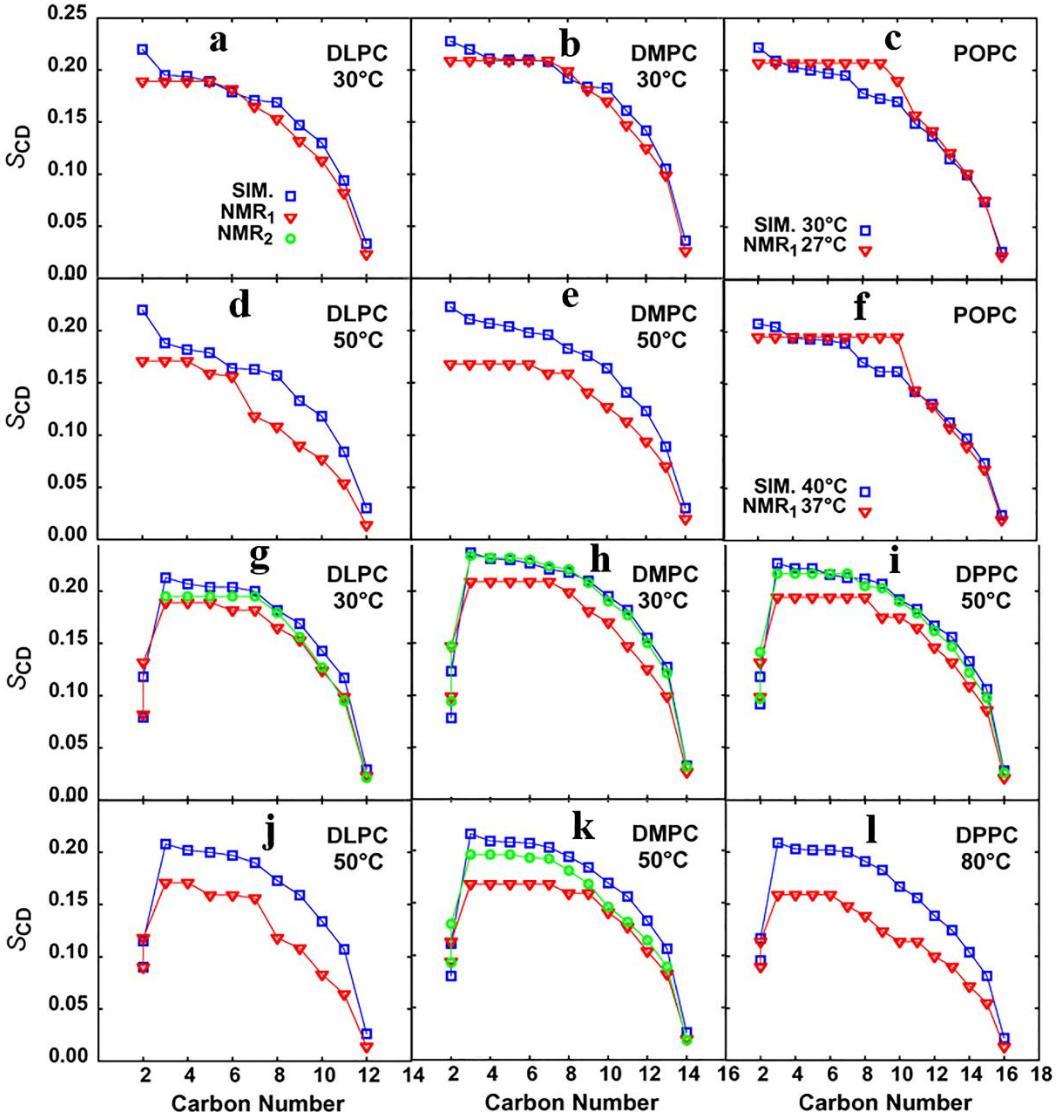

Figure A.9 Comparison of simulated S_{CD} with experimental data. NMR$_1$ [11] and NMR$_2$ [98] of *sn*-1 S_{CD} at relatively low temperatures, e.g. a) DLPC at 30 °C and b) DMPC at 30 °C, and c) POPC at 30 °C, and relatively high temperature, e.g. d) DLPC at 50 °C e) DMPC at 50 °C f) POPC at 40 °C. *sn*-2 S_{CD} with experimental data at relatively low temperatures, e.g. a) DLPC at 30 °C and b) DMPC at 30 °C, and c) DPPC at 50 °C, and relatively high temperature, e.g. d) DLPC at 50 °C e) DMPC at 50 °C f) DPPC at 80 °C. Both the simulation and experimental data are sorted from C2 to end for *sn*-1, and C3 to end for sn-2.

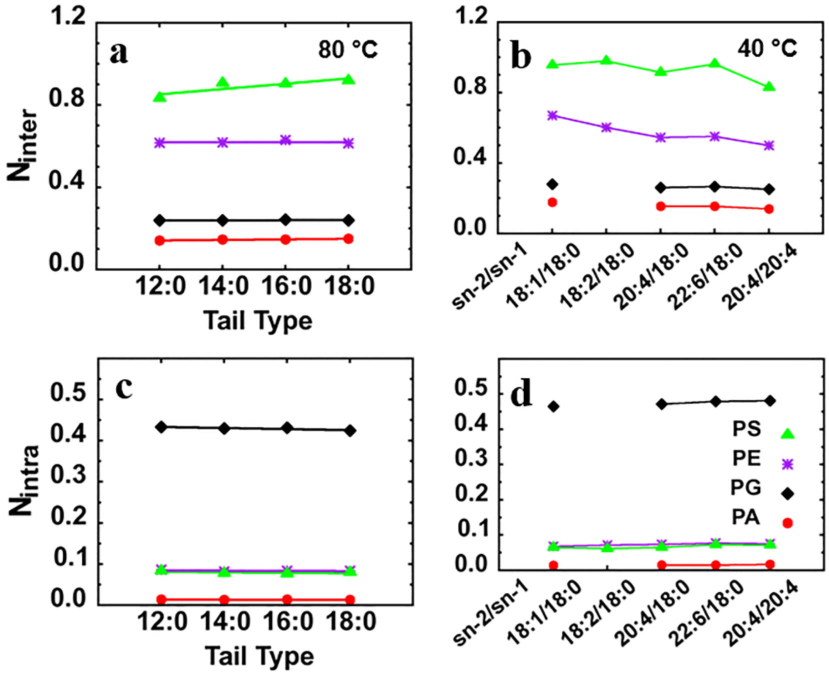

Figure A.10 The number of inter-lipid hydrogen bonds of varied a) tails of saturated lipids at 80 °C b) tails of mixed and poly-unsaturated lipids at 40 °C, and the number of intra-lipid hydrogen bonds of varied c) tails of saturated lipids at 80 °C d) tails of mixed and poly-unsaturated lipids at 40 °C.

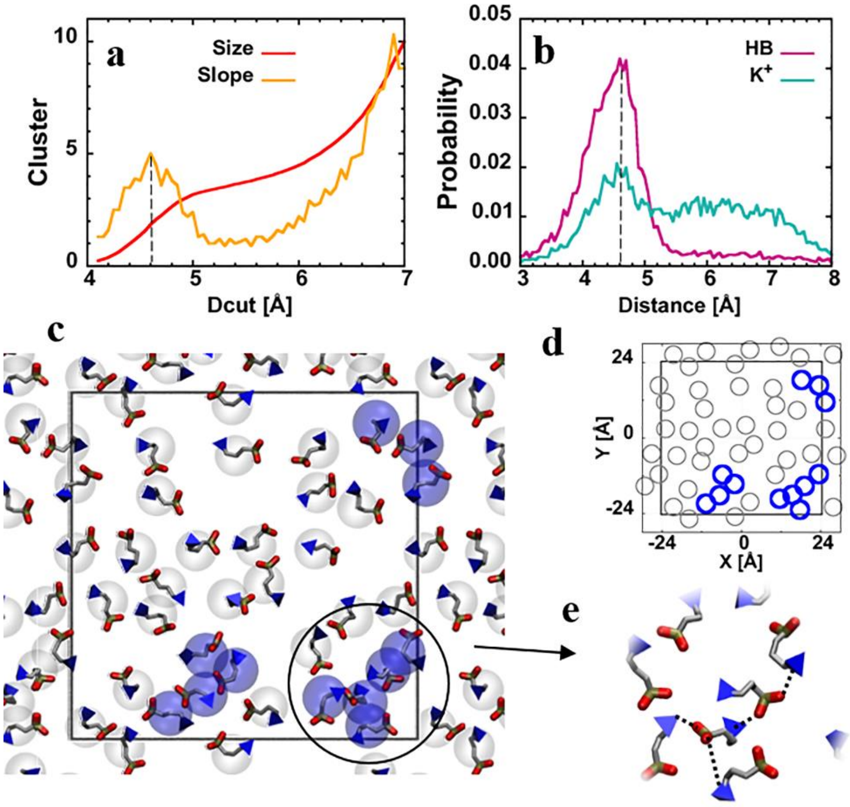

Figure A.11 a) The number of DLPA per cluster and its slope (the change of cluster size per Å) as function of cutoff distance b) The probability distribution of distance of DLPA head group atoms for lipids involved in hydrogen bonding and K$^+$-lipid interactions (considering the K$^+$ and O$^-$ within 4 Å**)** c) The snap shot of the top leaflet of DLPC at 30 °C at the end of simulation. The orange and red in licorice represent the phosphate, the blue triangle represents N$^+$, the blue sphere indicates the head group forms the cluster, and clear sphere indicates the ones that do not form the cluster. d) The schematic plot of clusters of DLPC at 30 °C, based on the center of mass of PC head group e) Zoom-in view of a DLPC cluster. The black dot lines indicate the possible electrostatic interactions.

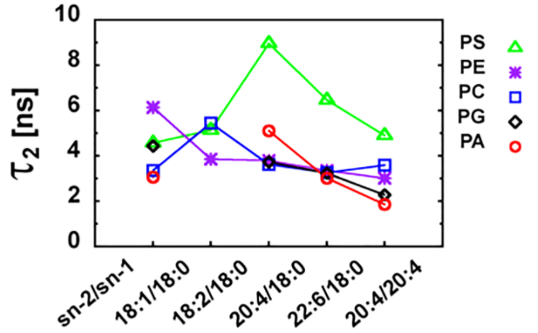

Figure A.12 The C_{22}-C_{32} orientation correlation time τ_2 of poly-unsaturated lipids at 40 °C.

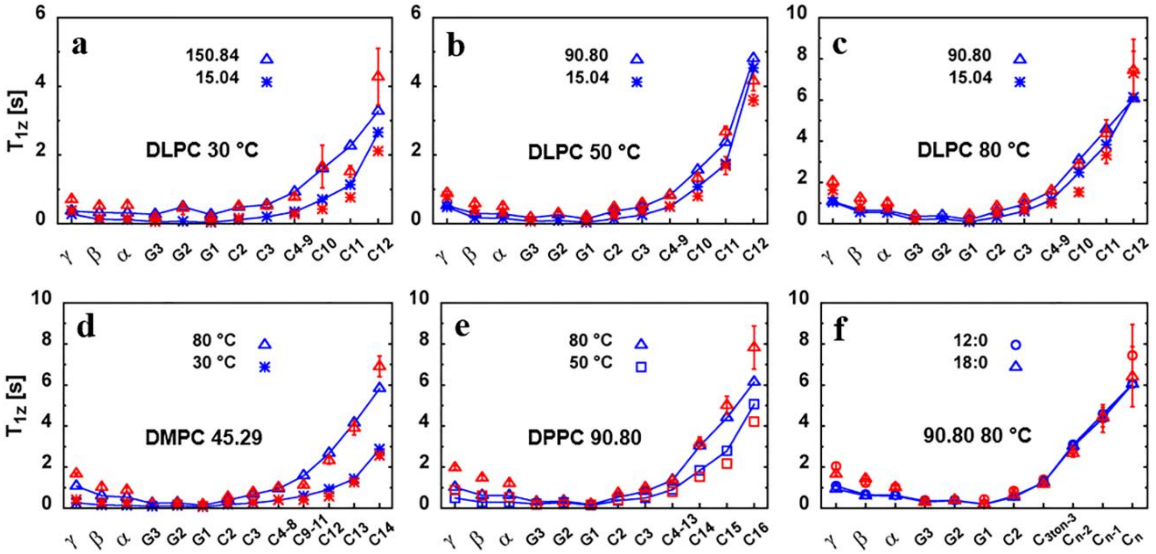

Figure A.13 The comparison of simulated spin-lattice relaxation time T_1 to experimental data [11] of a-c) DLPC at 30, 50, 80 °C with varied ω_C d) DMPC with ω_C = 45.29 MHz and e) DPPC with wc = 90.80 MHz at varied temperatures f) T_1 of DLPC and DSPC with ω_C = 90.80 MHz at 80 C. Uncertainty for lower frequencies and temperatures are not shown as they are less than the symbol size.

Appendix B

Table B.1 The comparison of DLiPC sorted simulated relaxation time T_1 (s) to the experimental data with hydrogen Larmor frequency ω_C=90.80 MHz at 30 °C.

Index	Sim	Exp
1	0.315 ± 0.004	0.272 ± 0.022
2	0.394 ± 0.004	0.385 ± 0.026
3	0.534 ± 0.007	0.512 ± 0.037
4	0.556 ± 0.005	0.593 ± 0.029
5	0.874 ± 0.008	0.759 ± 0.010
6	0.973 ± 0.011	0.964 ± 0.011
7	1.079 ± 0.006	0.964 ± 0.011
8	1.239 ± 0.014	0.987 ± 0.053
9	1.460 ± 0.028	1.015 ± 0.019
10	1.623 ± 0.021	1.687 ± 0.042
11	2.216 ± 0.021	2.264 ± 0.045
12	4.052 ± 0.038	4.292 ± 0.404

Table B.2 The comparison of DLiPC simulated relaxation time T_1 (s) of each carbon number to the corresponded original assigned experimental data with hydrogen Larmor frequency ω_C=90.80 MHz at 30 °C.

Carbon	Sim	Exp
C2	0.315 ± 0.004	0.272 ± 0.022
C3	0.394 ± 0.004	0.385 ± 0.026
C4-6	0.534 ± 0.007	0.987 ± 0.053
C7-8	0.556 ± 0.005	1.015± 0.019
C9	0.973 ± 0.011	0.964 ± 0.011
C10	0.877 ± 0.004	--
C11	0.874 ± 0.008	0.759 ± 0.010
C12	1.582 ± 0.008	--
C13	1.460 ± 0.028	0.964 ± 0.011
C14	1.079 ± 0.006	0.593 ± 0.029
C15	1.239 ± 0.014	0.512 ± 0.037
C16	1.623 ± 0.021	1.687 ± 0.042
C17	2.216 ± 0.021	2.264 ± 0.045
C18	4.052 ± 0.038	4.292 ± 0.404

Table B.3 Number of hydrogen bonds in soybean membranes.

Type	Hypocotyl			Root		
Intra-lipid	Lipids	N_{HB}	N_{HB}/lipid	Lipids	N_{HB}	N_{HB}/lipid
	PLPE	2.170 ± 0.720	0.083 ± 0.001	PLPA	0.330 ± 0.108	0.016 ± 0.001
	L2PE	1.088 ± 0.333	0.078 ± 0.006	LLPA	0.337 ± 0.119	0.017 ± 0.001
	LLPE	0.818 ± 0.275	0.082 ± 0.006	PLPE	1.457 ± 0.437	0.073 ± 0.007
	PLPI	4.636 ± 1.587	0.258 ± 0.01	LLPE	1.426 ± 0.474	0.071 ± 0.001
	L2PI	1.980 ± 0.654	0.248 ± 0.012	LLPS	0.561 ± 0.197	0.070 ± 0.005
				PLPI	2.382 ± 0.779	0.238 ± 0.006
Inter-lipid	PLPC	9.621 ± 3.419	0.401 ± 0.028	PLPA	12.489 ± 4.086	0.624 ± 0.013
	L2PC	17.110 ± 5.573	0.389 ± 0.011	LLPA	12.288 ± 3.933	0.614 ± 0.022
	PLPE	20.773 ± 6.818	0.799 ± 0.01	PLPC	7.446 ± 2.528	0.338 ± 0.011
	L2PE	11.598 ± 3.964	0.828 ± 0.039	LLPC	7.051 ± 2.468	0.321 ± 0.017
	LLPE	8.345 ± 2.828	0.835 ± 0.031	PLPE	18.200 ± 6.149	0.910 ± 0.017
	PLPI	18.540 ± 6.184	1.030 ± 0.036	LLPE	19.094 ± 6.499	0.955 ± 0.024
	L2PI	8.660 ± 2.871	1.082 ± 0.013	LLPS	9.222 ± 3.034	1.153 ± 0.024
	SITO	10.052 ± 3.388	0.239 ± 0.004	PLPI	9.683 ± 3.090	0.968 ± 0.037
	STIG	3.315 ± 1.116	0.237 ± 0.006	SITO	8.975 ± 2.952	0.236 ± 0.004
				STIG	4.602 ± 1.525	0.230 ± 0.0002
Overall	PLPC	9.621 ± 3.419	0.401 ± 0.028	PLPA	12.819 ± 4.194	0.640 ± 0.014
	L2PC	17.110 ± 5.573	0.389 ± 0.011	LLPA	12.625 ± 4.052	0.631 ± 0.023
	PLPE	22.943 ± 7.538	0.882 ± 0.011	PLPC	7.446 ± 2.528	0.338 ± 0.011
	L2PE	12.686 ± 4.297	0.906 ± 0.045	LLPC	7.051 ± 2.468	0.321 ± 0.017
	LLPE	9.163 ± 3.103	0.917 ± 0.037	PLPE	19.657 ± 6.586	0.983 ± 0.024
	PLPI	23.176 ± 7.771	1.288 ± 0.046	LLPE	20.520 ± 6.973	1.026 ± 0.025
	L2PI	10.640 ± 3.525	1.330 ± 0.025	LLPS	9.783 ± 3.231	1.223 ± 0.029
	SITO	10.052 ± 3.388	0.239 ± 0.004	PLPI	12.065 ± 3.869	1.206 ± 0.043
	STIG	3.315 ± 1.116	0.237 ± 0.006	SITO	8.975 ± 2.952	0.236 ± 0.004
				STIG	4.602 ± 1.525	0.230 ± 0.0002

Table B.4 Clustering fractions of soybean membranes. Yc is the fraction of lipids forming the cluster, Xo is the overall composition fraction, and Xc is the average fraction of lipid in the cluster.

Hypocotyl					Root				
Lipids	Yc	Xo	Xc	Xc-Xo	Lipids	Yc	Xo	Xc	Xc-Xo
SITO	0.684	0.210	0.276	0.066	SITO	0.660	0.190	0.236	0.046
STIG	0.701	0.070	0.095	0.025	STIG	0.644	0.100	0.121	0.021
PLPC	0.416	0.120	0.096	-0.024	PLPA	0.507	0.100	0.095	-0.01
PLPE	0.465	0.130	0.117	-0.013	PLPC	0.456	0.110	0.094	-0.02
PLPI	0.530	0.090	0.092	0.002	PLPE	0.517	0.100	0.097	-0.003
LLPE	0.522	0.050	0.05	0.000	PLPI	0.514	0.050	0.048	-0.002
DLiPC	0.407	0.220	0.172	-0.048	LLPA	0.440	0.100	0.083	-0.02
DLiPE	0.480	0.070	0.065	-0.005	LLPC	0.440	0.110	0.091	-0.02
DLiPI	0.486	0.040	0.037	-0.003	LLPE	0.507	0.100	0.095	-0.01
					LLPS	0.524	0.040	0.039	-0.001

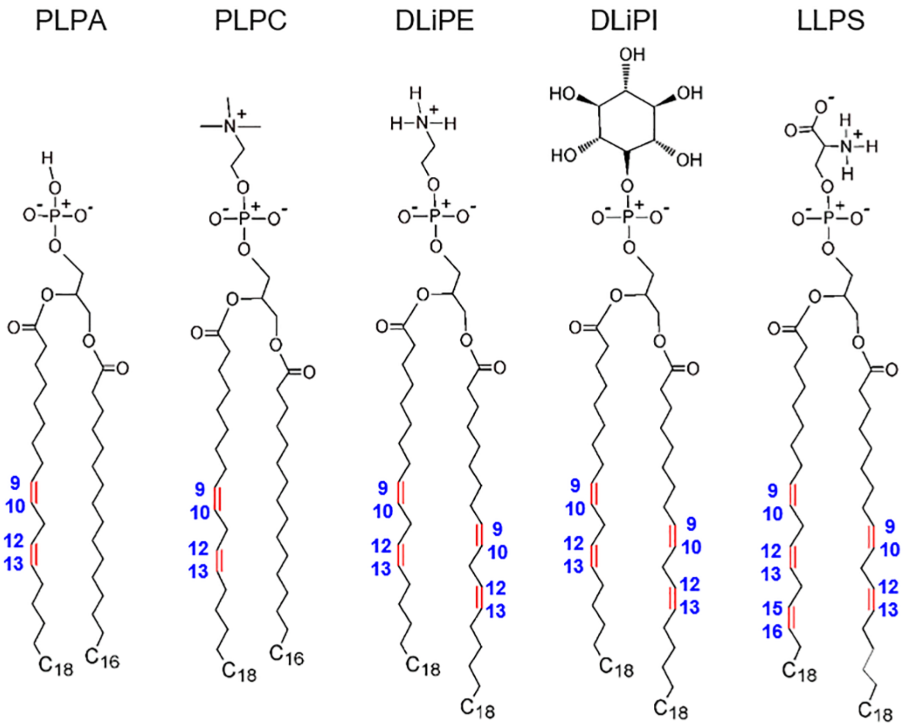

Figure B.1 Chemical structures of a few exemplary and lipids with the various head group and tail types in the two soybean membrane systems. The double bond is shown in red and the positions are shown in blue texts.

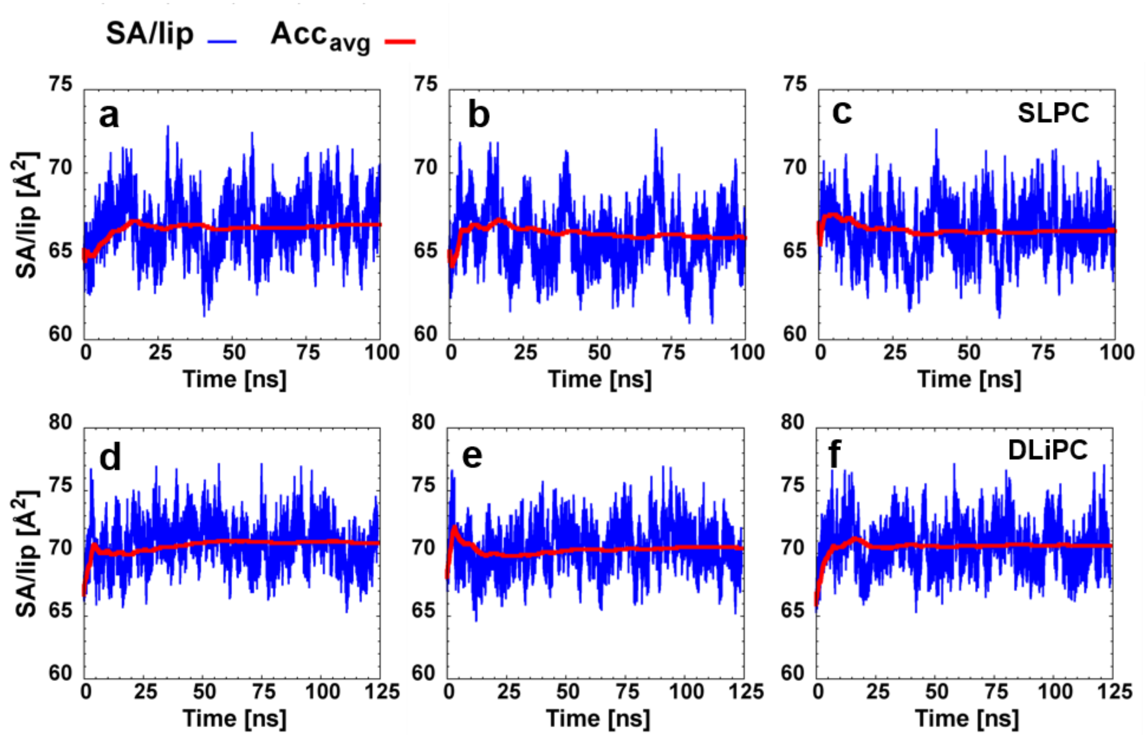

Figure B.2 Surface area per lipid and the accumulative average (Å^2) as a function of time of the three replicate runs of a-c) SLPC, and d-f) DLiPC.

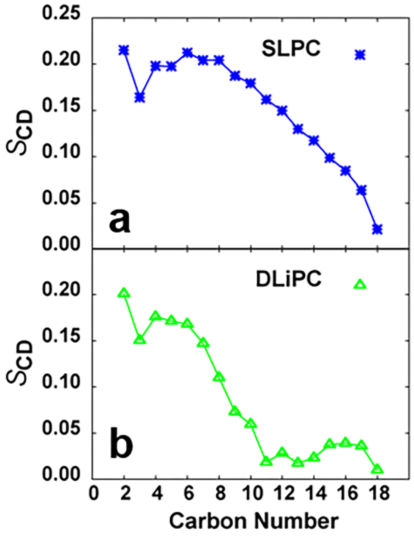

Figure B.3 The averaged deuterium order parameter (S_{CD}) of *sn*-1 chain for a) SLPC and b) DLiPC.

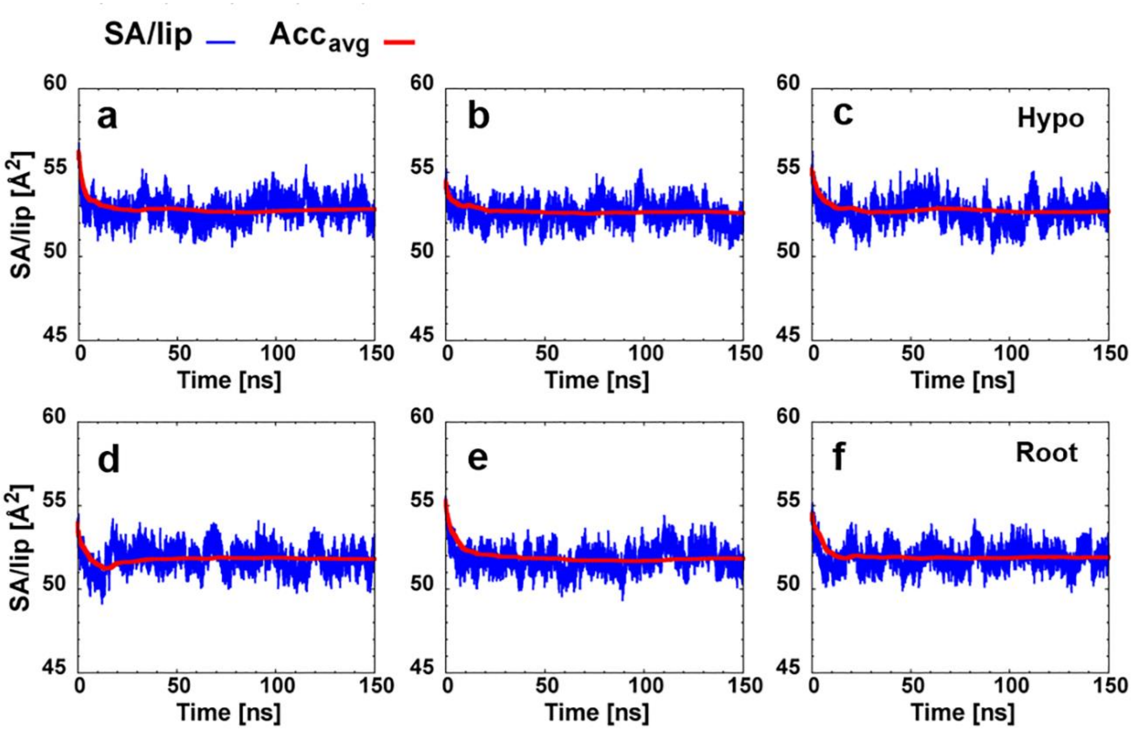

Figure B.4 Surface area per lipid and the accumulative average (Å^2) as a function of time of the three replicate runs of a-c) hypocotyl and d-f) root membranes.

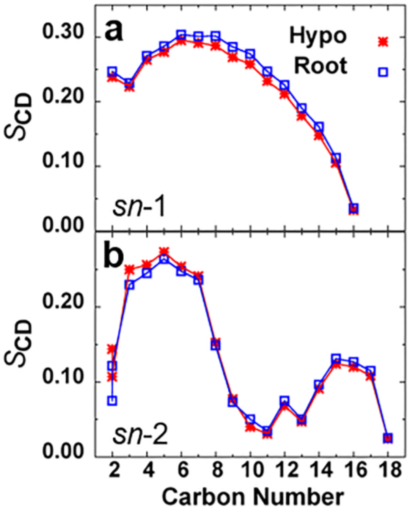

Figure B.5 The S_{CD} of PLPC in hypocotyl and root membrane in a) *sn*-1 and b) *sn*-2 chain.

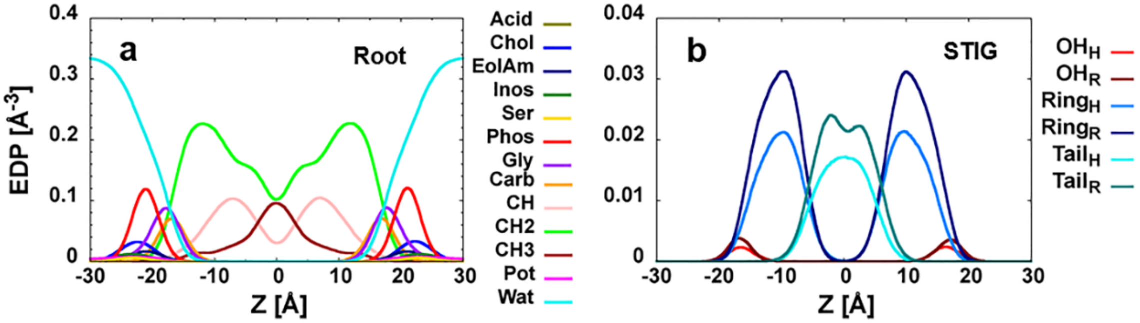

Figure B.6 a) The component electron density profiles (EDP) of root membrane, which includes choline (Chol), ethanolamine (EolAm), inositol (Inos), serine (Ser), phosphate (Phos), glycerol (Gly), carbonyl (Carb), methine (CH), methylene (CH2), methyl (CH3), potassium ion (Pot), and water (Wat). b) The group EDP of stigmasterol, which includes OH, ring, and tail in hypocotyl and root membrane.

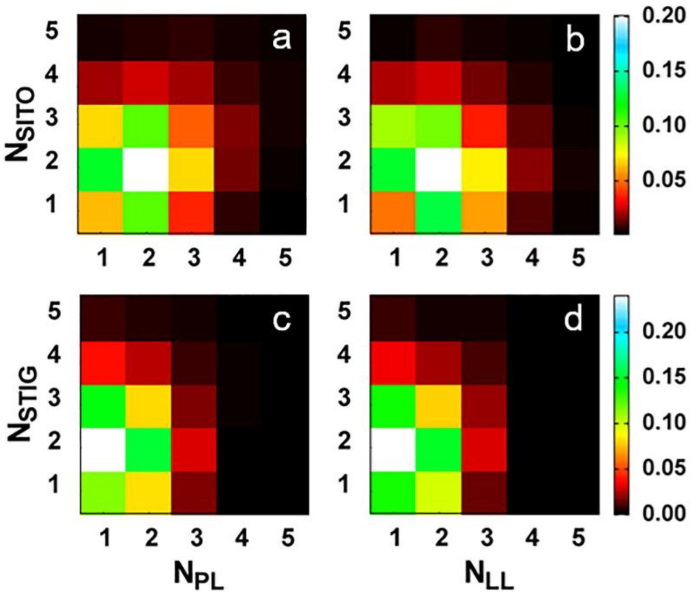

Figure B.7 Cluster composition of sterols and glycerol phosphate lipids in the root membrane.

Appendix C

Table C.1 Systems of LacY^{IM-EX} with three protonation states in the explicit membrane.

Protonation	# Lipids Top./Bot.	# Regular Waters	# Pore Water	# K^+ Ions	# Total Atoms
E269	151/154	19038	91	51	101381
H322	151/154	19034	96	51	101384
E325	151/154	19037	88	51	101369

Table C.2 Molecular dynamics dummy spin-labels (MDDS) simulations [88] resulted main peaks of structural averages of spin label mean distances (d_{ON}) of all residue pairs of LacY compared to C_α distances ($d_{C\alpha}$) in Å. The different states of structures from the simulation are based on pore radii. The standard errors are calculated based on six or ten structures of each state. The mismatch of 1 Å for S136/Q340 and N137/Q340 in 6th columns and 5th columns are due to round off in the 5th column.

Structures	Inward		Occluded		Outward		Outw. -Inw.		Diff.
Pairs	$d_{C\alpha}$	d_{ON}	$d_{C\alpha}$	d_{ON}	$d_{C\alpha}$	d_{ON}	$\Delta d_{C\alpha}$	Δd_{ON}	Δd_{ON} - $\Delta d_{C\alpha}$
V105/T310	40±0.8	41±1.4	38±0.4	47±0.8	43±0.6	50±0.7	3	9	6
I164/T310	32±0.8	37±1.9	30±0.4	37±0.7	30±0.6	40±1.0	-2	2	4
I164/S375	37±1.1	46±1.1	36±0.4	49±0.4	39±0.7	54±1.0	3	8	5
R73/S401	45±0.5	49±2.0	43±0.5	53±1.3	35±0.2	42±0.5	-10	-7	3
R73/Q340	34±0.5	43±1.5	33±0.4	46±0.6	28±0.2	42±0.3	-6	-2	4
S136/Q340	35±0.7	47±0.7	27±0.3	39±0.3	24±0.3	36±0.4	-10	-11	0
N137/Q340	32±0.7	36±0.5	28±0.3	33±0.9	23±0.3	26±0.4	-10	-10	-1
S136/S401	43±0.7	45±1.3	36±0.5	39±1.0	31±0.2	38±0.5	-12	-7	5
N137/S401	40±0.6	39±1.4	36±0.5	35±1.1	29±0.3	27±0.4	-11	-12	-1

Table C.3 All peaks of structurally averaged spin label distances of nine residue pairs of LacY (with the bold font indicating the main distance peak) obtained from MDDS simulations and DEER data [62] in Å.

	MDDS			DEER [62]
Pairs	d_{IW}	d_{OW}	Δd (d_{OW}- d_{IW})	Δd (d_{OW} - d_{IW})
V105/T310	**41**	45 **50**	4 **9**	5 14
I164/T310	**37** 43	35 **40**	-3 -8 **2** -4	10 14
I164/S375	**46** 50	49 **54**	4 -1 **8** 3	4 12
R73/S401	37 43 **49**	42	5 -1 **-7**	-21 -16
R73/Q340	35 49 **43**	39 **42**	3 -10 6 -8 -5 **-2**	-14
S136/Q340	44 **47**	33 **36**	**-11** -14 -7	-11
N137/Q340	**36**	26	**-10**	-18
S136/4S01	**45** 50 55	38	**-7** -13 -17	-4
N137/S401	33 **39**	27	-5 **-12**	-13

Table C.4 The correlation coefficient among the main peaks the structurally averaged distance values for nine residue pairs of LacY for C_α-C_α distance ($D_{C\alpha}$) and MDDS interlabel distances (D_{ON}), C_α-C_α distance ($D_{C\alpha}$) and DEER interlabel distances, and MDDS (D_{ON}) and DEER interlabel distances of inward-facing and outward-facing LacY.

States	$D_{C\alpha}$-D_{ON}	$D_{C\alpha}$-DEER	D_{ON}-DEER
iw+ow	0.76±0.16	0.76±0.16	0.70±0.18
iw	0.66±0.19	0.67±0.19	0.70±0.18
ow	0.89±0.11	0.73±0.17	0.75±0.17

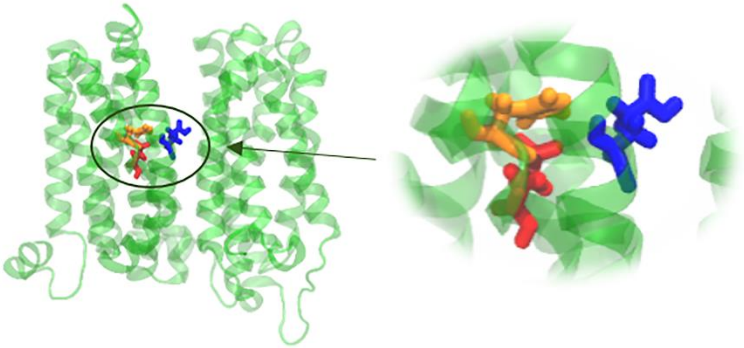

Figure C.1 A side view snapshot of three protonated residues in the built LacY^{IM-EX} in the explicit membrane system, with E269 (blue), H322 (orange) and E325 (red). The N- and C- domain are flipped for a better view of the three residues.

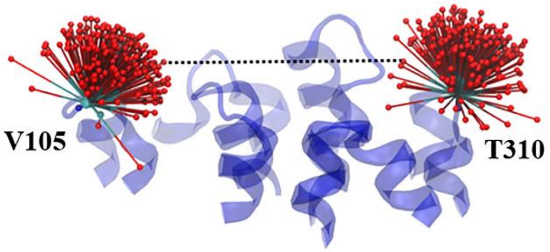

Figure C.2 The snapshot that shows the movement of the dummy ON spin labels on residue V105 and T310 of inward-facing LacY during MDDS simulation. Only 200 out of 20000 frames are shown for clarity

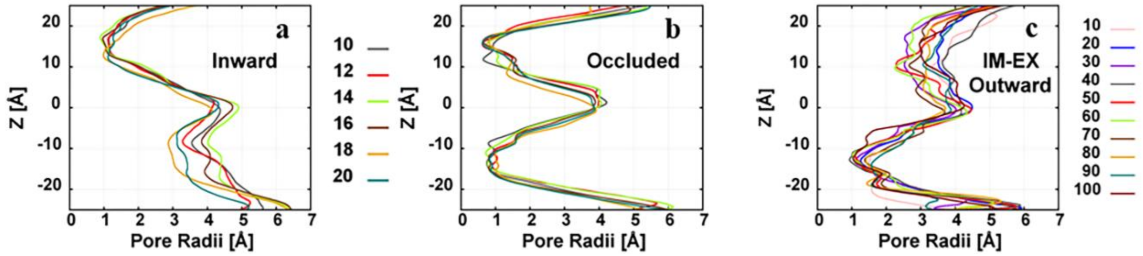

Figure C.3 LacY with E269 protonated, a) inward-facing from equilibrated six time frames (ns) from implicit simulation b) occluded from equilibrated six time frames (ns) from implicit simulation c) outward-facing LacY$^{IM\text{-}EX}$ from equilibrated ten time frames (ns) from the explicit simulation. Z in the plot represents the z-coordinate of the axis along the bilayer normal. Positive z corresponds to the periplasmic side of Lacy, and negative z corresponds to the cytoplasmic side of LacY.

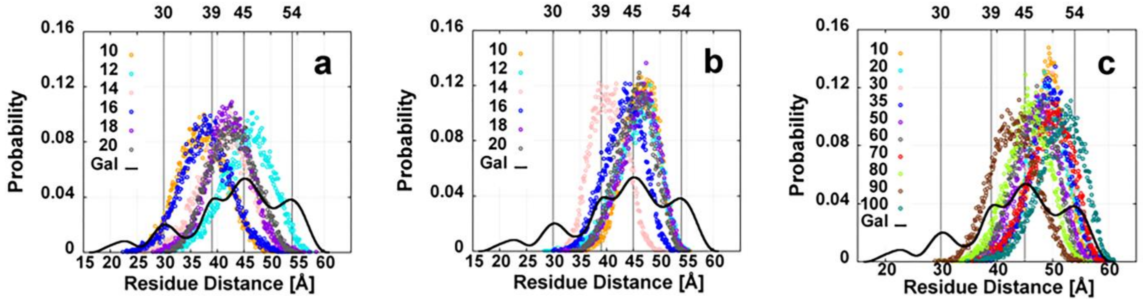

Figure C.4 The raw MDDS resulted residue distances of pair V105/T310 on the periplasmic side from a) inward-facing b) occluded and c) outward-facing LacY compared to DEER data of LacY with binding sugar [62]. The black curve is DEER data of LacY with galactosides, the black vertical lines are DEER mean distance peak(s).

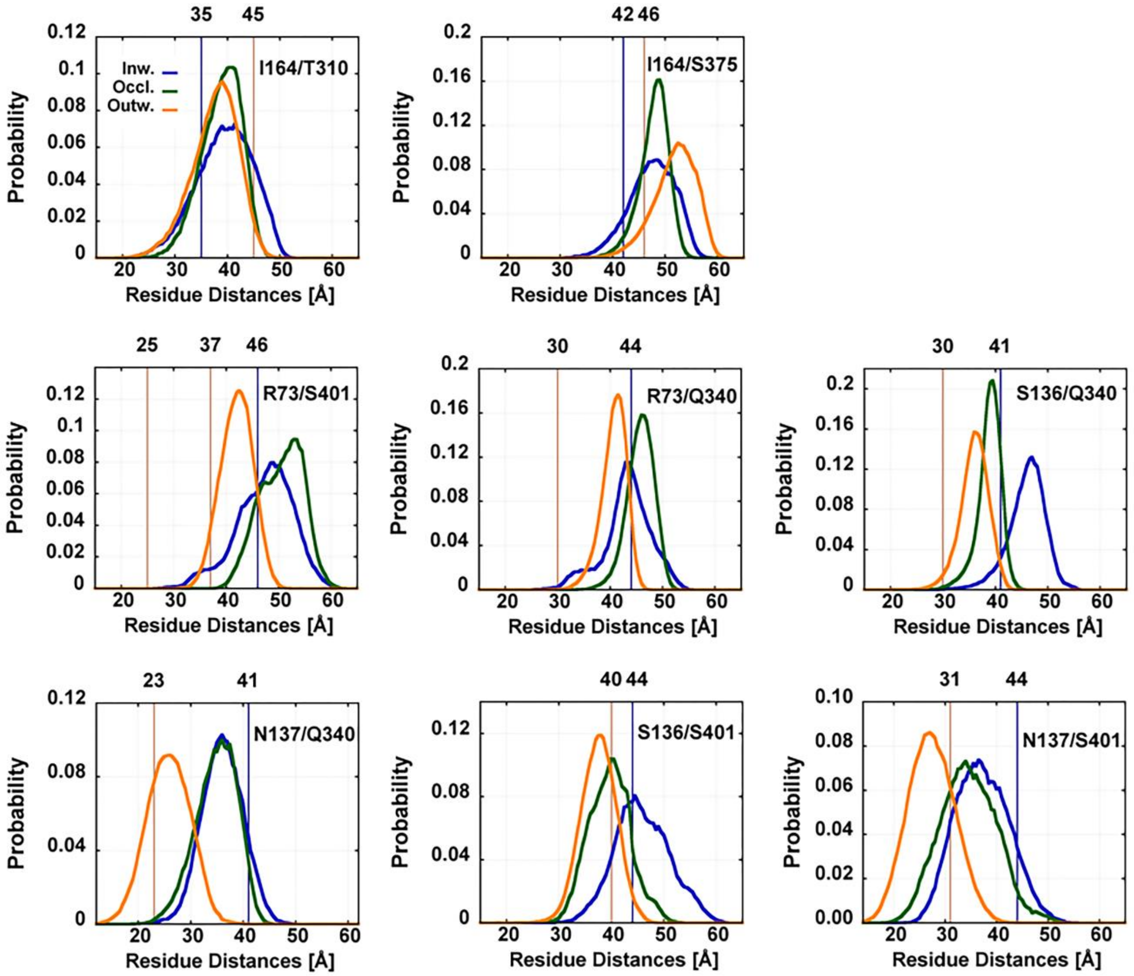

Figure C.5 a) MDDS resulted paired spin label distance distribution of overall structural averaged inward-facing, occluded and outward-facing LacY compared to DEER distance distribution main mean peaks [62]. The solid vertical lines are the DEER mean distances that are interpreted to be the spin label distances for the inward-facing and outward-facing states of LacY.

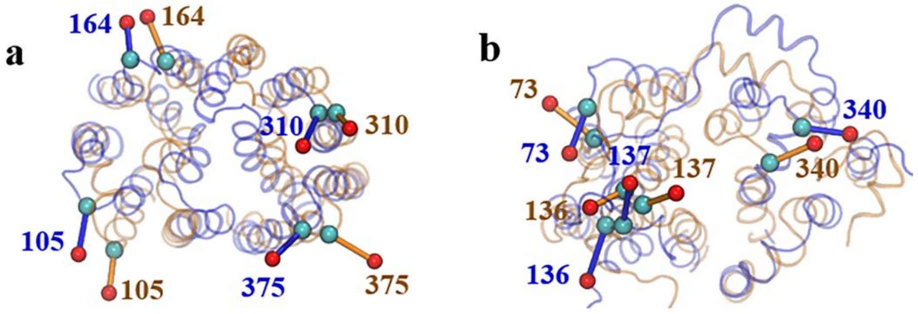

Figure C.6 The top view on the a) periplasmic side and b) cytoplasmic side of LacY that show the orientation of spin labels in the averaged mean distance.

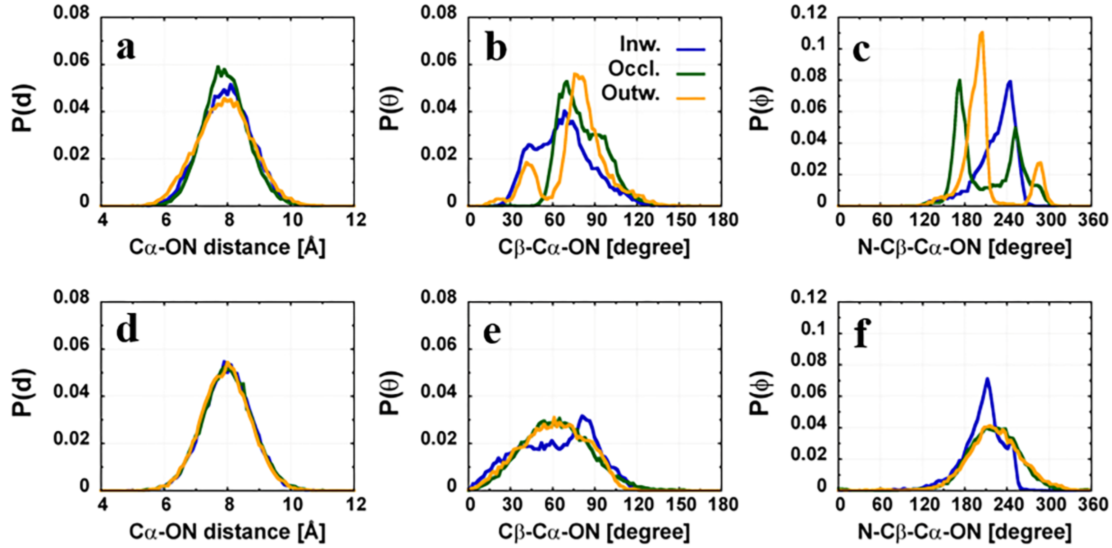

Figure C.7 Comparison of distribution of the C_α-ON distances, C_β-C_α-ON bond angles, and N-C_β-C_α-ON dihedral angles obtained from MDDS simulations in inward-facing(blue), occluded (green), and outward-facing states (orange) for residue S136 (a-c) and N137 (d-f)

References

[1] D.L. Nelson, A.L. Lehninger, M.M. Cox, Lehninger Principles of Biochemistry, W. H. Freeman, 2008.

[2] P.L. Yeagle, Lipid regulation of cell-membrane structure and function, Faseb Journal, 3 (1989) 1833-1842.

[3] L.S. Vermeer, B.L. de Groot, V. Reat, A. Milon, J. Czaplicki, Acyl chain order parameter profiles in phospholipid bilayers: computation from molecular dynamics simulations and comparison with H-2 NMR experiments, Eur Biophys J Biophy, 36 (2007) 919-931.

[4] M.J. Ueda, T. Ito, T.S. Okada, S.I. Ohnishi, A correlation between membrane fluidity and the critical temperature for cell adhesion, The Journal of cell biology, 71 (1976) 670-674.

[5] P. Yeagle, The membranes of cells, Academic Press, 1993.

[6] N. Kucerka, M.P. Nieh, J. Katsaras, Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature, Bba-Biomembranes, 1808 (2011) 2761-2771.

[7] N. Kucerka, B. van Oosten, J.J. Pan, F.A. Heberle, T.A. Harroun, J. Katsaras, Molecular Structures of Fluid Phosphatidylethanolamine Bilayers Obtained from Simulation-to-Experiment Comparisons and Experimental Scattering Density Profiles, J Phys Chem B, 119 (2015) 1947-1956.

[8] J.J. Pan, F.A. Heberle, S. Tristram-Nagle, M. Szymanski, M. Koepfinger, J. Katsaras, N. Kucerka, Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering, Bba-Biomembranes, 1818 (2012) 2135-2148.

[9] J.J. Pan, X.L. Cheng, L. Monticelli, F.A. Heberle, N. Kucerka, D.P. Tieleman, J. Katsaras, The molecular structure of a phosphatidylserine bilayer determined by scattering and molecular dynamics simulations, Soft Matter, 10 (2014) 3716-3725.

[10] N. Kucerka, J. Gallova, D. Uhrikova, P. Balgavy, M. Bulacu, S.J. Marrink, J. Katsaras, Areas of Monounsaturated Diacylphosphatidylcholines, Biophys J, 97 (2009) 1926-1932.

[11] A. Leftin, M.F. Brown, An NMR database for simulations of membrane dynamics, Bba-Biomembranes, 1808 (2011) 818-839.

[12] J.F. Nagle, R.T. Zhang, S. TristramNagle, W.J. Sun, H.I. Petrache, R.M. Suter, X-ray structure determination of fully hydrated L(alpha) phase dipalmitoylphosphatidylcholine bilayers, Biophys J, 70 (1996) 1419-1431.

[13] H.I. Petrache, K.C. Tu, J.F. Nagle, Analysis of simulated NMR order parameters for lipid bilayer structure determination, Biophys J, 76 (1999) 2479-2487.

[14] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, C. Mondragon-Ramirez, I. Vorobyov, D.J. Tobias, A.D. MacKerell, R.W. Pastor, Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types., J. Phys. Chem. B, 114 (2010) 7830-7843.

[15] J.B. Klauda, V. Monje, T. Kim, W. Im, Improving the CHARMM Force Field for Polyunsaturated Fatty Acid Chains, J Phys Chem B, 116 (2012) 9424-9431.

[16] J.B. Lim, B. Rogaski, J.B. Klauda, Update of the Cholesterol Force Field Parameters in CHARMM, J Phys Chem B, 116 (2012) 203-210.

[17] C.M. MacDermaid, H.K. Kashyap, R.H. DeVane, W. Shinoda, J.B. Klauda, M.L. Klein, G. Fiorin, Molecular dynamics simulations of cholesterol-rich membranes using a coarse-grained force field for cyclic alkanes, J. Chem. Phys., 143 (2015) 243144.

[18] E.L. Wu, Y.F. Qi, S. Park, S.S. Mallajosyula, A.D. MacKerell, J.B. Klauda, W. Im, Insight into Early-Stage Unfolding of GPI-Anchored Human Prion Protein, Biophys J, 109 (2015) 2090-2100.

[19] B. Rogaski, J.B. Klauda, Membrane-Binding Mechanism of a Peripheral Membrane Protein through Microsecond Molecular Dynamics Simulations, Journal of Molecular Biology, 423 (2012) 847-861.

[20] V. Monje-Galvan, J.B. Klauda, Peripheral membrane proteins: Tying the knot between experiment and computation, Biochim. Biophys. Acta, Biomembr., 1858 (2016) 1584-1593.

[21] X.H. Zhuang, J.B. Klauda, Modeling structural transitions from the periplasmic-open state of lactose permease and interpretations of spin label experiments, Bba-Biomembranes, 1858 (2016) 1541-1552.

[22] X. Cheng, S. Jo, H.S. Lee, J.B. Klauda, W. Im, CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems, Journal of Chemical Information and Modeling, 53 (2013) 2171-2180.

[23] X. Cheng, J.K. Kim, Y. Kim, J.U. Bowie, W. Im, Molecular dynamics simulation strategies for protein-micelle complexes, Bba-Biomembranes, 1858 (2016) 1566-1572.

[24] E.L. Wu, O. Engstrom, S. Jo, D. Stuhlsatz, M.S. Yeom, J.B. Klauda, G. Widmalm, W. Im, Molecular Dynamics and NMR Spectroscopy Studies of *E. coli* Lipopolysaccharide Structure and Dynamics, Biophys J, 105 (2013) 1444-1455.

[25] E.L. Wu, P.J. Fleming, M.S. Yeom, G. Widmalm, J.B. Klauda, K.G. Fleming, W. Im, *E. coli* Outer Membrane and Interactions with OmpLA, Biophys J, 106 (2014) 2493-2502.

[26] P. Khakbaz, J.B. Klauda, Probing the importance of lipid diversity in cell membranes via molecular simulation, Chemistry and Physics of Lipids, 192 (2015) 12-22.

[27] K.R. Pandit, J.B. Klauda, Membrane models of E-coli containing cyclic moieties in the aliphatic lipid chain, Bba-Biomembranes, 1818 (2012) 1205-1210.

[28] X.H. Zhuang, J.R. Makover, W. Im, J.B. Klauda, A systematic molecular dynamics simulation study of temperature dependent bilayer structural properties, Bba-Biomembranes, 1838 (2014) 2520-2529.

[29] B.D. Madej, C.J. Dickson, A.A. Skjevik, R.M. Betz, R.C. Walker, K. Teigen, Lipid14: The Amber lipid force field, Abstracts of Papers of the American Chemical Society, 248 (2014).

[30] J.P.M. Jambeck, A.P. Lyubartsev, An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes, Journal of Chemical Theory and Computation, 8 (2012) 2938-2948.

[31] O. Berger, O. Edholm, F. Jahnig, Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature, Biophys J, 72 (1997) 2002-2013.

[32] T.J. Piggot, A. Pineiro, S. Khalid, Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study, Journal of Chemical Theory and Computation, 8 (2012) 4593-4609.

[33] J.P.M. Jambeck, A.P. Lyubartsev, Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids, J Phys Chem B, 116 (2012) 3164-3179.

[34] N. Schmid, A.P. Eichenberger, A. Choutko, S. Riniker, M. Winger, A.E. Mark, W.F. van Gunsteren, Definition and testing of the GROMOS force-field versions 54A7 and 54B7, Eur Biophys J Biophy, 40 (2011) 843-856.

[35] K. Pluhackova, S.A. Kirsch, J. Han, L.P. Sun, Z.Y. Jiang, T. Unruh, R.A. Bockmann, A Critical Comparison of Biomembrane Force Fields: Structure and Dynamics of Model DMPC, POPC, and POPE Bilayers, J Phys Chem B, 120 (2016) 3888-3903.

[36] J.B. Klauda, B.R. Brooks, R.W. Pastor, Dynamical motions of lipids and a finite size effect in simulations of bilayers, Journal of Chemical Physics, 125 (2006).

[37] C.E. Whitman, R.L. Travis, Phospholipid Composition of a Plasma Membrane-Enriched Fraction from Developing Soybean Roots, Plant Physiology, 79 (1985) 494-498.

[38] M.R. Wenk, Lipidomics: new tools and applications, Cell, 143 (2010) 888-895.

[39] M.E. Siwko, S.J. Marrink, A.H. de Vries, A. Kozubek, A.J. Schoot Uiterkamp, A.E. Mark, Does isoprene protect plant membranes from thermal shock? A molecular dynamics study, Biochim Biophys Acta, 1768 (2007) 198-206.

[40] A. Surjus, M. Durand, Lipid changes in soybean root membranes in response to salt treatment, Journal of Experimental Botany, 47 (1996) 17-23.

[41] M. Goodrich-Tanrikulu, R.L. Travis, Plasma membrane phospholipid and sterol synthesis in soybean hypocotyl segments undergoing auxin-induced elongation, Protoplasma, 185 (1995) 83-92.

[42] H. Jung, Jung, K. and Kaback, H. R. , A conformational change in the lactose permease of Escherichia coli is induced by ligand binding or membrane potential, Protein Science, 3 (1994) 6.

[43] H.R. Kaback, Molecular-biology and energetics of membrane-transport, Journal of Cellular Physiology, 89 (1976) 575-593.

[44] M.G. Madej, H.R. Kaback, Evolutionary mix-and-match with MFS transporters II, Proceedings of the National Academy of Sciences of the United States of America, 110 (2013) E4831-E4838.

[45] Y. Yin, M.O. Jensen, E. Tajkhorshid, K. Schulten, Sugar binding and protein conformational changes in lactose permease, Biophys J, 91 (2006) 3972-3985.

[46] H.R. Kaback, Structure and mechanism of the lactose permease, Comptes Rendus Biologies, 328 (2005) 557-567.

[47] L. Guan, H.R. Kaback, Lessons from lactose permease, in: Annual Review of Biophysics and Biomolecular Structure, vol. 35, 2006, pp. 67-91.

[48] P.Y. Pendse, B.R. Brooks, J.B. Klauda, Probing the periplasmic-open state of lactose permease in response to sugar binding and proton translocation, Journal of Molecular Biology, 404 (2010) 506-521.

[49] M.G. Madej, S.N. Soro, H.R. Kaback, Apo-intermediate in the transport cycle of lactose permease (LacY), Proceedings of the National Academy of Sciences of the United States of America, 109 (2012) E2970-E2978.

[50] D.L. Foster, T. Goldkorn, N. Carrasco, H.R. Kaback, Structure of the lac carrier protein from escherichia-coli, Biophys J, 41 (1983) A277-A277.

[51] S. Radestock, L.R. Forrest, The alternating-access mechanism of MFS transporters arises from inverted-topology repeats, Journal of Molecular Biology, 407 (2011) 698-715.

[52] L. Guan, O. Mirza, G. Verner, S. Iwata, H.R. Kaback, Structural determination of wild-type lactose permease, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007) 15294-15298.

[53] J. Abramson, I. Smirnova, V. Kasho, G. Verner, H.R. Kaback, S. Iwata, Structure and mechanism of the lactose permease of Escherichia coli, Science, 301 (2003) 610-615.

[54] H. Kumar, V. Kasho, I. Smirnova, J.S. Finer-Moore, H.R. Kaback, R.M. Stroud, Structure of sugar-bound LacY, Proceedings of the National Academy of Sciences of the United States of America, 111 (2014) 1784-1788.

[55] X. Jiang, I. Smirnova, V. Kasho, J.P. Wu, K. Hirata, M. Ke, E. Pardon, J. Steyaert, N. Yan, H.R. Kaback, Crystal structure of a LacY-nanobody complex in a periplasmic-open conformation, Proceedings of the National Academy of Sciences of the United States of America, 113 (2016) 12420-12425.

[56] I. Smirnova, V. Kasho, J. Sugihara, J.Y. Choe, H.R. Kaback, Residues in the H+ Translocation Site Define the pK(a) for Sugar Binding to LacY, Biochemistry, 48 (2009) 8852-8860.

[57] H.R. Kaback, I. Smirnova, V. Kasho, Y.L. Nie, Y.G. Zhou, The alternating access transport mechanism in LacY, Journal of Membrane Biology, 239 (2011) 85-93.

[58] Z.Y. Liu, M.G. Madej, H.R. Kaback, Helix Dynamics in LacY: Helices II and IV, Journal of Molecular Biology, 396 (2010) 617-626.

[59] Y.G. Zhou, L. Guan, J.A. Freites, H.R. Kaback, Opening and closing of the periplasmic gate in lactose permease, Proceedings of the National Academy of Sciences of the United States of America, 105 (2008) 3774-3778.

[60] Y.G. Zhou, Y.L. Nie, H.R. Kaback, Residues Gating the Periplasmic Pathway of LacY, Journal of Molecular Biology, 394 (2009) 219-225.

[61] D.S. Majumdar, I. Smirnova, V. Kasho, E. Nir, X.X. Kong, S. Weiss, H.R. Kaback, Single-molecule FRET reveals sugar-induced conformational dynamics in LacY, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007) 12640-12645.

[62] I. Smirnova, V. Kasho, J.Y. Choe, C. Altenbach, W.L. Hubbell, H.R. Kaback, Sugar binding induces an outward-facing conformation of LacY, Proceedings of the National Academy of Sciences of the United States of America, 104 (2007) 16504-16509.

[63] D. Frenkel, B. Smit, Understanding Molecular Simulation: From Algorithms to Applications, Academic Press, 2002.

[64] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, P.A. Kollman, A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol 117, pg 5179, 1995), J Am Chem Soc, 118 (1996) 2309-2309.

[65] B.R. Brooks, C.L. Brooks, A.D. Mackerell, L. Nilsson, R.J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A.R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R.W. Pastor, C.B. Post, J.Z. Pu, M. Schaefer, B. Tidor, R.M. Venable, H.L. Woodcock, X. Wu, W. Yang, D.M. York, M. Karplus, CHARMM: The biomolecular simulation program, Journal of Computational Chemistry, 30 (2009) 1545-1614.

[66] H.J.C. Berendsen, D. Vanderspoel, R. Vandrunen, GROMACS - A message-passing parallel molecular-dynamics implementation, Computer Physics Communications, 91 (1995) 43-56.

[67] S. Plimpton, Fast parallel algorithms for short-range molecular-dynamics, Journal of Computational Physics, 117 (1995) 1-19.

[68] J.C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, K. Schulten, Scalable molecular dynamics with NAMD, Journal of Computational Chemistry, 26 (2005) 1781-1802.

[69] A. Kukol, Molecular Modeling of Proteins, Humana Press, UK, 2008.

[70] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, All-atom empirical potential for molecular modeling and dynamics studies of proteins, J Phys Chem B, 102 (1998) 3586-3616.

[71] P.J. Steinbach, B.R. Brooks, New spherical-cutoff methods for long-range forces in macromolecular simulation, Journal of Computational Chemistry, 15 (1994) 667-683.

[72] W.C. Swope, H.C. Andersen, P.H. Berens, K.R. Wilson, A computer-simulation method for the calculation of equilibrium-constants for the formation of physical clusters of molecules - application to small water clusters, Journal of Chemical Physics, 76 (1982) 637-649.

[73] X.W. Wu, B.R. Brooks, Force-momentum-based self-guided Langevin dynamics: A rapid sampling method that approaches the canonical ensemble, Journal of Chemical Physics, 135 (2011).

[74] S.E. Feller, Y. Zhang, R.W. Pastor, B.R. Brooks, Constant pressure molecular dynamics simulation: The Langevin piston method, Journal of Chemical Physics, 103 (1995) 4613-4621.

[75] G.J. Martyna, D.J. Tobias, M.L. Klein, Constant pressure molecular dynamics algorithms, Journal of Chemical Physics, 101 (1994) 4177-4189.

[76] J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen, Numerical-integration of Cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes, Journal of Computational Physics, 23 (1977) 327-341.

[77] S. Jo, J.B. Lim, J.B. Klauda, W. Im, CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes, Biophys J, 97 (2009) 50-58.

[78] T. Jo S Fau - Kim, V.G. Kim T Fau - Iyer, W. Iyer Vg Fau - Im, W. Im, CHARMM-GUI: a web-based graphical user interface for CHARMM.

[79] S. Jo, T. Kim, W. Im, Automated builder and database of protein/membrane complexes for molecular dynamics simulations, Plos One, 2 (2007).

[80] J. Lee, X. Cheng, J.M. Swails, M.S. Yeom, P.K. Eastman, J.A. Lemkul, S. Wei, J. Buckner, J.C. Jeong, Y.F. Qi, S. Jo, V.S. Pande, D.A. Case, C.L. Brooks, A.D. MacKerell, J.B. Klauda, W. Im, CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field, Journal of Chemical Theory and Computation, 12 (2016) 405-413.

[81] E.L. Wu, X. Cheng, S. Jo, H. Rui, K.C. Song, E.M. Davila-Contreras, Y.F. Qi, J.M. Lee, V. Monje-Galvan, R.M. Venable, J.B. Klauda, W. Im, CHARMM-GUI membrane builder toward realistic biological membrane simulations, Journal of Computational Chemistry, 35 (2014) 1997-2004.

[82] S.R. Durell, B.R. Brooks, A. Bennaim, Solvent-Induced Forces between Two Hydrophilic Groups, Journal of Physical Chemistry, 98 (1994) 2198-2202.

[83] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of Simple Potential Functions for Simulating Liquid Water, Journal of Chemical Physics, 79 (1983) 926-935.

[84] T. Darden, D. York, L. Pedersen, Particle Mesh Ewald - an NLog(N) Method for Ewald Sums in Large Systems, Journal of Chemical Physics, 98 (1993) 10089-10092.

[85] L.J. Berliner, J. Grunwald, H.O. Hankovszky, K. Hideg, A novel reversible thiol-specific spin label - papain active-site labeling and inhibition, Analytical Biochemistry, 119 (1982) 450-455.

[86] G. Jeschke, A. Koch, U. Jonas, A. Godt, Direct conversion of EPR dipolar time evolution data to distance distributions, Journal of Magnetic Resonance, 155 (2002) 72-82.

[87] G. Jeschke, DEER Distance Measurements on Proteins, Annual Review of Physical Chemistry, Vol 63, 63 (2012) 419-446.

[88] S.M. Islam, B. Roux*, Simulating the Distance Distribution between Spin-Labels Attached to Proteins, The Journal of Physical Chemistry B, 119 (2015) 11.

[89] K. Kazmier, S. Sharma, S.M. Islam, B. Roux, H.S. McHaourab, Conformational cycle and ion-coupling mechanism of the Na+/hydantoin transporter Mhp1, Proceedings of the National Academy of Sciences of the United States of America, 111 (2014) 14752-14757.

[90] S.M. Islam, R.A. Stein, H.S. McHaourab, B. Roux, Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme, J Phys Chem B, 117 (2013) 4740-4754.

[91] X.W. Wu, B.R. Brooks, Self-guided Langevin dynamics simulation method, Chemical Physics Letters, 381 (2003) 512-518.

[92] M.A. Olson, S. Chaudhury, M.S. Lee, Comparison between self-guided Langevin dynamics and molecular dynamics simulations for structure refinement of protein loop conformations, Journal of Computational Chemistry, 32 (2011) 3014-3022.

[93] J.F. Nagle, S. Tristram-Nagle, Structure of lipid bilayers, Biochimica Et Biophysica Acta-Reviews on Biomembranes, 1469 (2000) 159-195.

[94] B.W. Koenig, K. Gawrisch, Specific volumes of unsaturated phosphatidylcholines in the liquid crystalline lamellar phase, Bba-Biomembranes, 1715 (2005) 65-70.

[95] N. Kučerka, J.F. Nagle, J.N. Sachs, S.E. Feller, J. Pencer, A. Jackson, J. Katsaras, Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data, Biophys. J., 95 (2008) 2356-2367.

[96] A.R. Braun, J.N. Sachs, J.F. Nagle, Comparing Simulations of Lipid Bilayers to Scattering Data: The GROMOS 43A1-S3 Force Field, J. Phys. Chem. B, 117 (2013) 5065-5072.

[97] S. Lee, A. Tran, M. Allsopp, J.B. Lim, J. Hénin, J.B. Klauda, CHARMM36 United-Atom Chain Model for Lipids and Surfactants, J. Phys. Chem. B, 118 (2014) 547-556.

[98] J.P. Douliez, A. Leonard, E.J. Dufourc, Restatement of order parameters in biomembranes - calculation of c-c bond order parameters from c-d quadrupolar splittings, Biophys J, 68 (1995) 1727-1739.

[99] N. Kucerka, J. Katsaras, J.F. Nagle, Comparing Membrane Simulations to Scattering Experiments: Introducing the SIMtoEXP Software, Journal of Membrane Biology, 235 (2010) 43-50.

[100] C.B. Barber, D.P. Dobkin, H. Huhdanpaa, The Quickhull algorithm for convex hulls, Acm Transactions on Mathematical Software, 22 (1996) 469-483.

[101] J.B. Klauda, N. Kucerka, B.R. Brooks, R.W. Pastor, J.F. Nagle, Simulation-based methods for interpreting X-ray data from lipid bilayers, Biophys J, 90 (2006) 2796-2807.

[102] R.S. Armen, O.D. Uitto, S.E. Feller, Phospholipid component volumes: Determination and application to bilayer structure calculations, Biophys J, 75 (1998) 734-744.

[103] H.I. Petrache, S.E. Feller, J.F. Nagle, Determination of component volumes of lipid bilayers from simulations, Biophys J, 72 (1997) 2237-2242.

[104] N. Kucerka, J.F. Nagle, J.N. Sachs, S.E. Feller, J. Pencer, A. Jackson, J. Katsaras, Lipid bilayer structure determined by the simultaneous analysis of neutron and x-ray scattering data, Biophys J, 95 (2008) 2356-2367.

[105] J.H. Ipsen, O.G. Mouritsen, M. Bloom, Relationships between lipid-membrane area, hydrophobic thickness, and acyl-chain orientational order - the effects of cholesterol, Biophys J, 57 (1990) 405-412.

[106] H. Schindler, J. Seelig, Deuterium order parameters in relation to thermodynamic properties of a phospholipid bilayer. Statistical mechanical interpretation, Biochemistry, 14 (1975) 2283-2287.

[107] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12 (2011) 2825-2830.

[108] H.-P.K. Martin Ester, Jörg Sander, Xiaowei Xu, A density-based algorithm for discovering clusters in large spatial databases with noise, Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining, 96 (1996) 226-231.

[109] X. Zhuang, E.M. Dávila-Contreras, A.H. Beaven, W. Im, J.B. Klauda, An extensive simulation study of lipid bilayer properties with different head groups, acyl chain lengths, and chain saturations, BBA - Biomembranes, 1858 (2016) 12.

[110] S.P. James, Statistical Mechanics: Entropy, Order Parameters, and Complexity, Oxford University Press, 2006.

[111] R.M. Venable, A.J. Sodt, B. Rogaski, H. Rui, E. Hatcher, A.D. MacKerell, R.W. Pastor, J.B. Klauda, CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature, Biophys J, 107 (2014) 134-145.

[112] J.B. Klauda, M.F. Roberts, A.G. Redfield, B.R. Brooks, R.W. Pastor, Rotation of lipids in membranes: Molecular dynamics simulation, P-31 spin-lattice relaxation, and rigid-body dynamics, Biophys J, 94 (2008) 3074-3083.

[113] J.B. Klauda, N.V. Eldho, K. Gawrisch, B.R. Brooks, R.W. Pastor, Collective and noncollective models of NMR relaxation in lipid vesicles and multilayers, J Phys Chem B, 112 (2008) 5924-5929.

[114] G. Lipari, A. Szabo, Model-Free Approach to the Interpretation of Nuclear Magnetic-Resonance Relaxation in Macromolecules .1. Theory and Range of Validity, J Am Chem Soc, 104 (1982) 4546-4559.

[115] M. Ottiger, A. Bax, Determination of relative N-H-N N-C ', C-alpha-C ', andC(alpha)-H-alpha effective bond lengths in a protein by NMR in a dilute liquid crystalline phase, J. Am. Chem. Soc., 120 (1998) 12334-12341.

[116] O.S. Smart, J.M. Goodfellow, B.A. Wallace, The pore dimensions of gramicidin A, Biophys J, 65 (1993) 2455-2460.

[117] S.J. Weiner, P.A. Kollman, D.A. Case, U.C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, A new force-field for molecular mechanical simulation of nucleic-acids and proteins, J Am Chem Soc, 106 (1984) 765-784.

[118] N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, E. Teller, Equation of state calculations by fast computing machines, Journal of Chemical Physics, 21 (1953) 1087-1092.

[119] S. Kirkpatrick, C.D. Gelatt, Jr., M.P. Vecchi, Optimization by simulated annealing, Science, 220 (1983) 671-680.

[120] W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics, J. Mol. Graphics, 14 (1996) 33-38.

[121] T. Williams, C. Kelley, Others, gnuplot 4.2 An Interactive Plotting Program, (2009).

[122] J.B. Lim, J.B. Klauda, Branching at the Iso- and Anteiso- Positions in Complex Chlamydia Membranes: A Molecular Dynamics Study, Biochim. Biophys. Acta, Biomembr., 1808 (2011) 323-331

[123] R.W. Benz, F. Castro-Roman, D.J. Tobias, S.H. White, Experimental validation of molecular dynamics simulations of lipid bilayers: A new approach, Biophys J, 88 (2005) 805-817.

[124] N. Kucerka, Y.F. Liu, N.J. Chu, H.I. Petrache, S.T. Tristram-Nagle, J.F. Nagle, Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using X-ray scattering from oriented multilamellar arrays and from unilamellar vesicles, Biophys J, 88 (2005) 2626-2637.

[125] S. Lee, A. Tran, M. Allsopp, J.B. Lim, J. Hénin, J.B. Klauda, CHARMM36 United Atom Chain Model for Lipids and Surfactants, The Journal of Physical Chemistry B, 118 (2013) 547-556.

[126] S.S. Qin, Z.W. Yu, Y.X. Yu, Structural Characterization on the Gel to Liquid-Crystal Phase Transition of Fully Hydrated DSPC and DSPE Bilayers, J Phys Chem B, 113 (2009) 8114-8123.

[127] J.B. Klauda, V. Monje, T. Kim, W. Im, Improving the CHARMM Force Field for Polyunsaturated Fatty Acid Chains, J. Phys. Chem. B, 116 (2012) 9424-9431.

[128] J.E. Baenziger, H.C. Jarrel, I.C.P. Smith, Molecular Motions and Dynamics of a Diunsaturated Acyl Chain in a Lipid Bilayer: Implications for the Role of Polyunsaturation in Biological Membranes, Biochemistry 31 (1992) 3377-3385.

[129] J. Zajicek, J.F. Ellena, G.D. Williams, M.A. Khadim, a.M.F. Brown, Molecular Dynamics Of Vesicles Of Unsaturated Phosphatidylcholines Studied By 13C NMR Spin-Lattice Relaxation, Collection of Czechoslovak Chemical Communications, 60 (1995) 719-735.

[130] S. Finger, A. Kerth, M. Dathe, A. Blume, The efficacy of trivalent cyclic hexapeptides to induce lipid clustering in PG/PE membranes correlates with their antimicrobial activity, Bba-Biomembranes, 1848 (2015) 2998-3006.

[131] P. Wadhwani, R.F. Epand, N. Heidenreich, J. Burck, A.S. Ulrich, R.M. Epand, Membrane-Active Peptides and the Clustering of Anionic Lipids, Biophys J, 103 (2012) 265-274.

[132] R.M. Epand, R.F. Epand, Lipid domains in bacterial membranes and the action of antimicrobial agents, Bba-Biomembranes, 1788 (2009) 289-294.

[133] H. Koldso, D. Shorthouse, J. Helie, M.S.P. Sansom, Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers, Plos Computational Biology, 10 (2014).

[134] P.J. Franco, E.A. Matzke, J.L. Johnson, B.M. Wiczer, R.J. Brooker, A suppressor analysis of residues involved in cation transport in the lactose permease: Identification of a coupling sensor, Journal of Membrane Biology, 211 (2006) 101-113.

[135] S.E. Feller, D.X. Yin, R.W. Pastor, A.D. MacKerell, Jr., Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: Parameterization and comparison with diffraction studies, Biophys. J., 73 (1997) 2269-2279.

[136] J.B. Michael Schlenkrich, Alexander D. MacKerellJr., Martin Karplus, An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications, in: K.M. Merz, Roux, Benoit (Ed.) Biological Membranes: A Molecular Perspective from Computation and Experiment Birkhauser, Boston, 1996, pp. 31-81.

[137] R.B. Best, X. Zhu, J. Shim, P.E.M. Lopes, J. Mittal, M. Feig, A.D. MacKerell, Jr., Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting

Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles, Journal of Chemical Theory and Computation, 8 (2012) 3257-3273.

[138] J.B. Klauda, R.M. Venable, J.A. Freites, J.W. O'Connor, D.J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A.D. MacKerell, R.W. Pastor, Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types, J Phys Chem B, 114 (2010) 7830-7843.

[139] B. Roux, S.M. Islam, Restrained-ensemble molecular dynamics simulations based on distance histograms from double electron-electron resonance spectroscopy, J Phys Chem B, 117 (2013) 4733-4739.

[140] G. Jeschke, V. Chechik, P. Ionita, A. Godt, H. Zimmermann, J. Banham, C.R. Timmel, D. Hilger, H. Jung, DeerAnalysis2006 - a comprehensive software package for analyzing pulsed ELDOR data, Applied Magnetic Resonance, 30 (2006) 473-498.

[141] Y. Polyhach, G. Jeschke, Prediction of favourable sites for spin labelling of proteins, Spectroscopy-an International Journal, 24 (2010) 651-659.

[142] J. Holyoake, M.S.P. Sansom, Conformational change in an MFS protein: MD simulations of LacY, Structure, 15 (2007) 873-884.

[143] M.O. Jensen, Y. Yin, E. Tajkhorshid, K. Schulten, Sugar transport across lactose permease probed by steered molecular dynamics, Biophys J, 93 (2007) 92-102.

[144] I. Smirnova, V. Kasho, X.X. Jiang, E. Pardon, J. Steyaert, H.R. Kaback, Outward-facing conformers of LacY stabilized by nanobodies, Proceedings of the National Academy of Sciences of the United States of America, 111 (2014) 18548-18553.

[145] I.V. Borovykh, S. Ceola, P. Gajula, P. Gast, H.J. Steinhoff, M. Huber, Distance between a native cofactor and a spin label in the reaction centre of Rhodobacter sphaeroides by a two-frequency pulsed electron paramagnetic resonance method and molecular dynamics simulations, Journal of Magnetic Resonance, 180 (2006) 178-185.

[146] K. Sale, L.K. Song, Y.S. Liu, E. Perozo, P. Fajer, Explicit treatment of spin labels in modeling of distance constraints from dipolar EPR and DEER, J Am Chem Soc, 127 (2005) 9334-9335.

[147] M. Souaille, B. Roux, Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations, Computer Physics Communications, 135 (2001) 40-57.

Publications and Presentations

Publications

Appeared in print:

1. Xiaohong Zhuang, Judah R. Makover, Wonpil Im, and Jeffery B. Klauda, *A systematic molecular dynamics simulation study of temperature dependent bilayer structural properties*. BBA-Biomembranes, 2014, **1838**(10): 2520-2529.
2. Xiaohong Zhuang and Jeffery B. Klauda, *Modeling structural transitions from the periplasmic-open state of* lactose *permease and interpretations of spin label experiments.* BBA-Biomembranes, 2016, **1858** (7 Pt A):1541-52
3. Xiaohong Zhuang, Eder M. Dávila-Contreras, Andrew H. Beaven, Wonpil Im, and Jeffery B. Klauda, *An extensive simulation study of lipid bilayer properties with different head groups, acyl chain lengths, and chain saturations.* BBA-Biomembranes, 2016,**1858**(7):3093-3104

Accepted book chapter

4. Pouyan Khakbaz, Viviana Monje-Galvan, Xiaohong Zhuang, and Jeffery B. Klauda, “Modeling Lipid Membranes”, *Handbook of Hydrocarbon and Lipid Microbiology Series: Biogenesis of Fatty Acids, Lipids and Membranes*, Otto Geiger, Springer, 2016

Submitted:

5. Xiaohong Zhuang, Anna Ou, and Jeffery B. Klauda, *Simulations of simple linoleic acid-containing lipid membranes to models of soybean plasma membranes*. BBA-Biomembranes. 2016.

Will be submitted:

6. Mark Adams, Xiaohong Zhuang, Rui Ponte, Wonpil Im, and Jeffery B. Klauda, *CHARMM36 United-Atom chain force field for general use for glycerophospholipids, sphingomyelins, and ceramides.* (In preparation)
7. Xiaohong Zhuang and Jeffery B. Klauda, *Stability of Outward-facing Models of Lactose Permease*. 2017 (In preparation)

Conference presentations

Delivered:

Oral Presentations:

2015 *American Chemical Society Annual Meeting (Denver, CO)*
Title: Molecular dynamics simulations on the periplasmic-open state lactose permease

Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

2014 *American Institute of Chemical Engineers Annual Meeting (Atlanta, GA)*
Title: Molecular dynamics simulations on the periplasmic-open state lactose permease
Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

Poster Presentations:

2016 *International Conference on Molecular Simulation (Shanghai, China)*

Title: Simulation study on the interpretations of spin label experiments and conformational changes of lactose permease
Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

2016 *Biophysical Society Annual Meeting (Los Angeles, CA)*

Title: Probing conformational changes of secondary active transport proteins
Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

2015 *Biophysical Society Annual Meeting (Baltimore, MD)*

Title: Molecular dynamics simulations on the periplasmic-open state lactose permease
Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

2014 *Biophysical Society Annual Meeting (San Francisco, CA)*

Title: Temperature dependence of bilayer structural properties studied with molecular dynamics simulations
Authors: Xiaohong Zhuang, Dr. Jeffery B. Klauda
Speaker: Xiaohong Zhuang

Scheduled (abstract accepted):

2017 *Biophysical Society Annual Meeting (Baltimore, MD)* *(Feb.11-15, 2017)*

Title: Simulations of linoleoyl-containing pure lipid bilayer and soybean plasma membranes
Authors: Xiaohong Zhuang, Anna Ou, Dr. Jeffery B. Klauda
Speaker: Dr. Jeffery B. Klauda