ABSTRACT

Title of Dissertation:	PARAMETERIZATION OF THE CHARMM LIPID FORCE FIELD AND APPLICATIONS TO MEMBRANE MODELING
	Yalun Yu, Doctor of Philosophy, 2022
Dissertation directed by:	Professor, Jeffery B. Klauda, Chemical and Biomolecular Engineering Department

Computational modeling of lipids at the atomistic level provides insights into the chemical physics of biological membranes and opens the possibility to model membrane-protein interactions. This dissertation presents contributions to the CHARMM/Drude family of lipid force fields and applications of the CHARMM36 lipid force field to model membranes.

Long-range Lennard-Jones interactions are critical for membrane simulations but were excluded from the CHARMM lipid force field for historical reasons. Re-parameterization of the CHARMM36 (C36) lipid force field for phosphatidylcholine, phosphatidylethanolamine, phosphatidylglycerol, and ether lipids is performed to incorporate these interactions through the Lennard-Jones particle-mesh Ewald (LJ-PME) method. The resulting force field is denoted C36/LJ-PME. C36/LJ-PME is in excellent agreement with experimental structure data for lipid bilayers and reproduces the experimental compression isotherm of monolayers. A semi-automated protocol is developed and used during this parameterization and significantly accelerates the whole process.

The same protocol is used for the optimization of the Drude polarizable lipid force field. The optimization of this force field focuses on the structural and mechanical properties of bilayers

and *ab initio* results of model compounds representing the lipid headgroup. Long-range dispersion interactions are incorporated into the force field as well. The resulting force field is validated against more structural and dynamic properties of bilayers and the compression isotherm of monolayers and demonstrates significant improvements over the past versions of the force field.

In addition to these fully atomistic models, this dissertation also discusses the update to the CHARMM36 united atom chain model. Both the original model (C36UA) and the revised model (C36UAr) adopt the all-atom C36 lipid force field parameters for the headgroup and a united atom representation for the chain. The update focuses on the Lennard-Jones parameters of the hydrocarbon chain and related dihedrals. Bulk liquid properties (density, heat of vaporization, isothermal compressibility, and diffusion constant) of linear alkanes and alkenes and *ab initio* torsional scans are used as initial fitting targets. Bilayer surface area is used to fine-tune the dihedral parameters. Bilayer simulations of various headgroups and tails using C36UAr demonstrate significant improvements over C36UA from a structural perspective.

The last part of this dissertation presents the applications of the C36 lipid force field. The inner membrane of *Pseudomonas aeruginosa* (*P. aeruginosa*) is modeled in two modes (planktonic and biofilm) to study the influence of lipid composition on bilayer structural and mechanical properties. The hydrophobic thicknesses of the model membrane agree with the *P. aeruginosa* transmembrane proteins in the Orientations of Proteins in Membranes (OPM) database. Symmetric and asymmetric models for the *Arabidopsis thaliana* plasma membrane are modeled. Molecular dynamics (MD) simulations indicate that the outer leaflet is more rigid and tightly packed to the inner leaflet. The interplay between glycolipids and sterols is found to be critical in lipid clustering and a possible mechanism for lipid phase separation has been proposed.

PARAMETERIZATION OF THE CHARMM LIPID FORCE FIELD AND APPLICATIONS TO MEMBRANE MODELING

by

Yalun Yu

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
Doctor of Philosophy
2022

Advisory Committee:
Professor Jeffery B. Klauda, Chair
Professor Silvina Matysiak
Professor Richard Pastor
Professor Sergei Sukharev
Professor Pratyush Tiwary

© Copyright by
Yalun Yu
2022

Dedication

For my parents, and Xiangyu (Judy) Li

Chapter 1
Yalun Yu, Min-Kang Hsieh, Jeffery B. Klauda, and Richard W. Pastor

Chapter 2
Yalun Yu, Andreas Krämer, Richard M. Venable, Andrew C. Simmonett, Alexander D. MacKerell Jr., Jeffery B. Klauda, Richard W. Pastor, and Bernard R. Brooks.

Chapter 3
Yalun Yu, Jonathan Thirman, Richard M. Venable, Jeffery B. Klauda, Richard W. Pastor, Alexander D. MacKerell Jr., and Benoît Roux.

Chapter 4
Yalun Yu, and Jeffery B. Klauda

Chapter 5
Yalun Yu, and Jeffery B. Klauda

Acknowledgements

I would like to acknowledge my advisors Jeffery Klauda and Richard Pastor. They introduced me to force field development, which I found later is exactly my interest. They value my interest in method development. They were also flexible with my work schedule and gave me a lot of support and care beyond research. They also let me know my strengths and weaknesses, therefore I was able to build confidence in myself. I also enjoyed the discussions with Alex MacKerell and Benoît Roux and benefited from their experience in the Drude model. I am also very grateful for Andreas Krämer, Andy Simmonett and Rick Venable. Andreas guided me through the development stage of FFLiP. Andy was always a good friend to talk about science. Rick influenced me by his knowledge of CHARMM and his clean scripts and documentations. I also want to thank Yihang Wang for being a good friend and also a reliable resource to ask research-related questions.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1. Organization of this dissertation

This dissertation covers four main areas of research: Parameterization and validation of the CHARMM36 (C36) additive lipid force field (FF) with explicit treatment of long-range dispersion (Chapter 2); parameterization and validation of the Drude polarizable lipid FF with explicit treatment of long-range dispersion (Chapter 3); parameterization and validation of the C36 united atom chain model for lipids (Chapter 4); modeling realistic bio-membranes using the C36 lipid FF (Chapter 5). Chapter 1 focuses on background of lipid membranes and lipid FF development, with a special interest in the CHARMM family of lipid FF. Chapter 6 summarizes the work.

This dissertation is composed of both previously published material and new findings. Previously published sections are listed below with the corresponding 6 publications.

Chapter 1: Introduction; Sections 1.3.2 and 1.4

Hsieh, M.-K.; Yu, Y.; Klauda, J. B. All-Atom Modeling of Complex Cellular Membranes. *Langmuir* **2022**.

Chapter 2: Parameterization and validation of the CHARMM36 lipid force field with explicit treatment of long-range dispersion

Yu, Y.; Krämer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D.; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of

Long-Range Dispersion. *Journal of Chemical Theory and Computation* **2021**.

Yu, Y.; Krämer, A.; Venable, R. M.; Brooks, B. R.; Klauda, J. B.; Pastor, R. W. CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *Journal of Chemical Theory and Computation* **2021**.

Chapter 4: Update to the CHARMM36 united atom chain model

Yu, Y.; Klauda, J. B. Update of the CHARMM36 United Atom Chain Model for Hydrocarbons and Phospholipids. *Journal of Physical Chemistry B* **2020**.

Chapter 5: Modeling realistic bio-membranes

Yu, Y.; Klauda, J. B. Modeling Pseudomonas aeruginosa inner plasma membrane in planktonic and biofilm modes. *The Journal of Chemical Physics* **2018.**

Yu, Y.; Klauda, J. B. Symmetric and Asymmetric Models for the *Arabidopsis thaliana* Plasma Membrane: A Simulation Study. *Journal of Physical Chemistry B* **2021**.

1.2. Lipids and their biological significance

Cell membranes are composed of a variety of lipids and proteins where they interact with each other to fulfill their roles. They can provide protection to the cell as a plasma membrane (PM) or define cellular organelles, e.g., endoplasmic reticulum (ER), trans-Golgi Network (TGN) and mitochondria.[1] The major building blocks of

cell membranes, lipids, are amphiphilic molecules with a hydrophilic head and hydrophobic tails, which enables it to act as a barrier between different biological environments where specific chemical reactions may happen to increase biochemical efficiency and to restrict dissemination of reaction products.[1] In addition to the barrier function, lipids are also involved in budding, tubulation, fission and fusion of cells. They also provide the cell with potential for signal recognition and transduction in nutritional uptake, environmental responses, and developmental signaling.[1-3] Finally, lipids are reservoirs for energy and fatty acid and sterol components needed for membrane biogenesis.[1]

The diversity of lipids provides the basis for their various functions noted above. Fahy et al.[4] divide lipids into eight categories (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides) based on their chemical structures. This dissertation focuses on the modeling (FF parameterization) of glycerophospholipids and the application to realistic membranes containing glycerophospholipids, sphingolipids, sterol lipids (see examples in Figure 1).

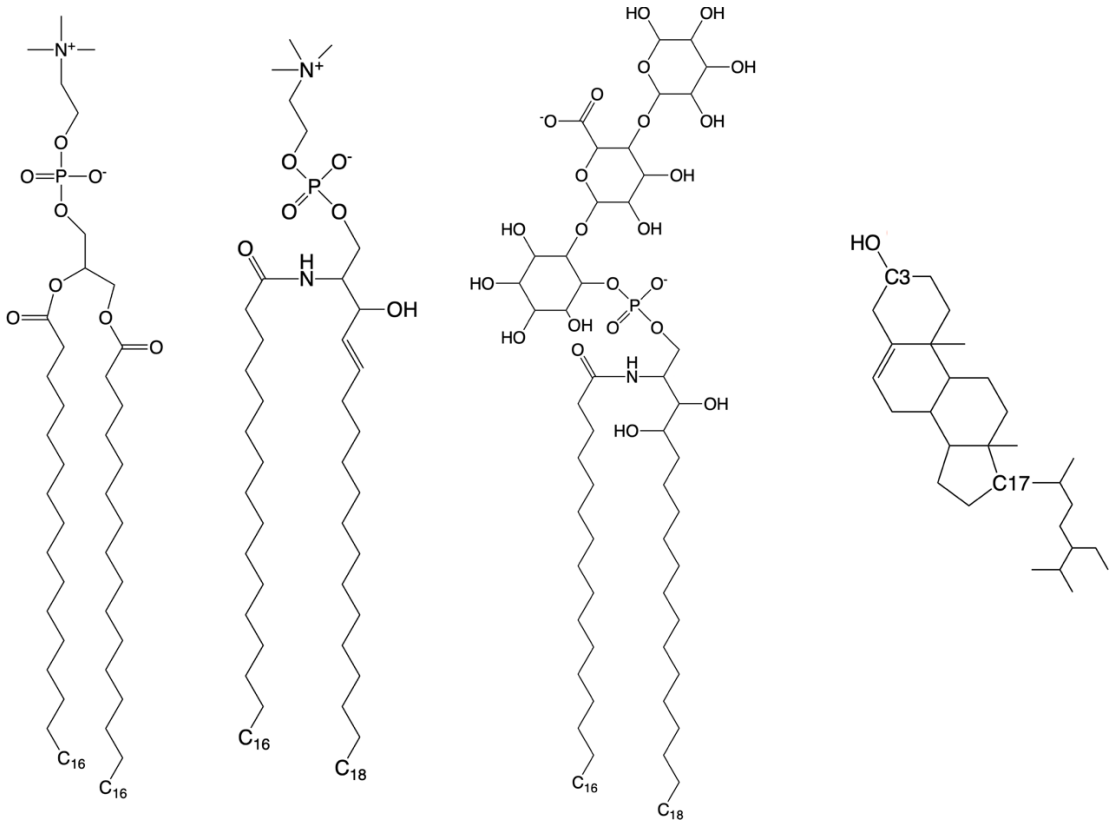

Figure 1. (Left to right) Dipalmitoylphosphatidylcholine (DPPC), N-palmitoyl-D-erythro-sphingosylphosphorylcholine (PSM), a glycosyl-inositol-phosphoceramide (GIPC), and sitosterol (SITO).

Glycerophospholipids can be further classified by their headgroups and tails. Common headgroups include phosphatidylcholine (PC), phosphatidylethanolamine (PE), phosphatidylserine (PS), phosphatidylinositol (PI), phosphatidic acid (PA), and phosphatidylglycerol (PG). Functions of glycerophospholipids are tightly associated with their headgroups. For example, through interactions mediated by their headgroups, which can be reversibly phosphorylated to generate multiple species, PIs play an important role in controlling membrane interfaces. PEs are cone shaped lipids and this is essential for the functional embedding of membrane proteins and for

membrane fusion,[3] PS lipids, on the other hand, plays a critical part in apoptosis as their massive translocation to the outer leaflet of the PM ensures the recognition and uptake apoptotic cells by phagocytes.[5] PC lipids are the most abundant glycerophospholipids in mammalian and plant cells. They have a cylindrical shape due to their large headgroup and hence a low spontaneous curvature, which makes them optimal for assembly into lipid bilayers.

Tails of phospholipids vary widely among different species and organisms. They can be saturated, e.g., palmitoyl acid (16:0) and myristic acid (14:0), or unsaturated, such as oleic acid (18:1) and palmitoleic acid (16:1). There are also polyunsaturated fatty acids such as docosahexaenoic acid (DHA, 22:6), which is essential for brain and the nerve system. Both chain length and unsaturation modify the melting point of single-component bilayers. Melting point increases as chain length increases or as degree of unsaturation decreases.[6] Chain length also directly influences the hydrophobic thickness of the membrane, thus changing the energy penalty of membrane-associated proteins and their functions.[7, 8] Lipid tails may also contain cyclic rings such as those seen in the *Escherichia coli* (*E. coli.*) PM. The function of these lipids is unclear, but simulation studies[9, 10] have shown that increased concentration of cyclic-containing lipids leads to decreased lipid packing and increased area compressibility modulus (K_A). Moreover, chain branching is common in bacteria such as *Chyladmida*.[11] A simulation study has revealed that chain branching increases membrane stiffness and decreases the axial relaxation rates of the lipid.[12]

Sphingolipids share many similarities with glycerophospholipids in structure. The major difference between them is that sphingolipids have a set of trans double

bonded carbons at the *sn*-1 chain linkage region (where the lipid headgroup connects to the chain) instead of the carbonyl group that is found in glycerophospholipids, whereas an amide bond is present at the *sn*-2 chain linkage compared to the ester group in glycerophospholipids (Figure 1). Sphingolipids usually have a long to very long fatty acyl chain. The sphingosine chain is often modified by double bonds and hydroxylation. The headgroup of sphingolipids can also be modified by sugars to form glycolipids, which are abundant in the PM of plants.[13] Sphingolipids and sterols are enriched in the PM (especially the non-cytosolic leaflet) and endosomes.[3] They can form lipid rafts where signaling molecules or proteins reside.[2, 14] Since sphingolipids form a significant amount of hydrogen bonds with neighboring lipids and usually contain saturated tails, they rigidify the membrane.

Sterols and sterol conjugates are another important category of lipids. As mentioned above, sterols can form lipid rafts with sphingolipids. In fact, they interact favorably with lipids containing saturated tails, regulating lipid chain order, and rendering the membrane exceptionally sturdy.[14] A MD study of DPPC/Cholesterol (Chol) mixed bilayers showed that Chol results in significant ordering of the DPPC chains, increase K_A, and reduced lateral diffusion of both DPPC and Chol.[15] Through these interactions, sterols modify the thermotropic phase transition between liquid-disordered and solid-ordered phase by inducing an intermediate liquid-ordered phase. It was proposed this property of sterols might be related to the temperature adaption of plants.[2]

The lipid distribution within the cell is highly heterogeneous. The ER is the main site of lipid synthesis. It keeps low concentrations of sterols and complex

sphingolipids to maintain a loose packing the membrane, which is needed for the insertion and transport of newly synthesized lipids and proteins. The Golgi apparatus acts a sorting station for lipids and proteins and synthesizes ceramides and sphingolipids. The PM defines the boundary of the cell. In eukaryotes, the PM is enriched in sphingolipids and sterols to resist mechanical stress. PC lipids and sphingomyelins are more abundant in the outer leaflet while PE and PS lipids prefer the inner leaflet.[16] Gram-negative bacteria such as *E. coli* and *P. aeruginosa* have two membranes (inner and outer). The inner membrane is almost exclusively composed of PE and PG lipids, while the outer membrane is well known to be asymmetric with Lipopolysaccharides (LPS) on the outer leaflet. The structure of LPS contains a glycan polymer at the outermost domain, the core oligosaccharide, and lipid A having a phosphorylated glucosamine disaccharide with multiple acyl chains. The unique asymmetric outer plasma membrane of gram-negative bacteria functioning as a selective barrier and resistant shell allows the bacterium to survive in many harsh environments and resist antibiotics.

1.3. Lipid force fields

MD simulations provide detailed information of the structure and dynamics of biological systems with proteins, lipids, sugars, and nucleic acids. A crucial part of the simulation is the FF, which is the potential energy function to describe the interaction within the simulated system. This dissertation focuses on the classical description of the potential energy function for lipid molecules.

Since lipids are amphiphilic with a hydrophilic head and hydrophobic tails, the development of a lipid FF is either targeted at these two major segments[17, 18] or smaller

molecules to represent each segment.[19, 20] The parameterization procedure is usually a combination of theoretical calculations such as quantum mechanics (QM) and empirical adjustments using MD simulations. Based on the modeling strategy, popular lipid FFs can be divided into four major groups – additive atomistic FFs, polarizable FFs, united-atom (UA) FFs, and coarse-grained FFs. The parameters associated with these FFs are harmonic bonded terms, dihedral terms, and nonbonded terms (electrostatics and van der Waals). In atomistic FFs,[18, 19, 21-25] all atoms including hydrogens are assigned a mass and its own nonbonded parameters. Polarizable FFs can also model the induced dipoles for all the atoms in the system[26] or the heavy atoms.[27, 28] In a UA FF, selected hydrogens are combined into their bonded heavy atoms forming the united atoms,[29-33] so that the total mass and the effective nonbonded parameters are used for each united atom. For coarse-grained FFs, multiple heavy atoms are grouped together.[34, 35] The choice of the lipid FF in simulations depends on the trade-off between efficiency and accuracy, the type of problem being solved, the compatibility with the non-lipid part of the simulated system, and special considerations such as comparison with previous studies.

1.3.1. Existing lipid force fields

Several major FF communities have systematically parameterized their lipid parameters. Lipid14,[18] the lipid FF of the Amber community, has been validated for 12:0/12:0 PC (DLPC), 14:0/14:0 PC (DMPC), 16:0/16:0 PC (DPPC), 18:1(n-9)*cis*/18:1(n-9)*cis* PC (DOPC), 16:0/18:1(n-9)*cis* PC (POPC) and 16:0/18:1(n-9)*cis* PE (POPE). Cholesterol was also introduced in a later publication.[36] A recent update to the force field aimed to correct the phase transition temperature for several saturated lipids

and the headgroup order parameters was named Lipid21[37] and extended the FF to more lipid species. When using Lipid14 and Lipid21, a long-range analytical dispersion correction should be applied to the energy and pressure.[18, 37]

Another lipid FF compatible with Amber is Slipids.[17, 22, 23, 38] Although originated from the CHARMM lipid FF,[17, 19] it has adopted methods from the Amber community for partial charge refinement and the 1-4 scaling factors for the nonbonded interactions are consistent with Amber.[17] After the initial parameterization for PCs,[17] it has been extended to PEs,[22] PGs,[23] sphingolipids,[23] and lipid-cholesterol mixtures[23], all of which agreed favorably with experimental structural measurements. A more recent study[38] focused on the headgroup torsions improved the agreement with the experimental NMR deuterium order parameters (S_{CD}) for that region. In this FF, a long-range correction for the potential and pressure are also needed.

OPLS-AA is another major all-atom lipid FF, which is relatively new (first introduced in 2014[25]) and the first parameterization only included DPPC. This FF agrees well with experimental measurements of bilayer structures, especially for the headgroup S_{CD}, which had been neglected by most works at that time. Dispersion corrections for both energy and pressure are used in this FF.

Besides the above-mentioned all-atom FFs, the GROMOS community has developed several different branches of United-Atom (UA) lipid FFs. Popular ones include the 43A1-S3 set,[20, 31] the 54A7 set,[39, 40] and the CKP reparameterization.[41] While all belonging to the GROMOS family, they should not be mixed together. Parameterizations of these FFs have primarily focused on PCs and a systematic parameterization or validation on other lipid types is lacked, though scattered works

from different branches[42-44] can be found. In a comprehensive study by Piggot et al.,[41] the authors have shown that removal of the long-range dispersion correction increased the surface area per lipid (A_l) and substantially increased the lipid diffusion.

The constant partial charge approximation in the above-mentioned additive FFs can be problematic when a molecule changes its environments with varying dielectrics. For the lipid FF, specifically, a good example is the permeation of water molecules across a lipid bilayer, where the rate of crossing and partitioning of water within the bilayer is dominated by the free energy barrier at the bilayer interior,[45, 46] which is overestimated by additive FFs.[47-49] To overcome such issues, electric polarizability can be introduced through the fluctuating charges (FQs), the Drude oscillators and induced point dipoles (multipoles). As one most popular polarizable FF, the Amoeba FF has no formally published lipid FF, though two relatively crude parameterizations have been performed for special uses.[50, 51] In the CHARMM community, Patel et al.[45, 52] has developed parameters for DMPC and DPPC based on the FQ model. However, this model requires simulation conducted in the constant pressure, surface area, and temperature (NPAT) ensemble. Furthermore, the FQ model requires a simulation time step of 0.5 fs, which is computationally costly. The CHARMM community has generally switched to using the classical Drude oscillator model to build a polarizable lipid FF,[27, 28] which is compatible with the other parts of the Drude FF.[53]

Finally, some of the above-mentioned FFs use a long-range dispersion correction, which is isotropic. However, a solvated lipid bilayer/monolayer is a heterogeneous system where the contribution of the van der Waals interactions are distributed unevenly in space. Therefore, a more rigorous solution to the long-range

dispersion is the LJ-PME method,[54] where all the van der Waals interactions are calculated explicitly. It should also be noted that none of these FFs has included monolayers as their fitting targets, though monolayers are used heavily in experiments.[55] Chapters 2 and 3 show that by introducing LJ-PME, the isotherm of monolayer can be accurately modeled by both the CHARMM additive FF and the Drude polarizable FF after careful optimizations.

1.3.2. The CHARMM all-atom additive lipid FF

The potential energy for the all-atom CHARMM36 lipid FF can be written as

$$
\begin{aligned}
V(\hat{R}) = & \sum_{\text{bonds}} K_{b,ij}\left(r_{ij} - r_{0,ij}\right)^2 + \sum_{\text{angles}} K_{\theta,ijk}\left(\theta_{ijk} - \theta_{0,ijk}\right)^2 \\
& + \sum_{\text{Urey-Bradley}} K_{ub,i(i+2)}\left(r_{i(i+2)} - r_{0,i(i+2)}\right)^2 \\
& + \sum_{\text{improper}} K_{\phi,ijkl}\left(\phi_{ijkl} - \phi_{0,ijkl}\right)^2 \\
& + \sum_{\text{dihedrals}} \sum_{n} K_{\varphi,n}\left(1 + \cos\left(n\varphi_{ijkl} - \delta_n\right)\right) \\
& + \sum_{\text{nonbonded pairs}} \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{12} - 2\left(\frac{R_{\min,ij}}{r_{ij}}\right)^{6}\right] \\
& + \sum_{\text{electrostatic}} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \qquad (1)
\end{aligned}
$$

where r is the distance between two atoms, θ is the angle formed by three atoms, ϕ is the improper angle, φ is the dihedral angle, δ_n is the phase shift for multiplicity n. The

Urey-Bradley term is used to describe 1-3 interactions between atoms that are separated by two bonds. The improper term is used to describe the out-of-plane orientations of and an atom relative to a plane made by three other atoms. In the CHARMM FF, a minimal number of dihedral terms is used to best represent the torsional profile, though the multiplicity can vary from 1 to 6. For the van der Waals interaction, CHARMM uses the Lennard-Jones (LJ) potential where R_{min} is the distance at which the LJ potential (of a specific atom pair) reaches its minimum. The electrostatic interaction follows Coulomb's Law where q is particle charge of atoms and ε_0 is vacuum permittivity. Parameterization of the CHARMM lipid FF was initially based on small molecules include butane, methylacetate (MAS), dimethylphosphate (DMP), and tetramethylammonium (TMA). The force constants and equilibrium distances for bonds and angles were fitted to the experimental equilibrium geometries and vibrational spectra when available and/or QM vibrational frequencies. Dihedral parameters were fitted to the relative energies of different conformers and the transition barriers from QM calculations. Nonbonded terms were optimized based on water-model compound interaction energies and distances, dipole moments, macroscopic pure solvent properties, and solvation properties. The original version, CHARMM22 (C22),[56] can only be used with the NPAT ensemble since the computational power at that time was not enough to optimize the FF based on lipid bilayers. In the second version, CHARMM 27 (C27), LJ parameters were optimized using minimum interaction energies and geometries from QM calculations between model compounds and rare gases and experimental pure solvent properties.[57] The FF was subsequently used for a range of studies, from which a systematic overestimate of the chain S_{CD} was

observed. A follow-up optimization for torsion parameters of linear alkanes targeting high-level QM calculations on larger model compounds was made to correct this error and the resultant FF was named CHARMM27r (C27r). The most recent update to the FF is CHARMM36 (C36), which yields the correct bilayer surface area at zero surface tension and thus it can be used in the constant particle number, pressure, and temperature (NPT) simulations.

The C36 lipid FF uses the TIP3P water model[58] with a small modification to avoid instabilities during MD.[59] The water model is effectively part of the FF because the FF was optimized to balance the solute-solute, solvent-solute, and solvent-solvent interactions. Although the TIP3P water model has several deficiencies (for example, too fast diffusion, too low surface tension),[60] it is not recommended to use another water model with the CHARMM lipid FF.

The cutoff method for the LJ interactions is another part of the FF. The C36 lipid FF uses the particle-mesh Ewald (PME) method[61] for the electrostatic interactions and a force-based switching function ranging from 8 to 12 Å for the LJ. This treatment for the LJ interactions was needed at the time of parameterization since there was no efficient algorithm for the long-range dispersion. However, the nonbonded terms, especially the electrostatic parameters, were carefully optimized to match surface area per lipid (A_l) for a DPPC bilayer and the aqueous solvation free energies of methylacetate (MAS) and dimethylphosphate (DMP). The initial parameterization focused on DPPC but was also tested against 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), which is a monounsaturated lipid. The agreement with experimental A_l and headgroup structure (measured by S_{CD}) for these lipids was

outstanding. However, the isotherm of lipid monolayer A_l versus surface tension was not accurately modeled by the C36 FF due to the lack of long-range dispersion. In the following years, more lipid headgroups were added to the FF. These include a systematic parameterization of the non-bonded fix (NBFIX) terms between ions and the ester and carbonyl oxygens to incorporate phosphatidylglycerol (PG) and phosphatidylserine (PS) lipids, and the extensive tests on phosphatic acids (PA)[62] and phosphatidylinositols (PI).[63] Sphingolipids and ceramides have been parameterized and tested in a series of studies.[64-66] Following the re-optimization of the ether linkage to increase bilayer hydration,[67] parameters for plasmalogens (which contain a vinyl-ether and an ester bond at the *sn-1* and *sn-2* positions, respectively) were developed recently.[68] Glycolipids and lipopolysaccharides (LPS) have also been developed and tested[69-72] utilizing the CHARMM carbohydrate FF.[73-76] Aside from this headgroup diversity, work has also been focused on the varying lipid tails, i.e., polyunsaturated,[77] branched,[12] and cyclic-containing.[9, 78] The C36 lipid FF also supports various sterols.[79-81]

1.3.3. The Drude polarizable lipid FF

Although the partial charges in the CHARMM lipid FF were carefully optimized to account for induced polarization in a mean-field average way by targeting various condensed phase experimental data, this mean-field approximation is questionable in a highly heterogeneous membrane environment. In fact, the relative dielectric constant changes from 80 (water) to 2 (hydrophobic core) in just 20 Å. Work by Harder et al.[82] has indicated that a polarizable model is essential to reproduce the experimental monolayer-air dipole potential. In a more recent study, Venable et al.[46]

have shown that a polarizable FF can better predict the transfer free energy of permeants (polar and nonpolar ones) from water to alkane. This is expected because the dielectric constants of these permeants and solvent are more accurately modeled by the polarizable FF.[83]

As mentioned above, CHARMM treats atomic polarizability with the classical Drude oscillator model. In the CHARMM polarizable (henceforth Drude) FF, an auxiliary particle called the Drude particle is bonded to a polarizable atom through a harmonic potential and the position of the Drude particle relative to the attached atom is allowed to fluctuate according to the potential. Since the Drude particle takes a charge q_D and a mass of 0.4 amu from the attached atom, this fluctuation mimics the induced dipole in response to an electric field. Assuming the force constant of the harmonic potential is k_D, then the atomic polarizability α equals q_D^2/k_D. In the Drude FF, k_D is set as a constant and each polarizable atom has a unique atomic polarizability which can be parameterized. Atomic polarizability is isotropic in most cases but can be anisotropic when anisotropic polarizability is needed. Atoms described by anisotropic polarization are also generally augmented by "lone pairs" (virtual sites carrying negative charge) to further improve the description of electronic distribution around the polarizable atom.[53] In addition, an anharmonic restoring force (commonly known as the Drude hard wall) is introduced to damp atom-Drude separations beyond 0.2 Å.[84]

As a result, the potential energy of the Drude FF is described by Eq. 1 with additional terms

$$U_{Drude}\left(\hat{R}, \hat{R}_{Drude}\right) = \frac{1}{4\pi\varepsilon_0}\left(\sum \frac{q_{D,i}q_j}{\left\|r_{D,i} - r_j\right\|} + \sum \frac{q_{D,i}q_{D,j}}{\left\|r_{D,i} - r_{D,j}\right\|}\right) +$$

$$\frac{1}{2}\sum k_D \left\| r_{D,i} - r_j \right\|^2 \qquad (2)$$

where $q_{D,i}$ is the charge carried by the Drude particle and $r_{D,i}$ is the coordinate of the Drude particle. The first term accounts for the atom-Drude and Drude-Drude interactions while the second term represents the energy of self-polarization. One major difference from the additive formalism is the explicit inclusion of dipole-dipole interactions for atoms within three bonds. In contrast, nonbonded interactions between atoms connected by one or two bonds are excluded in the CHARMM additive FF but implicitly included in the bonded terms. It should be noted that these 1-2 and 1-3 dipole-dipole interactions are screened by

$$S_{ij}(r_{ij}) = 1 - \left[\left(1 + \frac{a_{ij} r_{ij}}{2(\alpha_i \alpha_j)^{\frac{1}{6}}}\right)\right] \exp\left[\frac{-a_{ij} r_{ij}}{(\alpha_i \alpha_j)^{\frac{1}{6}}}\right] \qquad (3)$$

where r_{ij} is the distance between atoms i and j, α_i and α_j are respective atomic polarizabilities, and a_{ij} is a damping constant calculated by $a_{ij} = a_i + a_j$ where a_i and a_j are the atom-specific Thole screening factors. This screening, together with the Drude hard wall, is essential to prevent polarization catastrophe.

The development of the Drude lipid FF began in the early 2010s.[27] Parameters for individual building blocks (model compounds) of the PC lipid were first obtained. Partial charges, atomic polarizabilities and Thole screening factors were determined using electrostatic potential (ESP) calculated at a set of grid points surrounding a molecule. Candidate LJ parameters were first screened out using QM gas phase interaction energies and then selected based on liquid molecular volumes and enthalpy

of vaporization. Dihedral parameters were optimized against gas phase QM conformational energies. Additional dihedral parameters necessary for connecting the small model compounds were optimized using larger model compounds. The initial parameterization only considered DPPC, but a recent update by Li et al.[28] has expanded the FF to PC lipids with saturated and unsaturated tails and PE lipids. It was claimed that changes in the dihedral parameters around the phosphate group and the glycerol backbone had improved the agreement with experimental deuterium order parameters. However, this FF still underestimates A_l for most lipids and substantially overestimates K_A's. Moreover, as will be shown in Chapter 3, structural data such as hydrogen order parameters can supplement deuterium order parameters to optimize the dihedral parameters.

1.4. Computational modeling of complex cellular membranes

MD simulations provide atomistic resolution to the physical properties of cellular membranes and give insight into membrane-protein interactions. Key to understanding the physical properties of realistic cellular membranes is obtaining the lipid composition of cells and organelles. Earliest studies were on overall lipid compositions of entire cells like that done by for yeast with mass spectroscopy.[85] Experimental techniques to pull out individual organelles of the cell allow for improved quantification of cellular distribution of lipids.[86] More recent lipidomic studies with enhanced resolution can provide detailed distributions of headgroups and associated tails.[85] Moreover, cell membrane leaflets for many organelles are known to consist of varied concentrations of lipid headgroups and acyl chains.[1, 87]

This dissertation focuses on the modeling of pure (no-protein) cellular membranes at the atomistic level, though coarse-grained (CG) and mesoscale models have been used by other researchers to study membrane and membrane-protein interactions.[88] Cell membranes are composed of many different lipids containing various headgroups and tails, so it is evident that we want to include more than one or two lipid components in our models. However, due to FF limitations and computational cost, membrane modeling had been limited to single-component bilayers in the early 1990s[89] and two to three lipid components in the early 2000s.[90, 91] In the last decade, efforts have been made to model more realistic membranes at longer time scales. First, sophisticated lipid FFs have been developed to support a larger diversity of lipids; Second, online membrane builders like the CHARMM-GUI Membrane Builder[92] and Insane[93] have significantly simplified the building procedure. Last, many advanced simulation programs now provide faster codes (especially those running on Graphical Processing Units) reaching speeds for ns/day.

Chapter 2: Parameterization and validation of the CHARMM36 lipid force field with explicit treatment of long-range dispersion

This chapter was edited from publications:

Yu, Y.; Krämer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D.; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of Long-Range Dispersion. *Journal of Chemical Theory and Computation* 2021.

Yu, Y.; Krämer, A.; Venable, R. M.; Brooks, B. R.; Klauda, J. B.; Pastor, R. W. CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *Journal of Chemical Theory and Computation* 2021.

2.1. Introduction

Since its release in 2010, the C36 lipid FF[19] has been extended to include anionic lipids,[94] ether lipids,[67] ceramides,[64, 65] glycolipids,[95] plasmalogens,[68] and polyunsaturated tails.[77] The C36 lipid FF is heavily utilized due to the diversity of lipids it covers and the well-parameterized potential parameters. Like other commonly used FFs, C36 was parameterized considering specific treatment of nonbonded interactions. The C36 lipid FF uses the PME method[61] for the electrostatic interactions and a force-based switching function ranging from 8 to 12 Å for the LJ potential. Hence, the long-range dispersion is excluded from the energy calculation. Although standard for C36 and most other FFs, this way of treating the LJ potential leads to an inconsistency

between bilayers and monolayers, because the acyl chain/air interface of monolayers, similar to alkane/air interfaces, requires long-range dispersion to reproduce experiment; i.e., the LJ interactions must be summed to infinity, similar to the electrostatic terms (Figure 2).[67, 96, 97] However, as also shown in Figure 2, any deviation from the 12 Å cutoff in LJ leads to dramatic change in bilayer surface areas, and as follows from the hexadecane results, surface tensions of alkane/air interfaces. This inconsistency was recognized in the original publication of C36,[19] but it was not possible to rectify the problem at that time because an efficient method for calculating long-range LJ interactions was not supported in CHARMM[98] and other major simulation programs. The situation has changed and the inclusion of long-range LJ terms using a Particle Mesh Ewald method similar to that used for electrostatics called LJ-PME has been added to Gromacs[54, 99], OpenMM,[100] and CHARMM.[101]

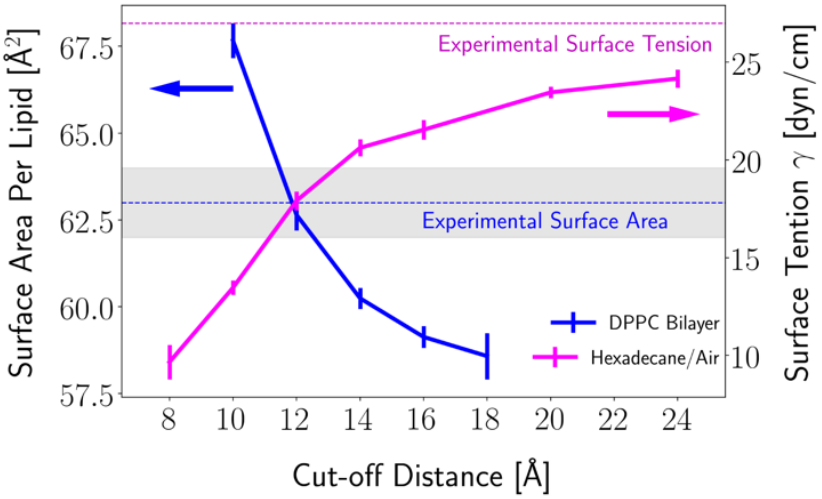

Figure 2. Influence of Lennard-Jones cutoff distance on the surface area per lipid of a DPPC bilayer at 323.15 K (blue line) and the surface tension of hexadecane/air interface at 298.15 K (magenta line) simulated with the C36 FF. Hexadecane/air data from Leonard et al. [101]

This chapter presents the re-parametrization of the C36 lipid FF with LJ-PME on common biologically relevant lipids and the extended validation of the newly optimized FF. The optimization of the new FF used a semi-automated workflow to fit the nonbonded parameters (partial atomic charges and LJ parameters) starting from the original C36 parameters. Since the hydrocarbon parameters utilized in lipid tails are well validated,[101] this re-parametrization is restricted to the headgroups. The water model, TIP3P,[58, 59] also remains unaltered to maintain consistency with the CHARMM-family of FFs. The experimental targets used in the parametrization include surface areas per lipid for bilayers, monolayer isotherms of the surface area, area compressibilities, NMR deuterium order parameters and radial distribution functions between water molecules and specific atoms in the headgroups. While the lipids used for the parametrization are PCs and ether lipids, the validation of the new FF also covers PEs and PGs (see Table 2). The optimized parameter set maintains the same level of accuracy for bilayer as C36, while correctly reproducing the pressure/area isotherm for monolayers.

A major challenge in lipid FF development is the complexity of the physics underlying the parametrization problem, and usually one can only focus on limited aspects of the system of interest under certain conditions (e.g., pressure, temperature, pH, ionic strength). Furthermore, a rigorous description of certain properties may not even be possible with a particular class of FF. For example, additive FFs are deficient in regions where the dielectric response is dominated by electronic polarization so that the permeability of water in bilayers is not well described by C36.[46, 47] Consequently, FF developers must decide the important scientific targets of their FFs, leading to

different parametrization strategies used by different communities. The development of the CHARMM lipid FF has historically utilized QM calculations and experimental data of model compounds before selected parameters are manually adjusted to reproduce experimental properties of lipid bilayers.[19, 56] Such an approach could neglect the interactions between functional groups within a lipid and from neighboring lipids in the condensed phase. As a result, additional efforts are usually required to refine the parameters. For example, developers of the C36 lipid FF tested several methyl acetate models before the final set was selected based on the accurate reproduction of bilayer experimental data.[19, 102] This procedure is consistent with the rest of the CHARMM additive FF allowing for application of the FF to complex, heterogeneous systems.

Additionally, both the optimization and validation of FFs require significant computational resources. MD simulations of lipid bilayers in the 1990s were restricted to small systems and sub-nanosecond time scales,[89, 103, 104] which greatly limited the development of reliable lipid FFs. This situation has significantly improved over the last two decades. With the increase of computer power and a variety of MD engines supporting parallel computing and/or graphics processing unit (GPU)-acceleration,[98, 100, 105, 106] a microsecond all-atom simulation of a thousand lipids is well within capabilities of most computer clusters. However, the atomistic simulation methods still suffer from the fact that the range of time and length scales by which different processes occur in membranes is vast.[107] For a simple calculation of bilayer surface area, the uncertainty from a simulation on the 100 ns timescale is typically less than 1%.[62, 108, 109] This error is acceptable in most circumstances but complicates optimization with

noisy parameter sensitivities computed through finite differences.[110, 111] Furthermore, simply obtaining the surface area within the experimental error (which can be larger than 1%) does not assure the accuracy of other properties. Finally, lipid FFs are frequently subjected to adjustments to study new experimental results and utilize simulation approaches, so efficient parametrization methods are essential.

To address these problems, physics-informed methods can be used to identify beneficial modifications of the parameters via reliable sensitivity evaluations and thereby guide the optimization process in a more efficient way.[112-115] This chapter introduces a semi-automated optimization approach for the CHARMM lipid FF. While both manual and automated FF parametrization have a long history, those two approaches have been developed mostly independently. Several authors have proposed automated strategies to fit force field parameters exclusively with respect to QM target data.[116-119] However, fully automated parametrization with respect to experimental observables was long limited to low-dimensional parameter spaces[44, 47-49] and has only recently been leveraged to parametrize full FFs in the context of the OpenForceField Initiative.[120, 121] Automated optimization procedures for united-atom lipid FFs have been presented in Refs.[30, 122] Such automated workflows have been developed for a gamut of numerical optimization algorithms, including gradient-based methods[112, 115, 123, 124] as well as global optimization methods based on evolutionary simplex,[30, 125] and metamodel-based optimization.[110, 114] However, FF optimization problems are often underdetermined, which can easily lead to unreasonable parameters that are generally not transferrable to simulation setups outside the training set. Therefore, the most used FFs to this day are still based on manual development.

The present work combines automated and manual approaches. While chemical intuition and consistency with the CHARMM FF are encoded in the form of restraints and suitable weights, new iterations of the FF are generated by an automated strategy based on thermodynamic reweighting. This approach allows the incorporation of intricacies specific to the lipid FF. While the primary goal in this dissertation is to incorporate the LJ-PME method into the C36 lipid FF, the general methodology is transferable to other modifications and other lipid FFs.

2.2. Methods

2.2.1 Lipids covered and nomenclatures

This study includes PC, PE, PG, and ether lipids. Optimized parameters are primarily tested on single component lipid bilayers and monolayers. 1,2-dipropionyl-*sn*-glycero-3-phosphocholine (C_3-PC) in solution phase is also simulated to study the headgroup hydrations. Table 1 provides the abbreviations and chain structures for the assorted lipids used in this study. The atom notations used by C36/LJ-PME are the same as C36 and follow the Sundaralingam nomenclature,[126] where the numbering of the glycerol carbons is inverted with respect to the *sn* nomenclature.[127, 128] Examples of atom notation based on DPPC and DHPC can be found in Figure 3. For consistency, CHARMM atom notations are used for all discussions in this dissertation.

Table 1. Name, abbreviation, and chain structure of lipids noted in this paper.

lipid name	abbreviation	*sn1*	*sn2*
1,2-dipropionyl-*sn*-glycero-3-phosphocholine	C_3-PC	3:0	3:0
1,2-dilauroyl-sn-glycero-3-phosphorylcholine	DLPC	12:0	12:0
1,2-dimyristoyl-sn-glycero-3-phosphorylcholine	DMPC	14:0	14:0
1,2-dipalmitoyl-sn-glycero-3-phosphocholine	DPPC	16:0	16:0
1,2-dimyristoyl-*sn*-glycero-3-phospho-(1'-rac-glycerol)	DMPG	14:0	14:0
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine	POPC	16:0	18:1
1,2-dioleoyl-sn-glycero-3-phosphocholine	DOPC	18:1	18:1
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine	POPE	16:0	18:1
1,2-dioleoyl-sn-glycero-3-phosphoethanolamine	POPG	16:0	18:1
1-stearoyl-2-docosahexaenoyl phosphatidylethanolamine	SDPE	18:0	22:6
1,2-di-O-hexadecyl-*sn*-glycero-3-phosphocholine	DHPC	16:0	16:0

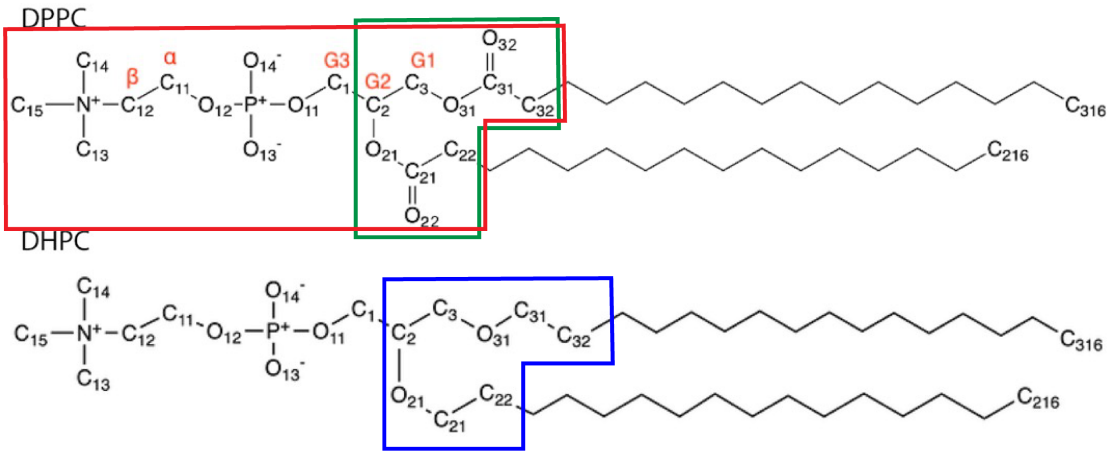

Figure 3. Chemical structure of DPPC and DHPC and regions of nonbonded parameterization. nomenclature follows CHARMM convention. The green box is the region for the *Linkage* optimization and the red box is the region for the Global optimization (see Section 2.2.3 for definitions of *Linkage* and *Global*). The blue box indicates the region for the ether optimization.

2.2.2. Training Set

The training targets were selected to be representative of lipid systems. There are two major considerations when selecting the training targets. The first is the lipid type. While various tails were used in every parametrization stage to explore effects of unsaturation and carbon number, different headgroups and linkages were parametrized sequentially. PC was the first headgroup optimized, and the ether linkage (DHPC) is optimized afterwards. PE and PG lipids were tested using the new parameters of the ester linkage region from the PC optimization and a decision on whether to further optimize them was made based on the tests. The reason behind this is that PE and PG headgroups share the same phosphate and glycerol backbone moieties with PC, hence it can be optimized easily after PC (if needed). For the ether lipid, DHPC, the only difference from DPPC is the linkage between the headgroup and tails, which can be parametrized separately as suggested by the parametrization of C36 lipid FF for ether lipids.[67]

The second consideration is what properties should be covered in the training set. As commonly used benchmarks for lipid FF development, surface areas and membrane thicknesses were included. The correct surface area is difficult to parametrize (because the large negative and positive oscillations in the tangential pressure must sum to zero[108]) and is critical to other properties. For this reason, relatively high weight factors were given to the surface areas in the optimizations. The monolayer surface areas at different surface tensions were also included in the training set, which historically has not been the case in parametrizations. This was inspired by the known inconsistency of bilayer and monolayer surface tensions for C36. Another

set of properties considered were the deuterium order parameters, S_{CD}. Order parameters are sensitive to bond orientations and chain order and are important metrics for membrane structure. In this study, order parameters from headgroup and tails were included if the experimental data was available. Since there is no published experimental data for ether lipids to the author's knowledge, only the PC optimization utilized order parameters.

The radial distribution functions (RDFs) between PC headgroup and water were recently determined by Foglia et al.[129] using C_3-PC in solution. Using neutron diffraction with isotopic substitution (NIDS) augmented by empirical potential structure refinement (EPSR) modeling, RDFs for unique atom-atom pairs between functional groups in C_3-PC and water were resolved. While the data itself is meaningful, the usefulness of it is debatable for several reasons. The first is that the TIP3P water model used by the CHARMM FF cannot reproduce experimental RDFs accurately, especially for the O-O pairs.[130] Secondly, electrostatic polarization is not included in the model, which may influence the interaction strengths between water and lipid atoms. Lastly, it was noted by Foglia *et al*. that EPSR only provides a model which is consistent with the diffraction data, but it is not necessarily unique. However, the inclusion of this set of data can avoid overfitting in parameters which may lead to unwanted strong interactions between water and lipid atoms when the optimizer tries to increase the hydration. To balance all the uncertainties, low weights were assigned to this data. Another concern is how to use the RDF data. A preliminary test using C36 indicated that integrals of the RDF curve (coordination numbers of water) are in good agreement with experiment, so that only the RMSD ranging from 2 to 6 Å were used.

Apart from properties extracted from single simulation, the training set also included three additional properties which were determined from multiple simulations. The most important one is the K_A of DPPC bilayer at 323.15 K. As discussed in previous publications,[19, 108] K_A can be determined by the fluctuation of the surface area, or through the surface tension dependence of the surface area at a given temperature as

$$K_A = \frac{k_B T \langle A \rangle}{\langle \delta A^2 \rangle} = A \left(\frac{d\gamma}{dA} \right)_T \quad (4)$$

where $\langle A \rangle$ is the average total area per leaflet, $\langle \delta A^2 \rangle$ is the mean square fluctuation, k_B is Boltzmann's constant, T is the temperature, and γ is the total surface tension tangential to the bilayer surface. The second one is the area difference between 323.15 K and 333.15 K for DPPC bilayer, which was used to extract the temperature dependence of the surface area. The last one is the area difference between bilayers of DMPC and DPPC, which was used to train the tail length dependence of the surface area. The weights for these properties were relatively low because their uncertainties were large from simulations.

Table 2 details the systems simulated in this re-parameterization. Table 3 shows the training targets associated with these systems.

Table 2. Systems simulated in this re-parameterization.

System	T (K)	Ensemble	N_{lipid}	$N_{water/lipid}$	Use
DPPC bilayer	323.15	NPT	72	30.4	optimization, validation
		NPγT (-5 dyn/cm)	72		optimization
		NPγT (5 dyn/cm)	72		optimization
		NPT	288		validation
		NPT	648		validation
		NPT	1152		validation
	333.15	NPT	72		optimization, validation
DPPC monolayer	321	$NL_z\gamma T$ (18 dyn/cm)	72	30.4	optimization, validation
		$NL_z\gamma T$ (40 dyn/cm)	72		optimization, validation
		$NL_z\gamma T$ (55 dyn/cm)	72		optimization, validation
DMPC bilayer	303.15	NPT	72	25.7	optimization, validation
DLPC bilayer	303.15	NPT	72	30.4	validation
POPC bilayer	303.15	NPT	72	31.1	optimization, validation
DOPC bilayer	303.15	NPT	72	33.5	validation
	303.15	NPT	288	30.4	validation
	303.15	NPT	648	33.5	validation
POPE bilayer	303.15	NPT	80	32.0	validation
	308.15	NPT	80	32.0	validation
DMPG bilayer	303.15	NPT	72	75.0	validation
	323.15	NPT	72		validation
	333.15	NPT	72		validation
POPG bilayer	303.15	NPT	72	45.0	validation
DHPC bilayer	321	NPT	80	30.0	optimization, validation
	333	NPT	80		optimization, validation
C_3-PC solution	298.15	NPT	9	250	optimization, validation

Table 3. Training target data, along with their scaling factors and weight factors. Properties cover surface area per lipid (A_l), deuterium order parameters (S_{CD}), overall bilayer thickness (D_B), and radial distribution function (RDF). S_{CD} (tail) refer to chain order parameters from the 4th carbon in each tail with available experimental data. Scaling and weight factors are defined Section 2.2.7.

property	system, temperature (K)	target value	scaling factor	weight factor – *Global*	weight factor – *Linkage*
A_l	DPPC Bilayer, 323.15	63.1 (Å^2)[131]	60 Å^2	20	20
S_{CD} (tail)		ref [132, 133]	0.15	0.05 × 11	0.05 × 11
S_{CD} (headgroup)		ref[134]	0.15	0.5 × 11	0.5 × 11
D_B		39.0 (Å)[131]	40 Å	5	5
ΔA_l (DMPC-DPPC)	DPPC Bilayer, 323.15 / DMPC Bilayer, 303.15	2.5 (Å^2)[131, 135]	2 Å	1	1
A_l	DMPC Bilayer, 303.15	60.6 (Å^2)[135]	60 Å^2	10	10
D_B		36.7 (Å)[131]	40 Å	5	10
A_l	DPPC Bilayer, 333.15	65.0 (Å^2)[131]	60 Å^2	10	10
D_B		38.1 (Å)[131]	40 Å	5	5
A_l	DMPC Bilayer, 303.15	60.6 (Å^2)[135]	60 Å^2	10	10
D_B		36.7 (Å)[131]	40 Å	5	10
A_l	DHPC Bilayer, 333.15	67.2 Å^2 [136]	60 Å^2	15	
D_B		36.8 Å [136]	40 Å	5	
A_l	DHPC Bilayer, 321.15	65.1 Å^2 [137]	60 Å^2	10	
A_l	POPC Bilayer, 303.15	64.3 (Å^2)[131]	60 Å^2	10	
A_l	DPPC Monolayer (18 dyn/cm), 321	54 (Å^2)[131]	60 Å^2	5	5
A_l	DPPC Monolayer (40 dyn/cm), 321	64 (Å^2)[131]	60 Å^2	5	5
A_l	DPPC Monolayer (55 dyn/cm), 321	80 (Å^2)[131]	60 Å^2	5	5
RDF (water atom ~ lipid atom) Maximum	C_3-PC in solution, 298.15	ref [129]	1	0.75 × 10	0
RDF (water atom ~ lipid atom) Minimum			0.5	0.75 × 6	0
RMSD of RDF from 0.2 to 0.6 nm			1	0	1.5
K_A of DPPC Bilayer at 323.15 K	DPPC Bilayer (5 dyn/cm), 323.15 / DPPC Bilayer (-5 dyn/cm), 323.15	231 (dyn/cm) [138]	200 dyn/cm	0.75	0.75
DPPC Bilayer Isobaric Expansion	DPPC Bilayer, 323.15 / DPPC Bilayer, 333.15	0.19 (Å^2/K) [131]	0.2 Å^2/K	1	1

2.2.3. Parameters to optimize

A previous study demonstrated that the C36 lipid FF reproduces experimental observables very well for linear alkanes when the long-range LJ interactions are included,[101] so that it is reasonable to maintain the alkane parameters and only modify the headgroup parameters for this re-parametrization. The next question concerns which part of the headgroup should be adjusted. To explore the outcomes of different choices, two separate optimizations were performed for the ester-linked PC lipids. In the first, only the nonbonded parameters of the glycerol and ester groups were optimized to maximize consistency with the rest of the CHARMM FF, e.g., the phosphate nonbonded parameters would remain the same for lipids, proteins, and DNA. Table 4 presents the nonbonded parameters explicitly optimized in both the *Global* and *Linkage* sets. The same information is visualized in Figure 3. The nonbonded parameters of the ether-linked lipids were selected based on the results of the *Global* and *Linkage* parameterizations and will be detailed in Section 2.3.5.

Table 4. Parameters subject to change for the two optimizations. Atom notations shown in Figure S1. Asterisk sign stands for A/B/C.

Optimization Name	*Global*	*Linkage*
Charges	N, H13*, H14*, H15*, C13, C14, C15, C12, H12*, C11, H11*, P, O13, O14, O11, O12, C1, HA, HB, C2, HS, O21, C21, O22, C22, H2R, H2S, C3, HX, HY, O31, C31, O32, C32, H2X, H2Y	C2, HS, O21, C21, O22, C22, H2R, H2S, C3, HX, HY, O31, C31, O32, C32, H2X, H2Y
ε	N, H13*, H14*, H15*, C13, C14, C15, H12*, C11, H11*, P, O11, O12, O21, C21, O22, O31, C31, C32	O21, O22, O31, C32
$R_{min}/2$	N, H13*, H14*, H15*, C13, C14, C15, H12*, C11, H11*, P, O11, O12, O21, C21, O22, O31, C31, C32	O21, O22, O31, C32

2.2.4. General optimization procedure

The optimization strategy (Figure 4), FFLiP (**F**orce **F**ield of **Li**pid **P**arametrization), was inspired by earlier gradient-based parametrization strategies,[123, 124] especially the ForceBalance protocol.[112] The starting point was the C36 set, which has been parametrized targeting reliable experimental data and high-level QM calculations. Lipid-containing systems were initially simulated with this parameter set. Upon completion of the simulations, equilibrium properties $\boldsymbol{f}^{sim}$ and potential energies U were calculated for each trajectory frame. The same trajectories were used for potential energy re-calculations in OpenMM 7.4.1[100] with the perturbed parameter sets. Parameter sensitivities (gradients) of the equilibrium properties were estimated through thermodynamic reweighting[139]

$$\boldsymbol{S}_{prop}(\boldsymbol{\lambda}, \boldsymbol{\delta\lambda}) = \frac{1}{|\boldsymbol{\delta\lambda}|}\left(\langle \boldsymbol{f}^{sim}\rangle_{\lambda+\delta\lambda} - \langle \boldsymbol{f}^{sim}\rangle_{\lambda}\right)$$

$$= \frac{1}{|\boldsymbol{\delta\lambda}|}\left(\frac{\langle \boldsymbol{f}^{sim} e^{-\beta(U_{\lambda+\delta\lambda}-U_{\lambda})}\rangle_{\lambda}}{\langle e^{-\beta(U_{\lambda+\delta\lambda}-U_{\lambda})}\rangle_{\lambda}} - \langle \boldsymbol{f}^{sim}\rangle_{\lambda}\right) \quad (5)$$

where $\boldsymbol{S}_{prop}$ is the sensitivity matrix of properties, $\boldsymbol{\lambda}$ is the original parameter set and $\delta\boldsymbol{\lambda}$ is the perturbation of that parameter set. $\langle\cdot\rangle_{\boldsymbol{\lambda}}$ denotes the average over trajectory, which is also the ensemble average of parameter set $\boldsymbol{\lambda}$, approximately. The approximate Jacobian $\boldsymbol{S}_{prop}$ was calculated in a parallel fashion because the partial derivatives can be computed independently. In the next step, the prediction for the best new set of nonbonded parameters $\boldsymbol{P} = \boldsymbol{\lambda} + \Delta\boldsymbol{P}$ is given by solving Eq. 6 in a least-square manner

$$\min! \|\boldsymbol{W}(\boldsymbol{S} \cdot \Delta\boldsymbol{P} - \boldsymbol{F})\| \tag{6}$$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_p & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{C36} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_{N_{prop}} \end{pmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{pmatrix} w'_1 & & \\ & \ddots & \\ & & w'_{N_{param}} \end{pmatrix} \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{prop} \\ \boldsymbol{S}_{param} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} s_{1,1} & \cdots & s_{1,N_{param}} \\ \vdots & \ddots & \vdots \\ s_{N_{prop},\,1} & \cdots & s_{N_{prop},N_{param}} \end{pmatrix} \\ \boldsymbol{I}_{N_{param}} \end{pmatrix}$$

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_{prop} \\ \boldsymbol{F}_{C36} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}^{exp} - \langle \boldsymbol{f}^{sim} \rangle_{\boldsymbol{\lambda}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $\boldsymbol{W}$ is a diagonal weight matrix composed of apparent weights for the target properties ($\boldsymbol{W}_p$) and weights used for regularization ($\boldsymbol{W}_{C36}$); $\boldsymbol{S}$ is the complete sensitivity matrix, which consists of the sensitivity matrix of the target properties and an identity matrix of size N_{param} (the total number of parameters); $\boldsymbol{F}$ is the target vector, which contains the deviations of simulated results from the experiments and N_{param} zeros to restrain the parameter set to C36. The least-square problem was solved by the "linalg.lstsq" method in NumPy, which minimizes the Euclidean 2-norm $\|\boldsymbol{W}(\boldsymbol{S} \cdot \Delta\boldsymbol{P} - \boldsymbol{F})\|$ when the equation is overdetermined.[140] While the number of parameters is not necessarily less than the number of experimental targets (which means underdetermination of the problem in the language of linear algebra), we can avoid this situation by adding parameter restraints (which correspond to $\boldsymbol{F}_{C36}$ in the $\boldsymbol{F}$ vector) as additional equations to make sure the number of equations are more than the

number of parameters. These additional equations all have the simple form of $\Delta P_i = 0$ (P_i stands for the i^{th} parameter) and correlate with the $\boldsymbol{W}_{\text{C36}}$ part in the weight matrix and the identity matrix $\boldsymbol{I}_{N_{param}}$ in the sensitivity matrix. They serve as restraints and ensure that Eq. 6 is overdetermined, and a least-square solution is available. Dihedral parameters associated with the changed nonbonded parameters were adjusted at the end of each optimization cycle through reweighting or *direct fitting* (see Section 2.2.5). A new set of simulations was subsequently run to test the new parameters. If the results were satisfactory, the optimization would be terminated. If not, the cycle would continue until optimal parameters were obtained.

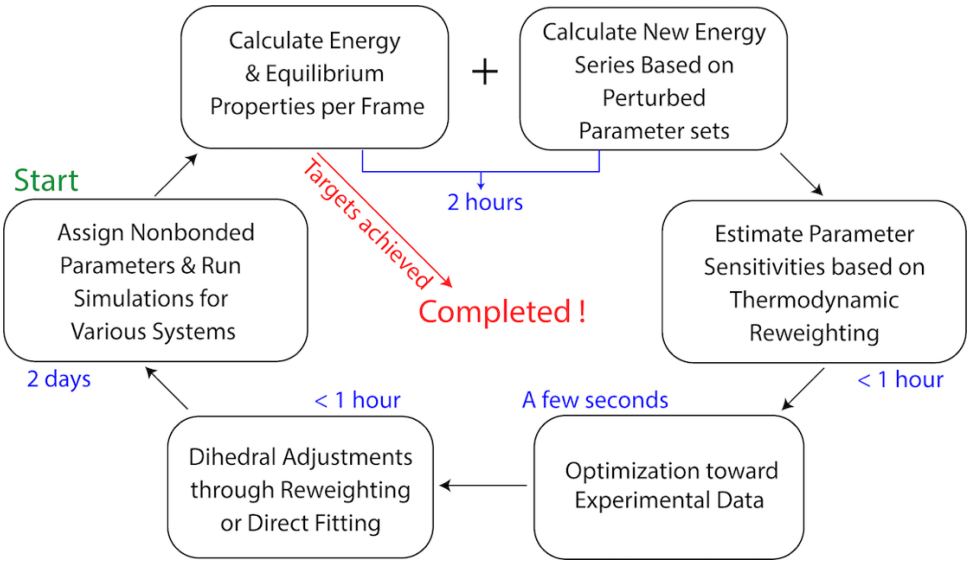

Figure 4. The optimization procedure and approximate time per step.

2.2.5. Dihedral fitting

Dihedral parameters play a critical role in bond orientations and dynamics at the atomic level. The C36 lipid FF yields excellent agreement with the experimental

deuterium order parameters in the headgroup region,[42, 141] which can be credited to the extensive QM calculations of model compounds representative of the headgroup and linkage region. Although fitting to the same set of QM data is a reasonable strategy to take for C36/LJ-PME, running the QM calculations has the added difficulty because one needs to find the conformations that are representative of the lipid in a biological environment. In fact, the C36 dihedrals are good targets for two reasons. First, the potential energy scans are relatively easy to obtain and integrate into the workflow. Second, fitting to C36 is expected to reproduce the order parameters which are already in good agreement with experiments. When fitting to C36, two methods were used. The first is fitting to the potential energy scan (PES) of a dihedral using a model compound. The second is fitting through thermodynamic reweighting (Eq. 5). In the first method, all dihedrals but the scanned dihedral were unconstrained, while the scanned dihedral was restrained using a force constant of 10^3 kcal/mol during the energy minimization consisting of 200 steepest decent steps and 1000 adopted basis Newton-Raphson steps in CHARMM. In the second method, a 10-ns trajectory block from a DPPC bilayer simulation was used to calculate the torsion angles and the corresponding energy series. In addition, a trajectory sampled by C36 was used to obtain the reference torsion distributions. The dihedral parameters were changed iteratively using the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm implemented in scipy[142] until the convergence criteria were met. Table 5 shows the dihedrals of ester-linked lipids fitted in each optimization cycle and the corresponding fitting protocols. The choice of fitting protocol was conditioned on the availability of QM conformational energies from the

original publication of C36.[19] A Monte Carlo simulated annealing approach[143] was used for fitting to the QM/C36 PES.

Table 5. Dihedrals fitted in each optimization cycle and the corresponding fitting protocols. Atom names and model compounds are shown in Figure 5. CHARMM atom types included in parenthesis for each dihedral. Symbols for the dihedrals are consistent with C36 and are used henceforth.

Dihedral	Symbol	Fitting Protocol	Model Compound
O11-C1-C2-C3 (OSLP-CTL2-CTL1-CTL2)	θ_1	Reweighting to C36 Torsion Distribution in DPPC Bilayer Simulation	None
C2-C3-O31-C31 (CTL1-CTL2-OSL-CL)	γ_1		
C1-C2-O21-C21(CTL2-CTL1-OSL-CL)	β_1		
O11-C1-C2-O21 (OSLP-CTL2-CTL1-OSL)	θ_2		
O21-C2-C3-O31 (OSL-CTL1-CTL2-OSL)	θ_4		
P-O12-C11-C12 (PL-OSLP-CTL2-CTL2)	α_4	Least Squares Fit to C36 Potential Energy Scan	PMP
P-O11-C1-C2 (PL-OSLP-CTL2-CTL1)	α_1		PMP
O31-C31-C32-C33 (OSL-CL-CTL2-CTL2)	γ_3		EGLY
C31-C32-C33-C34/C21-C22-C23-C24 (CL-CTL2-CTL2-CTL2)	β_4 / γ_4	Least Squares Fit to QM Potential Energy Scan	IPB/PB

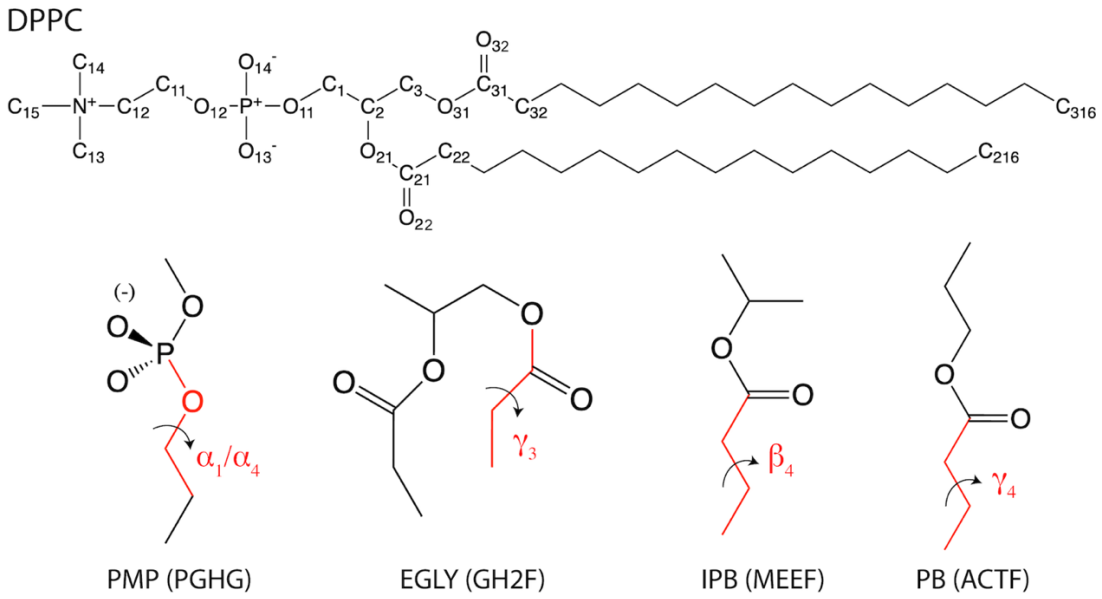

Figure 5. Chemical structure of DPPC and model compounds used for dihedral fitting: propylmethyphosphate (PMP), an esterfied glycerol-phosphate analogue (EGLY). isopropylbutyrate (IPB), and propylbutyrate (PB). The CHARMM residue names shown in parenthesis.

2.2.6. Parameter perturbation

As described in Section 2.2.4, thermodynamic reweighting allows fast evaluations of the parameter sensitivities. However, such evaluations require sufficient conformational overlap between the original and perturbed states for accurate prediction, i.e., the perturbation $\boldsymbol{\delta\lambda}$ in eq 5 must be sufficiently small to ensure this criterion is met. A detailed investigation of the magnitude of perturbation is presented in Section 2.3.2. Table 6 lists the perturbations used for all parameter types. The perturbations used for partial charges are 10^{-3} e. Percentage perturbations were used for the LJ parameters, since they differ substantially among different atom types. When

perturbing the partial charge of an atom, the charge was redistributed to selected neighboring atoms to ensure a fixed net charge. The rules for charge re-distribution are shown in Table 7, which takes the CHARMM definition of integer charge groups into consideration so that the re-distribution was always within a charge group. When there were N equivalent atom sites, for example, O22 and O32, perturbations to them occurred at the same time but the perturbation size for each was scaled by $1/N$ to limit the conformational change corresponding to these perturbations. In each optimization cycle, all parameters listed in Table 4 were perturbed and subject to potential changes, thereby providing access to all partial derivatives.

Table 6. Perturbation sizes used for partial charge, *e* and $R_{min}/2$ during optimization.

Parameter Type	Perturbation Size
q	10^{-3} e
e	0.1% of the parameter in the last optimization step
$R_{min}/2$	0.1% of the parameter in the last optimization step

Table 7-a. Rules for charge redistribution for the *Global* optimization. When the charge on the center atom(s) is changed, the opposite change is evenly distributed to the atom(s) to exchange charge with. The transfer of charge is always within an integer charge group defined in CHARMM. Equivalent atomic sites from group 3 and group 4 are listed in the same rows.

center atom(s)	atom(s) to exchange charge with
group 1	
N	C13, C14, C15, C12
H13*, H14*, H15*	H12*
C13, C14, C15	H13*, H14*, H15*
C12	H12*
group 2	
C11	P
H11*	C11
P	O13, O14, O11, O12
O13, O14	O11, O12
O11, O12	C1
C1	HA, HB
group 3 and group 4	
C2	HS
O21 / O31	C2 / C3
C21 / C22	(O21, O22, C22) / (O31, O32, C32)
O22 / O32	C21 / C31
C22 / C32	C21 / C31
(H2R, H2S) / (H2X, H2Y)	C22 / C32
C3	HX, HY

Table 7-b. Rules for charge redistribution for the *Linkage* optimization. Equivalent atomic sites from group 3 and group 4 are listed in the same rows.

center atom(s)	atom(s) to exchange charge with
group 3 and group 4	
C2 / C3	O21 / O31
HS	C2
O21 / O31	(C2, C21) / (C3, C31)
C21 / C22	(O21, O22, C22) / (O31, O32, C32)
O22 / O32	(O21, C21, C22) / (O31, C31, C32)
C22 / C32	C21 / C31
(H2R, H2S) / (H2X, H2Y)	C22 / C32

Table 7-c. Rules for charge redistribution for the ether optimization. Equivalent atomic sites are listed in the same rows.

center atom(s)	atom(s) to exchange charge with
C2 / C3	HS / (HX, HY)
O21 / O31	C2 / C3
C21 / C31	(O21, C22) / (O31, C32)
(H1R, H1S) / (H1X, H1Y)	C21 / C31
C22 / C32	C21 / C31
(H2R, H2S) / (H2X, H2Y)	C22 / C32

2.2.7. Regularization

As shown in Eq. 6, the optimal parameter set $\boldsymbol{P}$ is given by solving a linear equation which used the sensitivity matrix and a weight matrix. The weight matrix consisted of two parts. The i^{th} component of the first part, w_i, was set to be the weight factor of property i divided by the scaling factor of property i (see Table 3 for values). Weight factors were determined by considering the uncertainty of experimental data and the correlation between different properties. The weight factors for tail S_{CD} were significantly lower than the other properties because they are strongly correlated with the surface area. Scaling factors were used to make sure all training targets were at the same order of magnitude. While the user can set the weights for the training targets

(properties) according to their significances and the uncertainties from the experimental measurements, the second part of the weight matrix, $\boldsymbol{W}_{C36}$, is even more important because it determines how far the parameters can shift from the original C36 set. Such restraints, called regularizations, are usually applied to avoid overshooting of parameters during the fitting process. Incorporating the initial parameters into the target vector is equivalent to adding a harmonic restraint as in the L2 regularization used in ForceBalance,[112, 122] To construct a meaningful $\boldsymbol{W}_{C36}$, the relative magnitude of this submatrix (compared to $\boldsymbol{W}_{p}$) should be determined to assure that the change of the parameter set in each optimization step is suitably small. Moreover, the relative weights for charges and LJ parameters should be balanced properly, so that all parameters change in a physically meaningful way. The default weight factors for different parameter types in each optimization cycle can be found in Table 8. Apart from these, the standard deviations of parameter sensitivity estimated from three independent trajectory blocks were used to modify the default weight factors in each optimization step.

Table 8. Default weight factors for different parameter types in each optimization cycle.

optimization cycle	q	ε	$R_{\min}/2$
Global-1	4/e	6/percentage	6/percentage
Global-2	4/e	6/percentage	6/percentage
Global-3	16/e	24/percentage	24/percentage
Linkage-1	4/e	6/percentage	6/percentage
Linkage-2	4/e	6/percentage	6/percentage

To justify this modification, consider a property that is determined by only a subset of the system's coordinates, $\boldsymbol{r_1}$. Let $\boldsymbol{r}$ denote the full coordinate set, and $\boldsymbol{r_2}$

those coordinates not belonging to $\boldsymbol{r_1}$. The potential energy function, $U(\boldsymbol{r})$, can be decomposed as:

$$U(\boldsymbol{r}) = U(\boldsymbol{r_1}) + U(\boldsymbol{r_2}) + U(\boldsymbol{r_1}, \boldsymbol{r_2}) = U(\boldsymbol{r_1}) + U' \tag{7}$$

where $U(\boldsymbol{r_1})$ and $U(\boldsymbol{r_2})$ are contributions from subset $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$, respectively, and $U(\boldsymbol{r_1}, \boldsymbol{r_2})$ is the contribution from the cross-interactions between the two subsets. With this, an element of the sensitivity matrix in Eq. 5 can be reformatted accordingly as:

$$\begin{aligned} s(\lambda, \delta\lambda) &= \langle f^{sim} \rangle_{\lambda+\delta\lambda} - \langle f^{sim} \rangle_{\lambda} \\ &= \frac{\langle f^{sim} e^{-\beta[U_{\lambda+\delta\lambda}(\boldsymbol{r_1}) + U'_{\lambda+\delta\lambda} - U_{\lambda}(\boldsymbol{r_1}) - U'_{\lambda}]} \rangle_{\lambda}}{\langle e^{-\beta[U_{\lambda+\delta\lambda}(\boldsymbol{r_1}) + U'_{\lambda+\delta\lambda} - U_{\lambda}(\boldsymbol{r_1}) - U'_{\lambda}]} \rangle_{\lambda}} - \langle f^{sim} \rangle_{\lambda} \end{aligned} \tag{8}$$

Here, normal fonts are used for s, f^{sim} and λ instead of bold to indicate that we only focus on one particular property and only one parameter. Suppose that the subset $\boldsymbol{r_1}$ and $\boldsymbol{r_2}$ can be chosen in a way such that $U(\boldsymbol{r_1})$ is strongly coupled with the property f^{sim} while U' is only weakly coupled or is uncoupled. In such a case, if parameter λ's influence on $U(\boldsymbol{r_1})$ is substantial, then the property will be dominated by λ as they are strongly coupled with each other through $U(\boldsymbol{r_1})$. As a result, the gradient of f^{sim} in the direction of λ estimated by the reweighting would be meaningful. However, if $U(\boldsymbol{r_1})$ is not substantially influenced by λ, or if a change in λ leads to more change in U' rather than $U(\boldsymbol{r_1})$, the quality of the reweighting will be impaired since U' will act

as a noise to the total potential energy. Although this noise can be reduced by extending the simulation to acquire enough sampling, it is not practical for membrane systems (which contain more than 10^4 atoms). An alternative is to set customized weights for parameters according to the standard deviations of the sensitivities estimated from trajectory blocks. Specifically, a candidate w_λ for parameter λ is calculated through

$$w_\lambda = S_b * \frac{\sum_{prop}\left(w_{prop} \cdot \mathrm{STD}\left(s_{prop}\right)\right)}{\sum_{prop}\left(w_{prop}\right)} \tag{9}$$

where s_{prop} is the sensitivity for a particular property, w_{prop} is the effective weight (weight factor/scaling factor) for the property, which is also used in $\boldsymbol{W_p}$. STD means standard deviation, and S_b is a scaling factor used to balance $\boldsymbol{W_p}$ and $\boldsymbol{W}_{C36}$. The unit of w_λ is 1/e for charges and 1/percentage for ε and $R_{min}/2$. During the optimization, w_λ was used if it was greater than the default weight for the parameter type of λ, otherwise the default was kept. Moreover, if the calculated w_λ was larger than a preset upper bound, the parameter would not be allowed to change in the optimization step. Consequently, parameter changes leading to large uncertainties were avoided or at least reduced in a property-dependent manner. Therefore, confidence was increased for the more important properties. The S_b used in each optimization cycle can be found in Table 9, while the upper bounds were determined to be 20/e for charge or 20/percentage for ε and $R_{min}/2$. Much higher S_b and default weighting factors (Table 9) were used for the last (third) optimization cycle in the *Global* optimization, because the simulated

properties after the second cycle in that optimization were close to their target values so that harder restraints could be applied.

Table 9. Default weight factors for different parameter types in each optimization cycle.

optimization cycle	q	ε	$R_{min}/2$
Global-1	4/e	6/percentage	6/percentage
Global-2	4/e	6/percentage	6/percentage
Global-3	16/e	24/percentage	24/percentage
Linkage-1	4/e	6/percentage	6/percentage
Linkage-2	4/e	6/percentage	6/percentage

2.2.9. Simulation details

Table 2 lists the details of the systems simulated in this study. Initial structures of small (less than 100 lipids) bilayers and monolayers were built from CHARMM-GUI[92] Membrane Builder.[144-146] Systems containing 288 lipids were built by duplicating the 72-lipid systems at the end of 200 ns simulations 4 times. The initial coordinates of the 648-lipid and 1152-lipid systems were end-of-simulation frames taken from Venable et al.[108] and unpublished simulations using the C36 FF. The C_3-PC residue was created by modifying DMPC, removing fatty acid chain C4:C14 and attached H atoms, and converting C3 from methylene to methyl; no new or modified parameters were required. A single molecule was built from the internal coordinates of the residue definition, and then subjected to 80 ns of vacuum Self-Guided Langevin Dynamics (SGLD),[147] with a collision frequency of 1/ps and a dielectric constant of 80 to screen the charges. Nine configurations were randomly selected from the vacuum trajectory and placed on a grid with one molecule at the origin, and the other eight translated to the corners of a 20 Å cube centered on the origin, and randomly rotated. Water coordinates from an equilibrated cube with a 43 Å edge were read, and waters

closer than 2.3 Å to any C_3-PC molecule were deleted. The system was relaxed with 50 steps of steepest descent minimization, followed by 1000 steps of the adopted-basis Newton-Raphson minimization. All PC bilayers were fully hydrated using the same hydration numbers found in the original publication of C36.[19] The hydration numbers of PE and PG bilayers are consistent with two previous studies by Venable et al.[94, 108] DHPC bilayers were hydrated in the same way as Leonard et al.[67] For DPPC monolayers, lipid headgroups from different leaflets were separated by a water slab, and the tails were separated by vacuum which measures about 180 Å in the monolayer normal direction. For all systems, the modified TIP3P water model[58, 59] was used to keep consistency with the rest of the CHARMM FF. Simulations of PG bilayers included 0.1 M of NaCl. Bilayers were simulated in tetragonal boundary conditions in one of two ensembles: NPT (constant number, pressure of 1 atm, and temperature), or NPγT (constant number, normal pressure, surface tension, and temperature). Monolayers were simulated in the $NL_z\gamma T$ (constant number, cell length normal to the surface, surface tension, and temperature) ensemble.

In each optimization cycle, small PC bilayers and monolayers were simulated for 200 ns except the two NPγT simulations of DPPC bilayers used to calculate the area compressibility, which were simulated for 300 ns. The simulation of C_3-PC aqueous solution was 100 ns in each optimization cycle. All simulations were run in OpenMM 7.4.1 using the Langevin Integrator with a friction coefficient of 1/ps and a 2 fs timestep. Pressure was controlled by the Monte Carlo Membrane Barostat at 1 atmosphere for membranes and by the Monte Carlo Barostat at 1 atmosphere for the C_3-PC solution. The real-space cutoff was set to be 10 Å for LJ-PME. The Ewald error

tolerance was set to be 10^{-4}. The simulation method used for validation was the same as the optimization except the 288/648-lipid bilayers and an extra set of 72-lipid bilayers (DPPC and DOPC) used to study lipid diffusion, for which a Nosé-Hoover chain velocity Verlet integrator was applied. In addition to the LJ-PME simulations, two monolayers were simulated using the C36 FF and a force-switching function ranging from 8 to 12 Å for the LJ, and the NPγT ensemble was used in which γ was set to be 18 and 41 dyn/cm.

2.2.10. Property calculation

The following properties were evaluated during the optimization for the training set and after for the validation set:

1. A_l was calculated as the area of the simulation cell in the x-y plane (tangential to the membrane surface) divided by the number of lipids per leaflet.

2. Area compressibility, K_A. As discussed in Section 2.2.2 (Eq. 4), K_A can be calculated in two ways. For parameter optimization, K_A was obtained from $d\gamma/dA$, and the reweighting was done to the surface area of the two DPPC bilayers under applied surface tensions of ±5 dyn/cm.

3. Deuterium order parameters were computed from

$$S_{CD} = \left|\left\langle \frac{3}{2}\cos^2\theta - \frac{1}{2}\right\rangle\right| \quad (10)$$

where θ is the angle between the C-H vector and the bilayer normal, the angular bracket denotes the time and ensemble average.

4. The overall bilayer thickness, D_B, was estimated by the distance between the two midpoints (one for each side of the bilayer) of water electron density profile (EDP) along the bilayer normal, for which a bin size of 0.2 Å was used. EDPs for lipid atoms were also calculated in the validation process, from which the headgroup-to-headgroup distance (D_{HH}) and the hydrophobic thickness (D_C) were estimated. D_{HH} is defined as the distance between the peaks in the overall EDP, and D_C is half of the distance between the midpoints of EDP for the acyl chain. From EDPs, X-ray and neutron form factors were generated using the SIMtoEXP program,[148] which were compared directly to experiments.

Dynamic properties of lipid bilayers were studied as model validation. The first set of properties are the NMR spin lattice relaxation time T_1 of C-H bonds. T_1 was calculated assuming pure dipolar relaxation between the ^{13}C nucleus and its N attached protons using the formula[149]

$$\frac{1}{NT_1} = \left(\frac{h\gamma_C\gamma_H}{2\pi {r_{C-H}}^3}\right)^2 [J(\omega_H - \omega_C) + 3J(\omega_C) + 6J(\omega_H + \omega_C)] \qquad (11)$$

$$J(\omega) = \int_0^\infty \langle P_2(\hat{\mu}(0) \cdot \hat{\mu}(t))\rangle \cos(\omega t)\, \mathrm{d}t$$

where h is the Plank's constant, r_{C-H} is the effective C-H bond length, γ_C and γ_H are the gyromagnetic ratios of the ^{13}C and H nuclei, ω_C and ω_H are the angular Larmor frequencies. $J(\omega)$ is the spectral density of the second rank reorientational

correlation function (P_2) of the C-H bond direction, where $\hat{\mu}(t)$ is the unit vector along the C-H bond at time t.

Another set of properties are the lipid diffusion constants. Comparing simulated diffusion constants with experiments faces two major challenges. The first is the artifact introduced by the period boundary conditions (PBC) common to molecular dynamics (MD) simulations.[150] For membrane spanning proteins, Camley et al.[151] predicted that the diffusion constant would be underestimated by factors 3-10, and recent studies by Venable et al.[152] and Zgorski et al.[153] indicate that lipid diffusion could be affected by similar factors. To overcome this challenge, DPPC and DOPC bilayers of different sizes were simulated and the periodic Saffman-Delbrück (PSD) model developed by Camley et al.[151] was used to extrapolate diffusion constants at infinite system size. The second challenge is determining the viscosity of the lipid molecule in a membrane environment, which must be used as an input parameter in the PSD model. More details on this will be discussed in Sections 3.4 and 4. In diffusion constant calculations, the center of mass (COM) motion of the bilayer should be removed to get the motions of individual lipids relative to the bilayer. It was found in previous MD simulations[30, 154] that the two leaflets of small bilayers (36 lipids/leaflet) have fast relative motions, so that COM motions for individual leaflets were removed instead of the COM motion of the whole bilayer. Two-dimensional (2D) mean-squared displacement (MSD) for each lipid as a function of time was computed via a difference correlation function using the x and y components of the lipid COM.[152] The 2D diffusion constants were computed from the slopes of $\langle msd \rangle_{LIPID}$ vs. time, for 10 ns < t < (t_i – 10)/2, where t_i is the analysis interval (80 ns). The lower cutoff removes the sub-diffusive dynamics, which is

significantly faster than the diffusion in all-atom simulations.[152] Standard errors were estimated based on 3 replicas for each system, and each replica contains 4 (72- and 288-lipid systems) or 2 (648-lipid systems) trajectory blocks of 80 ns.

2.3. Results

2.3.1 Parameter sensitivities

The sensitivity analysis was performed on all training targets associated with the PC headgroup starting from the first optimization cycle. Sensitivities for the surface area of DPPC bilayer at 323.15 K are reported in Figure 6; examples for other training targets can be found in the published paper.[155] The sensitivities were calculated according to Eq. 5, and the perturbation sizes $\boldsymbol{\delta\lambda}$ were determined by monitoring the effects of the size (see Section 2.3.2). A perturbation size of 0.001 elementary charge (e) was chosen for partial charges (about 0.3% the average absolute value of partial charges from headgroup atoms), and percentage perturbations (0.1%) were used for $R_{\text{min}}/2$ and ε due to the wide distributions of these two parameters among different atom types.

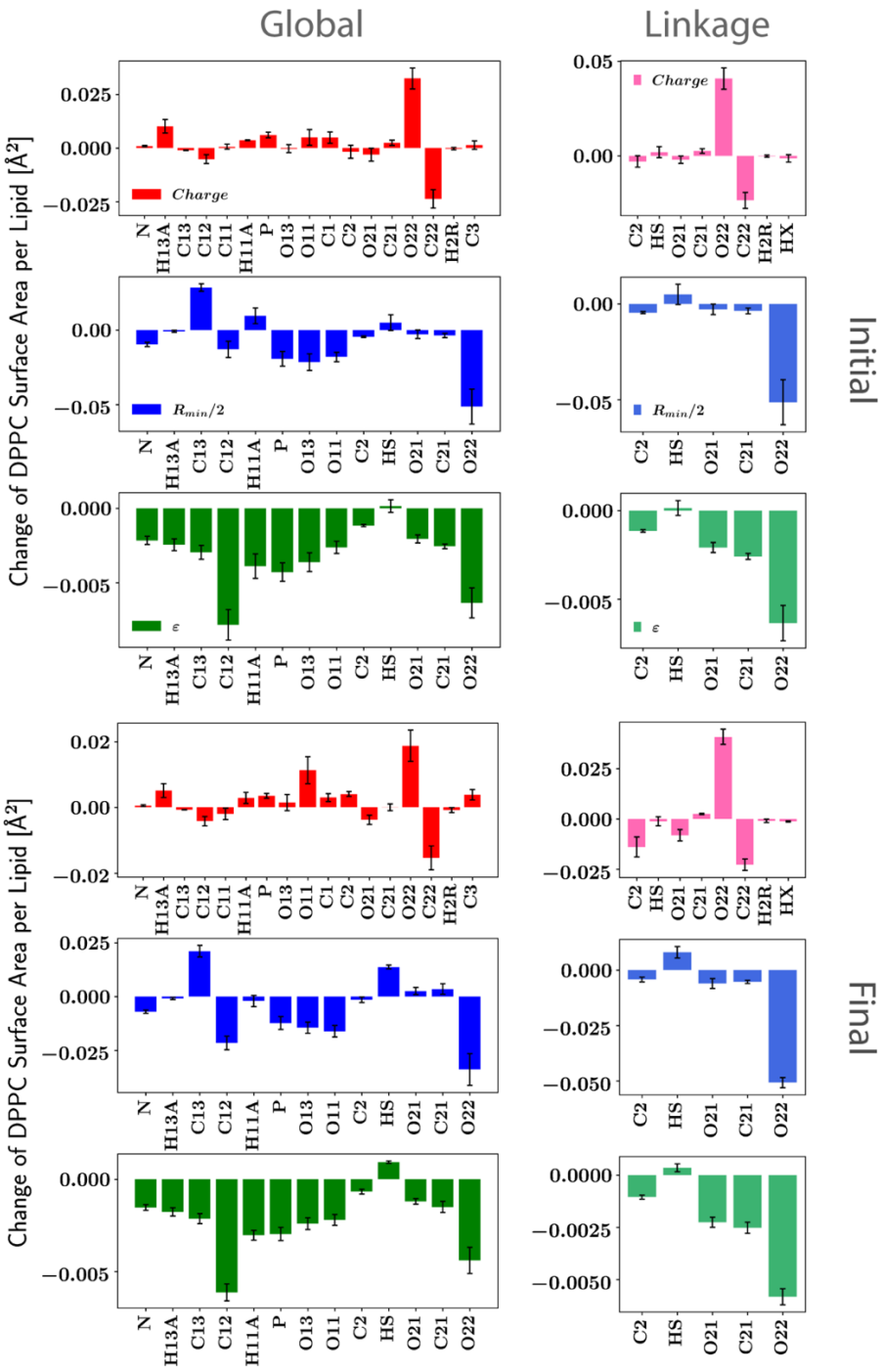

Figure 6. Parameter sensitivities for the surface area of DPPC bilayer at 323.15 K. Averages and standard deviation evaluated from three blocks of 50 ns shown for each parameter. The perturbation sizes used are 0.001 e (absolute value) for partial charge and 0.1% of the original parameter for $R_{min}/2$ and ε. Note that different scales were used for the plots of initial and final parameter sets.

It is evident from Figure 6 that different parameters have different effects on the surface area of DPPC bilayer and its standard deviations. While the perturbations used for the three nonbonded parameter types (charge, $R_{min}/2$ and ε) are comparable, ε has much less influence on the surface area compared to the other two, and the sensitivities are generally negative (when increasing the absolute value of ε). This is not surprising since ε does not substantially influence the minimum energy distance between two interacting groups, though the distance distribution for an atom pair is affected by the interplay of all nonbonded parameters (and bonded parameters if linked). The $R_{min}/2$ sensitivities do not have a uniform sign, but most are negative. This is counterintuitive as one might expect the volume to increase with larger R_{min}, hence an increased surface area. However, it is the intersection of the LJ surfaces of the two interacting groups that impact the minimum interaction distances such that the relationship of R_{min} to interaction distance is not necessarily directly correlated. This result highlights the complexity of interactions in lipid bilayers. For partial charges, the "hot spots" are the two carbonyl groups (C22-O22/C32-O32) at the head-tail linkage region. Moreover, the O22 $R_{min}/2$ also has a significant impact on the surface area. Because these atoms are located at the head-tail linkage, two optimizations were performed. As detailed in the Methods, *Linkage* restricted the changes to the glycerol backbone and the ester groups, while *Global* also allowed the phosphate and choline groups to vary. The surface area of DPPC was not the only target for parametrization, and the sensitivity for a single property can only tell us the importance of the parameter for that particular property, but not necessarily others. This is illustrated by Figure 7 and Figure S3 of the original publication.[155]

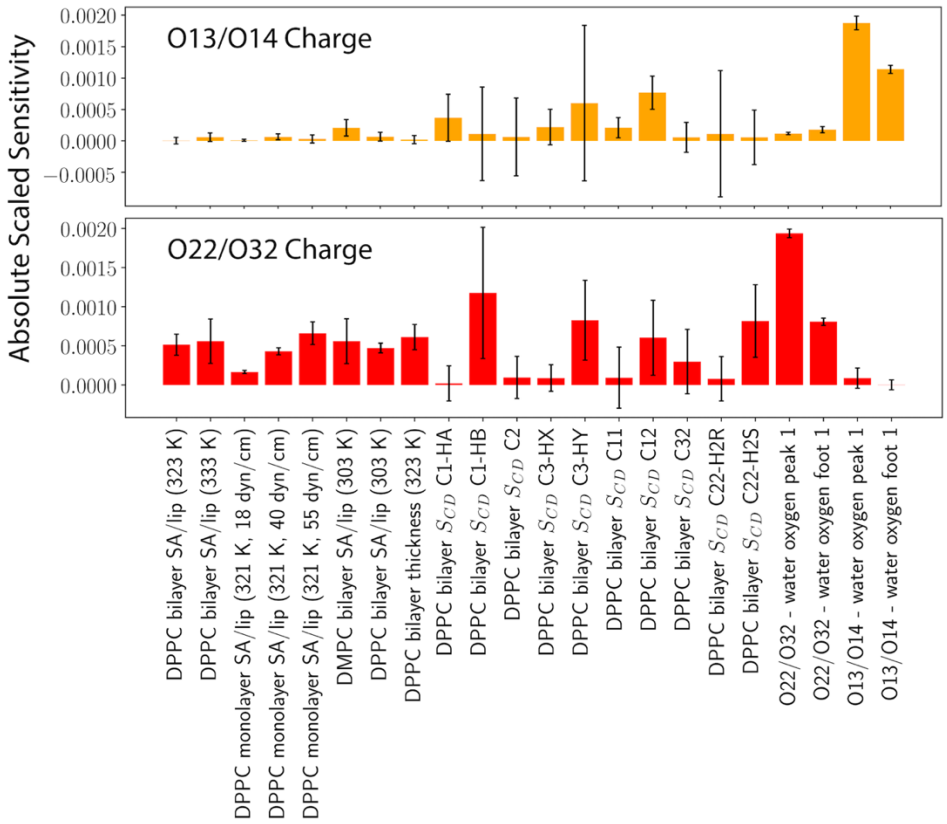

Figure 7. Absolute scaled parameter sensitivities of O13/O14 and O22/O32 partial charges for various properties.

Figure 7 plots the parameter sensitivities of O13/O14 and O22/O32 charges on a wide range of properties, which are scaled by the scaling factors in Table 3 after taking the absolute values. The parameter perturbations used are the same as those used in Figure 6. The standard deviations of the sensitivities for different properties vary significantly. For example, one can be almost certain about O13/O14 charges' influence on the RDF between O13/O14 and the water oxygen but cannot be so confident with the A_l of DPPC/DMPC bilayers. However, when parametrizing the force field, the influence of the O13/O14 charges on the A_l is nonnegligible, which makes the parametrization challenging.

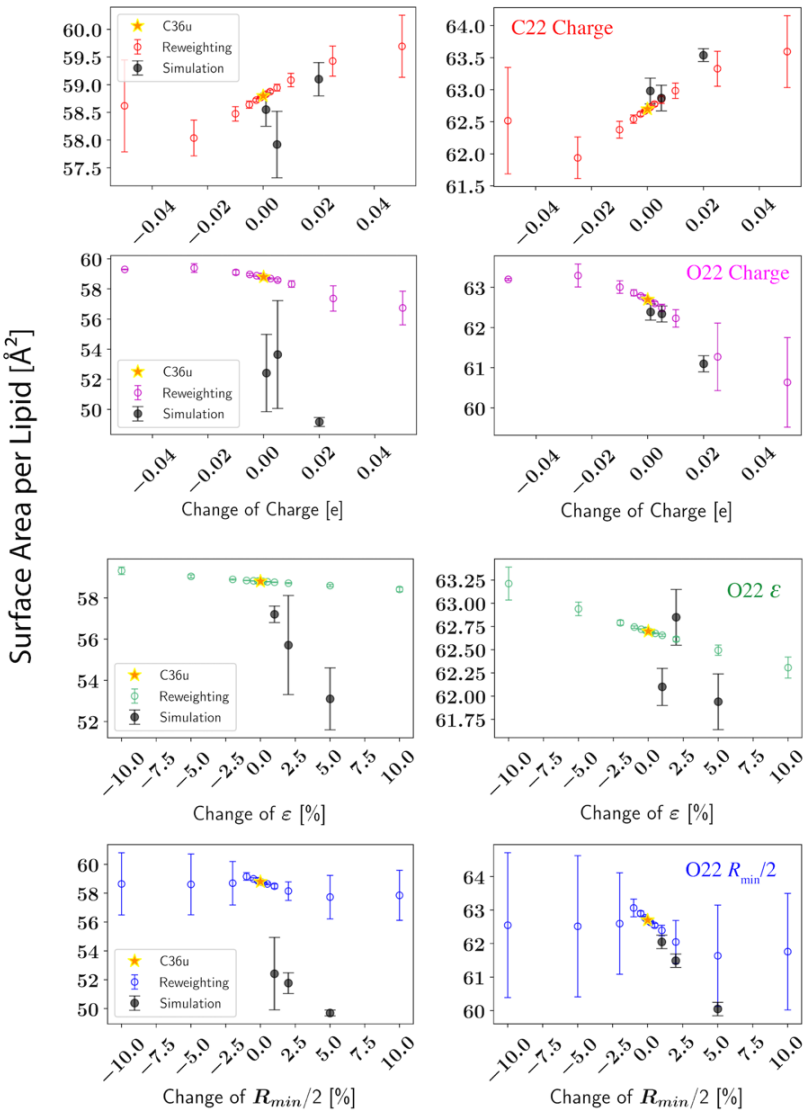

Figure 8. Simulated and reweighted surface area per lipid for DPPC bilayer at 323.15 K. All simulations are performed with LJ-PME. Yellow stars (C36u): simulations using the original C36 parameter set (left) or the final parameter set (right) from the *Global* optimization; black dots with error bars (Simulation): simulations using perturbed parameters; circles with error bars: predictions by reweighting the yellow-star simulations. The averages and standard deviations shown at error bars of the simulated results are based on 3 independent replicas of 300 ns (last 200 ns used for calculating the A_l), while the averages and standard deviations shown at error bars of the reweighted are from 3 trajectory blocks of 50 ns.

2.3.2. Scope of the linear prediction

Although Reweighting can potentially increase the efficiency of the parametrization, one should be careful about the parameter sensitivities predicted in this way. The reason is simple – to get meaningful sensitivities (gradients), the trajectory sampled by the simulation should represent an equilibrated state. And perhaps more importantly, the gradients calculated at a certain point in the parameter space are only accurate for a specific region around that point, which is hereafter called the linear region.

To identify the size of the linear region and to justify the reweighting strategy, reweightings using different perturbation sizes were performed; selected results are shown in Figure 8. For both the starting point (C36) and the end point (examples only shown for the *Global* parametrization), there are clearly linear regions, but the size of the linear region depends on the parameter. In the view of reweighting, the linear region is typically several hundredths of the elementary charge for partial atomic charges. Interestingly, this region is about 3% for $R_{min}/2$ but more than 10% for ε. Moreover, the statistical errors calculated based on 3 trajectory blocks indicate that reweighting for $R_{min}/2$ is no longer accurate when the perturbation exceeds 3%. Considering $R_{min}/2$ is directly related to the pairwise interaction between two atoms close to each other, changing $R_{min}/2$ would lead to a dramatic change in the conformational sampling so that the overlap between the sampled and reweighted states would be greatly reduced, thus leading to large statistical errors. As a result, weight factors for regularization were different for the three parameter types.

The linear region is harder to detect from direct simulations due to large statistical errors, while the physics-informed method (reweighting) can greatly reduce these errors if parameter perturbations allowing sufficient overlap between the original and the reweighted ensembles are used. However, simulations can still be used to determine the perturbation sizes and the upper bounds of parameter moves in each optimization step. Extreme cases can also be detected by direct simulations. For example, several simulations near the C36 parameter set condensed to the gel phase leading to huge statistical errors on the mean surface area. Since this only happened to the perturbed set, the microstates representing the gel phase were not captured by the simulation with the C36 set. Hence, the formation of a gel phase was not predicted by reweighting. Ideally, one would like to avoid such a scenario where the trajectory used to reweight does not sample some of important regions of the target state. However, the energy landscape of a lipid bilayer is so complex that it is almost certain to be nonergodic. Fortunately, as the optimization went on, the gel phase region of the parameter space was no longer sampled. Therefore, we can have more confidence in the reweighting. This is evident from the bottom row of Figure 8, where the reweightings agree well with the simulations, at least within the linear region.

2.3.3. Perturbation sizes and optimization weights.

Based on the linear region detected in Section 2.3.2, the perturbation sizes for the three parameter types can be readily determined. To ensure a well-behaved computation of sensitivities, perturbations were set to be much smaller than the sizes of the linear region (see Table 3). As noted in the Section 2.6, regularization on the parameter change is necessary to avoid overshooting and can be achieved by including

the parameter change into the objective function through a user-defined weight factor and a sensitivity of 1. Based on the results of Section 2.3.2 and a set of tests on different values, the default weight factors were determined to be 4/e for charges, and 6/percentage-change for ε and $R_{min}/2$ in the first optimization cycle, because they generated parameter changes within the linear region. Apart from the default values, a candidate weight factor w_λ for parameter λ were computed according to Eq. 9 and used it instead when larger than the default. Lastly, when w_λ was larger than 20/e for charge or 20/percentage for ε and $R_{min}/2$, the weight factor for parameter λ was set to be infinite, since a high w_λ indicates large noise in the reweighting.

2.3.4. Dihedral parameters.

Dihedrals of ester-linked lipids were fitted to either QM or C36. Conformational energies were obtained for α1, α4 and γ3 from C36 by scanning the dihedrals of interest from -180° to 180° in increments of 2°, while other dihedrals were relaxed during the scan to get the minimum energy conformation. QM energies from ref 4 were used for β4 and γ4. The scan of β4 was from -180° to 0° and the scan of γ4 was from 0° to 180°. Since the two dihedrals contain the same four atom types, the fitting was then conducted by giving each dihedral an equal weight. For dihedrals closer to the glycerol region, reweightings were used to match the C36 dihedral distributions in DPPC bilayer at 323.15 K. In addition to the standard dihedral fittings for each optimization cycle shown in Table 2, additional adjustments were made to N-C12-C12-O12 (α_5) and α_4 at the end of the *Linkage* optimization to match the experimental order parameters of carbon C11 and C12. The two torsions were selected by checking the correlations between the order parameters and all the headgroup torsions. Then, the

same procedure used for the nonbonded parametrization was used to optimize the torsions (reweighting + least-square fitting) because only small changes were expected. Multiplicities of all dihedrals were kept as they were in C36 for *Linkage* but allowed to expand for *Global*. The PES of dihedrals optimized through direct fits are plotted in Figure 9(a), in which the increments for $\alpha 1$, $\alpha 4$ and $\gamma 3$ are 10° while the increments for $\beta 4$ and $\gamma 4$ are those from the original scan. Figure 9(b) illustrates the quality of the reweighting for the last optimization cycle. Overall, the optimized set agrees very well with C36.

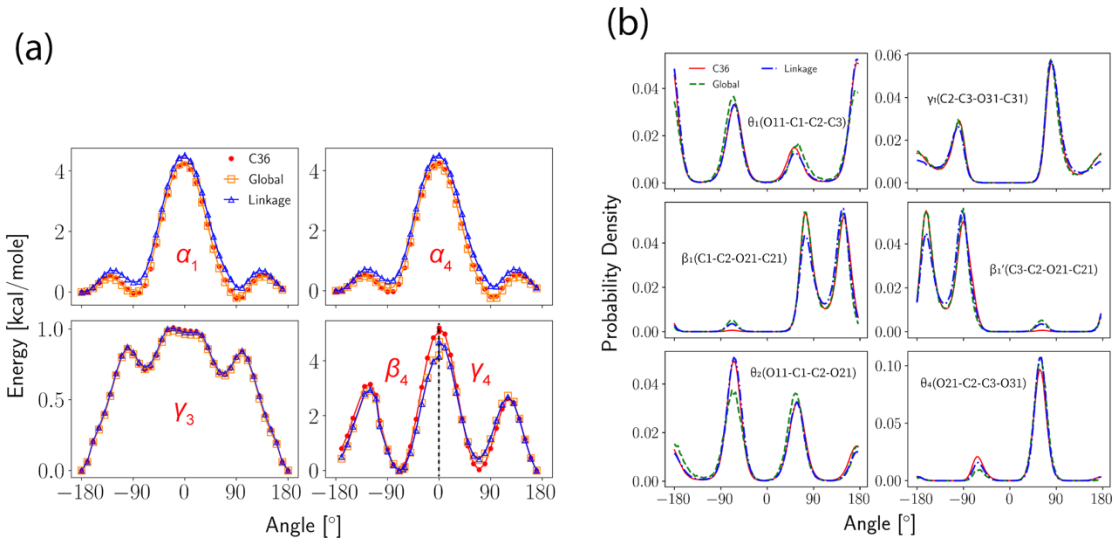

Figure 9. Dihedral fits. (a) Potential energy scans for $\alpha 1$, $\alpha 4$, $\gamma 3$, $\beta 4$ and $\gamma 4$. Reference energy is set to be the lowest energy of the *Linkage* parametrization; (b) Dihedral angle distributions for θ_1, γ_1, β_1' (and β_1), θ_2 and θ_4. The simulation used to reweight is not shown.

For the ether linkage, parameters for the C-O-O-C and O-C-C-O dihedrals were fitted to the QM potential energy scan of 1,2-diethoxyethane (C6H14O2; hereafter

DEOE) from Leonard et al.[67] and the optimized parameters are compared to the QM in Figure 10.

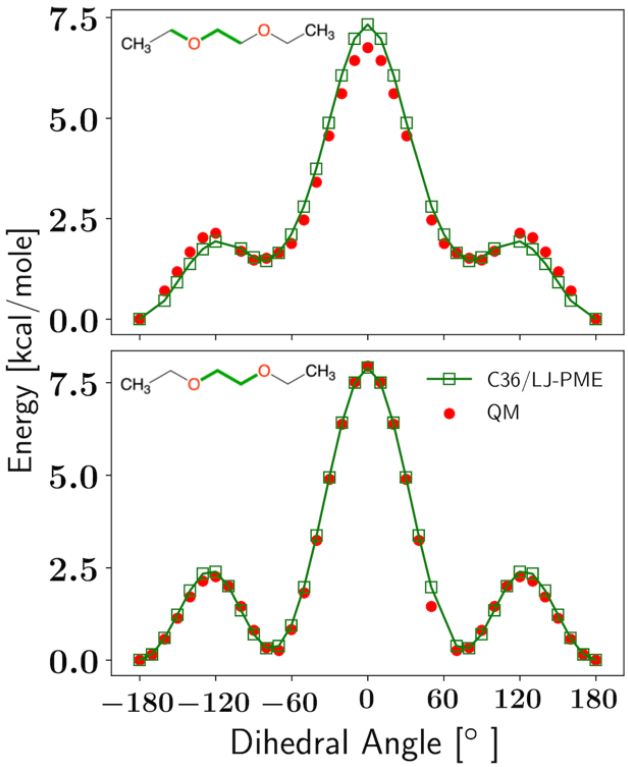

Figure 10. Dihedral fittings for the C-O-C-C and O-C-C-O dihedrals. The chemical structure of DEOE and the positions of the two dihedrals (green) are shown on the upper left of each panel. All dihedrals not used for the scan are fixed at 180°.

2.3.5. Nonbonded parameters

Three and two cycles were performed for the *Global* and *Linkage* optimization respectively until satisfactory agreement with the target data. For the ether linkage, only one cycle of optimization was performed. It should be noted that an extra cycle of optimization was always performed to ensure no further improvement can be made. In principle, the optimization protocol can be set to terminate automatically according to a preset convergence criterion, for example, a threshold for a function using the relative error

of the target properties and the change of the parameters as inputs. However, there are many properties involved in this parametrization and, therefore, determining a general meaningful threshold beforehand is not practical. This led us to use human interruptions when the properties are satisfactory. However, the program can be easily modified to allow automatic termination if needed. Table 10 compares the final nonbonded parameters of the DPPC headgroup to the original C36 FF. As will be shown in Section 2.3.6, there is no qualitative difference of membrane properties between the *Global* and *Linkage* optimizations, so that the Linkage optimization was chosen as the new C36/LJ-PME lipid FF. To keep consistent with the ester-linked lipids, only the partial charges of the ether linkage was optimized for the ether lipids and the comparison with the original C36 FF is presented in Table 11.

Table 10-a. Partial atomic charges (e) for the DPPC headgroup.

atom name	**atom type**	**C36**	***Global***	***Linkage***
N	NTL	-0.60	-0.6030	-0.6
C13	CTL5	-0.35	-0.3451	-0.35
H13A	HL	0.25	0.2542	0.25
H13B	HL	0.25	0.2542	0.25
H13C	HL	0.25	0.2542	0.25
C14	CTL5	-0.35	-0.3451	-0.35
H14A	HL	0.25	0.2542	0.25
H14B	HL	0.25	0.2542	0.25
H14C	HL	0.25	0.2542	0.25
C15	CTL5	-0.35	-0.3451	-0.35
H15A	HL	0.25	0.2542	0.25
H15B	HL	0.25	0.2542	0.25
H15C	HL	0.25	0.2542	0.25
C12	CTL2	-0.10	-0.0551	-0.1
H12A	HL	0.25	0.2027	0.25
H12B	HL	0.25	0.2027	0.25
C11	CTL2	-0.08	-0.0793	-0.08
H11A	HAL2	0.09	0.0664	0.09
H11B	HAL2	0.09	0.0664	0.09
P	PL	1.50	1.5308	1.5
O13	O2L	-0.78	-0.7816	-0.78
O14	O2L	-0.78	-0.7816	-0.78
O11	OSLP	-0.57	-0.5307	-0.57
O12	OSLP	-0.57	-0.5307	-0.57
C1	CTL2	-0.08	-0.1481	-0.08
HA	HAL2	0.09	0.0942	0.09
HB	HAL2	0.09	0.0942	0.09
C2	CTL1	0.17	0.0988	0.1339
HS	HAL1	0.09	0.1626	0.1023
O21	OSL	-0.49	-0.4919	-0.4739
C21	CL	0.90	0.8579	0.8445
O22	OBL	-0.63	-0.6560	-0.6272
C22	CTL2	-0.22	-0.1565	-0.1652
H2R	HAL2	0.09	0.0926	0.0928
H2S	HAL2	0.09	0.0926	0.0928
C3	CTL2	0.08	0.1139	0.0302
HX	HAL2	0.09	0.0738	0.1030

HY	HAL2	0.09	0.0738	0.1030
O31	OSL	-0.49	-0.4919	-0.4739
C31	CL	0.90	0.8579	0.8445
O32	OBL	-0.63	-0.6560	-0.6272
C32	CTL2	-0.22	-0.1565	-0.1652
H2X	HAL2	0.09	0.0926	0.0928
H2Y	HAL2	0.09	0.0926	0.0928

Table 10-b. ε (kcal/mol) for the DPPC headgroup. Only changed parameters are listed.

atom name	atom type	C36	*Global*	*Linkage*
O21	OSL	-0.1000	-0.0991	
C21	CL	-0.0700	-0.0693	
O22	OBL	-0.1200	-0.1185	-0.1192
O31	OSL	-0.1000	-0.0991	
C31	CL	-0.0700	-0.0693	
O32	OBL	-0.1200	-0.1185	-0.1192

Table 10-c. $R_{min}/2$ (Å) for the DPPC headgroup. Only changed parameters are listed.

atom name	atom type	C36	*Global*	*Linkage*
O21	OSL	1.650	1.6380	1.6809
C21	CL	2.000	1.9750	
O22	OBL	1.700	1.6789	1.6470
O31	OSL	1.650	1.6380	1.6809
C31	CL	2.000	1.9750	
O32	OBL	1.700	1.6789	1.6470

Table 11. Partial atomic charges (e) for the ether linkage.

atom name	C36	C36/LJ-PME
C2	0.13	0.1494
HS	0.01	0.0100
O21	-0.56	-0.6228
C21	0.40	0.4142
H1R	0.01	0.0108
H1S	0.01	0.0108
C22	-0.18	-0.1556
H2R	0.09	0.0916
H2S	0.09	0.0916
C3	0.08	0.0706
HX	0.02	0.0344
HY	0.02	0.0344
O31	-0.56	-0.6228
C31	0.40	0.4142
H1X	0.02	0.0208
H1Y	0.02	0.0208
C32	-0.18	-0.1556
H2X	0.09	0.0916
H2Y	0.09	0.0916

2.3.6. Equilibrium properties

Table 12 compares the A_l for bilayers including PE and PG lipids. The overall agreement with experiment for C36/LJ-PME is just as good as C36 in terms of bilayers. While A_l of DMPC bilayer is higher than experiment, X-ray and neutron form factors (Figure 11) agree well with experiment and are comparable to those calculated using

C36.[62] The situation is more complicated for DLPC, for which the neutron form factors deviate from experiment by noticeable amounts, indicating the potential deficiency of the FF to describe shorter tails. As the only ether lipid optimized and tested in this study, A_l of DHPC bilayers match the experimental data well and are comparable to C36.

Table 12. A_l from NPT and NPγT simulations. Available C36 and experimental data included for comparison. Standard errors are given in parentheses.

System	Temperature (K)	Ensemble	C36/LJ-PME (Å^2/lipid)	C36 (Å^2/lipid)	Experiment (Å^2/lipid)
DPPC bilayer	323.15	NPT	62.7 (0.2)	62.9 (0.1) [108]	63.1 (1.3) [156]
	333.15		63.3 (0.2)		65.0 (1.3) [156]
DMPC bilayer	303.15	NPT	62.4 (0.2)	61.5 (0.1) [108]	60.6 [135], 59.9 (1.2) [156]
DLPC bilayer	303.15	NPT	64.3 (0.4)	63.1 (0.3) [109]	60.8 (1.2) [156]
DOPC bilayer	303.15	NPT	69.4 (0.2)	68.9 (0.1) [108]	67.4 [157], 72.4 [158]
POPC bilayer	303.15	NPT	65.4 (0.5)	66.0 (0.1) [108]	64.4 (1.3) [156]
POPE bilayer	308.15	NPT	58.9 (0.4)	58.8 (0.1) [108]	58.0 (1.2) [159]
DMPG bilayer	303.15	NPT	59.8 (0.4)	64.0 (0.2) [94]	62.5 (1.3) [160, 161]
	323.15		65.3 (0.5)		66.0 (1.3) [160, 161]
	333.15		66.0 (0.4)		67.5 (1.4) [160, 161]
POPG bilayer	303.15	NPT	65.4 (0.3)	68.0 (0.1) [108]	64.3 (1.3) [160, 161]
DHPC bilayer	321.15	NPT	63.5 (0.2)	63.2 (0.3) [67]	65.1 [137]
	333.15		65.5 (0.3)	65.4 (0.2) [67]	67.2 [136]
DPPC monolayer	321.15	NPγT (γ=18 dyn/cm)	55.7 (0.2)	60.0 (0.2)	54.0 [162]
	321.15	NPγT (γ=41 dyn/cm)	63.9 (0.2)	72.1 (0.2)	64.0 [162]
	321.15	NPγT (γ=55 dyn/cm)	76.5 (0.3)		80.0 [162]

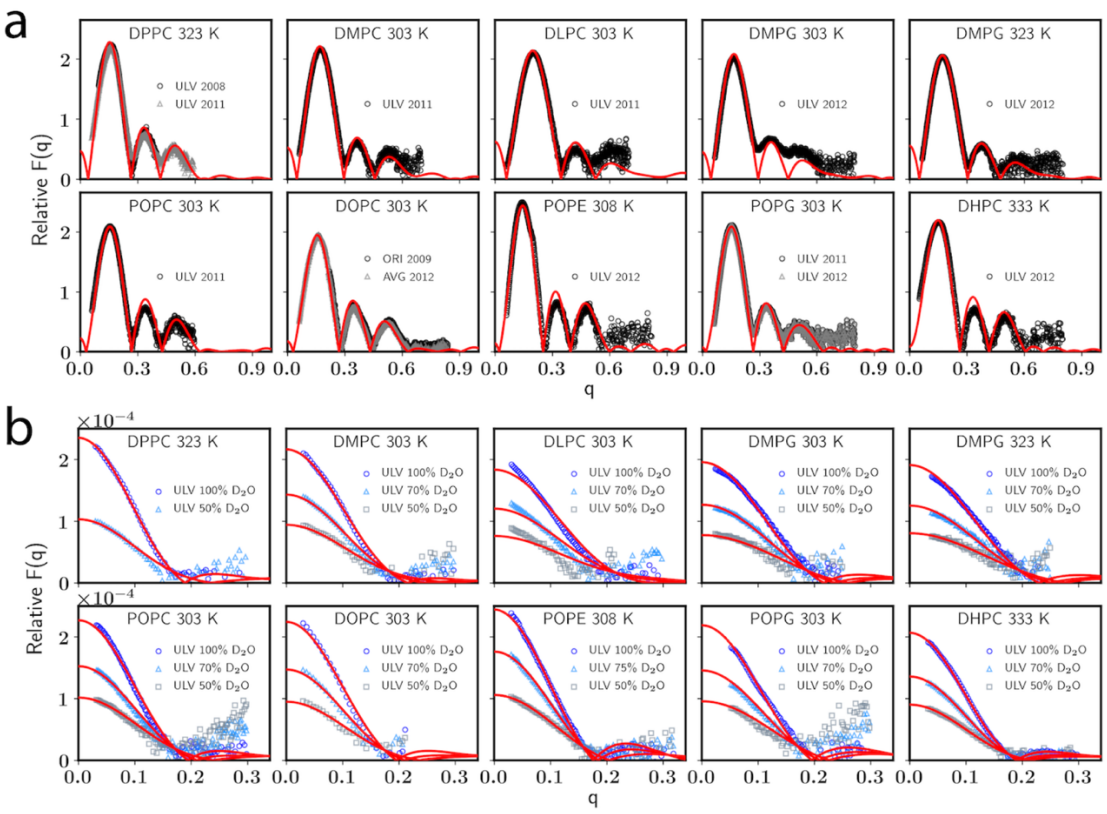

Figure 11. (a) X-ray and (b) neutron form factors for bilayers. C36/LJ-PME in red line and experiments[136, 156-160, 163, 164] in open symbol.

Table 13 compares membrane thicknesses calculated using C36/LJ-PME to experiments. Overall, C36/LJ-PME agrees well with experiments for all tested lipids, though under- and overestimations exist for some lipids. For DMPC, DLPC, POPC and DOPC bilayers, D_B is underestimated, which is consistent with their slightly overestimated A_l. While the average underestimation of D_B is small, this implies an overhydration is needed to generate the accurate A_l for bilayers. Nevertheless, comparison with the experimental form factors indicates that structures of these bilayers are well-modeled. The head-to-head distance D_{HH} and the hydrophobic thickness $2D_C$ follow the same trend of D_B for most lipids. One exception is POPE, for

which D_{HH} is higher than the experiment despite a close-to-experiment D_B. This indicates that headgroup size and intermolecular hydrogen bonds may influence D_{HH} and thus the difference between D_B and D_{HH}.

In addition, the overall bilayer structure measured by A_l (Table 12) and membrane thicknesses (Table 13) implies that PE and PG lipids are well modeled by C36/LJ-PME. However, there might be a potential issue for DMPG at relatively lower temperatures (e.g., 303.15 K) as indicated by the higher-than-experiment chain order parameters, which will be discussed in Section 2.4.

Table 13. Overall thickness (D_B), headgroup-headgroup distance (D_{HH}) and hydrophobic thickness ($2D_C$) for each lipid bilayer for C36/LJ-PME, C36 and experiments (Expt.). Standard errors given in parentheses. Experimental values with uncertainties for saturated PC lipids and POPC from 2011 (Kučerka et al.).[156] DOPC experimental values from earlier study of same group in 2008.[157] PE experimental values from their later study in 2015.[159] PG experimental values from their 2014 paper.[161] DHPC experimental values from two separate publications.[136, 137] C36 values are from Zhuang et al.[62, 109] and Leonard et al.,[67] and standard errors are less than 0.3 Å if not provided.

Lipid	Temperature (K)	D_B (Å)			D_{HH} (Å)		$2D_C$ (Å)		
		C36/LJPME	C36	Expt.	C36/LJPME	Expt.	C36/LJPME	C36	Expt.
DPPC	323.15	38.5 (0.1)	39.6	39.0 (0.8)	37.6 (0.2)	38.4	28.2 (0.1)	28.9	28.5 (0.6)
	333.15	38.3 (0.1)	38.9	38.1 (0.7)	38.0 (0.1)	34.6	28.1 (0.2)	28.5	27.9 (0.6)
DMPC	303.15	34.6 (0.2)	36.2	36.7 (0.7)	33.6 (0.2)	35.3	24.7 (0.1)	25.6	25.7 (0.5)
DLPC	303.15	29.7 (0.1)	31.0	32.6 (0.7)	29.8 (0.2)	29.8	20.3 (0.1)	20.9	21.7 (0.4)
POPC	303.15	37.1 (0.2)	37.4	39.1 (0.8)	37.3 (0.3)	36.5	27.7 (0.2)	28.1	28.8 (0.6)
DOPC	303.15	36.0 (0.2)		38.7	37.2 (0.1)	37.0	27.4 (0.1)		28.8
Absolute Deviation (PC)		4.8%	2.9%		1.9%		3.5%	2.0%	
DHPC	321.15	36.6 (0.3)	38.7 (0.1)		38.4 (0.4)	38.2	27.8 (0.3)	28.9 (0.1)	27.6
	333.15	35.7 (0.3)	37.7 (0.3)	36.8	37.8 (0.3)	38.0	27.3 (0.2)	28.4 (0.1)	27.1
Absolute Deviation (Ether)		3.0%	3.3%		0.5%		0.7%	4.8%	
POPE	308.15	40.7 (0.3)	43.0	40.5 (0.8)	39.8 (0.3)	38.3	31.1 (0.2)	32.6	32.1 (0.6)
Absolute Deviation (PE)		0.5%	6.2%		3.9%		3.1%	1.6%	
DMPG	303.15	34.2 (0.3)	33.3	33.8 (0.7)	34.5 (0.3)	34.6	25.6 (0.3)	24.0	24.5 (0.5)
	323.15	32.3 (0.3)	31.5	32.6 (0.7)	33.2 (0.3)	34.6	24.5 (0.3)	23.0	23.7 (0.5)
	333.15	32.2 (0.2)	31.1	32.0 (0.7)	32.8 (0.1)	33.8	24.2 (0.1)	22.8	23.4 (0.5)
POPG	303.15	36.8 (0.3)	36.2	37.6 (0.8)	37.2 (0.2)	36.6	28.3 (0.2)	27.2	28.5 (0.5)
Absolute Deviation (PG)		1.2%	2.9%		2.2%		3.0%	3.0%	

Another important metric of membrane structure is S_{CD}. C36 is the first lipid FF to accurately capture the chain order splitting for all carbons.[19, 141] Because the dihedral fitting for the PC headgroup was conducted in a way maximizing the consistency with C36, and post-fit refinement was used to further improve the agreement with experiment. Hence, headgroup S_{CD} for PCs are nearly identical as C36 and in perfect agreement with experiment (Figure 12, upper left). Furthermore, because the parametrization also included the nonbonded parameters of the first two carbons in the acyl chain and dihedral parameters extending to the fourth carbon, the splitting of C2 from the *sn*-2 chain is well preserved.

When it comes to PE and PG, the experimental data is less accessible. The lower left panel of Figure 12 compares the S_{CD} of POPE bilayer simulated using C36/LJ-PME with several independent experiments. The agreement with experiment is good for the headgroup region. When interpreting the experimental data for the *sn*-1 chain, extra caution should be exercised because the experiment assumed a monotonic variation of S_{CD} toward the disordered center of the bilayer.[165] However, the overall agreement for the *sn*-1 chain is decent, indicating that the structure of the hydrophobic core is well described by C36/LJ-PME, and that the overall structure of the bilayer is well captured. For the *sn*-2 chain, the drop of S_{CD} around the double bond (C9=C10) is evident for C36/LJ-PME and it agrees qualitatively with the experiment. S_{CD} for the POPG headgroup are also compared to experiment in Figure 12. There is no reported error from Borle et al.,[166] but the comparison with Ferreira's data indicates that further refinement of PG-specific parameters might be needed, and it will be a focus of future updates to the FF.

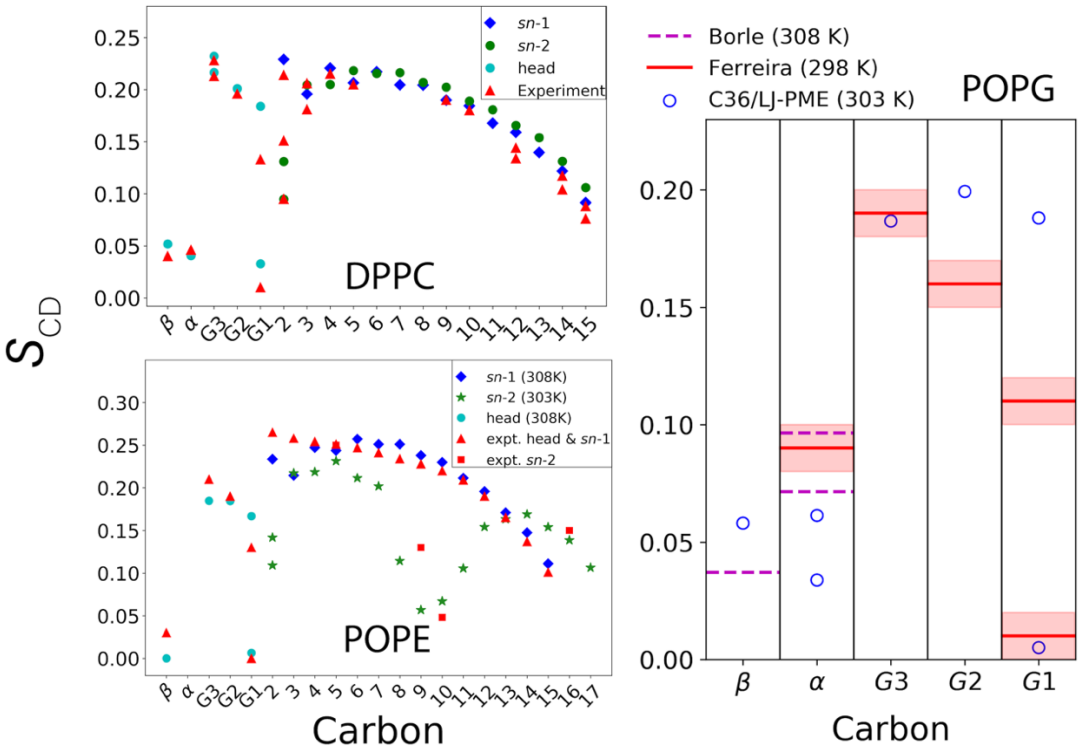

Figure 12. (Upper left) NMR deuterium order parameters (S_{CD}) for DPPC bilayer from experiments[132, 134, 167] and C36/LJ-PME at 323 K; (lower left) S_{CD} for POPE bilayer. Experimental temperatures are 308 K for the *sn*-1 chain,[165] 303 K for the *sn*-2 chain[168] and 310 K for the head group (Unpublished data from the NMRLipid Project[169]). Simulation temperatures shown in legend. (Right) S_{CD} for POPG head group, experimental data from Borle et al.[166] and the NMRLipid project.[169]

K_A calculated based on area fluctuation are listed in Table 14. Considering the standard errors, K_A from C36/LJ-PME simulations are highly consistent with C36 values, asserting that the re-parametrization only introduced minimal changes to the C36 lipid FF. K_A also compare favorably with available experimental values for PC and PE lipids, demonstrating the two headgroups were well parametrized. While a direct comparison with experiment is not available for DHPC, K_A for DHPC is close to

DPPC, which is reasonable considering the two molecules are very similar to each other in chemical structure.

Table 14. Area compressibility moduli (K_A) in dyn/cm for selected bilayers compared to C36 simulations using same system sizes and experimental values. Standard errors are given in parentheses.

lipid	temperature (K)	C36/LJ-PME	C36	experiment
DPPC	323.15	240 (20)	230 (20)[108]	231[138]
DMPC	303.15	210 (20)	210 (30)[108]	234[170]
DLPC	303.15	260 (20)		
DOPC	303.15	300 (40)	280 (10)[108]	300[171]
POPC	303.15	230 (30)	240 (10)[108]	180-330[172]
POPE	308.15	260 (40)	280 (20)[108,a]	233[173]
DMPG	303.15	200 (20)		
POPG	303.15	250 (20)	220 (20)[108]	
DHPC	321.15	240 (30)	214 (18)[67]	
	333.15	240 (30)	230 (15)[67]	

[a]Data obtained at 310 K.

2.3.7. Dynamic properties

The last part of the validation focuses on dynamic properties of lipid bilayers. The NMR spin-lattice relaxation times of C-H bonds are shown in Figure 13 for DMPC. Overall, C36/LJ-PME agrees well with the experiment at all three carbon Larmor frequencies, except carbon G2 (C2) under high frequency (150.84 MHz). The same trend has been detected for C36 in previous publications[62, 134] and will be further discussed in Section 2.4. In fact, the overall pattern for the headgroup T_1 from C36/LJ-PME is very similar to C36, despite the small differences for the α and β carbons. Again, this similarity is related to the dihedral fitting protocol maximizing the

consistency with C36. It should be noted that dihedrals in the glycerol region are highly coupled and direct fits to QM data are impractical. An alternative approach was taken by Klauda et al. in the original C36 FF,[19] where the order parameters of that region were used as additional targets to parametrize the dihedrals. This can potentially lead to unphysical description of the energy landscape in that region since the dihedral parameters that can generate the experimental order parameters might not be unique. In terms of the tail region, better agreement with experiment is seen for carbons closer to the headgroup, while the last three carbons (C12, C13 and C14) are underestimated at high frequency (150.84 MHz).

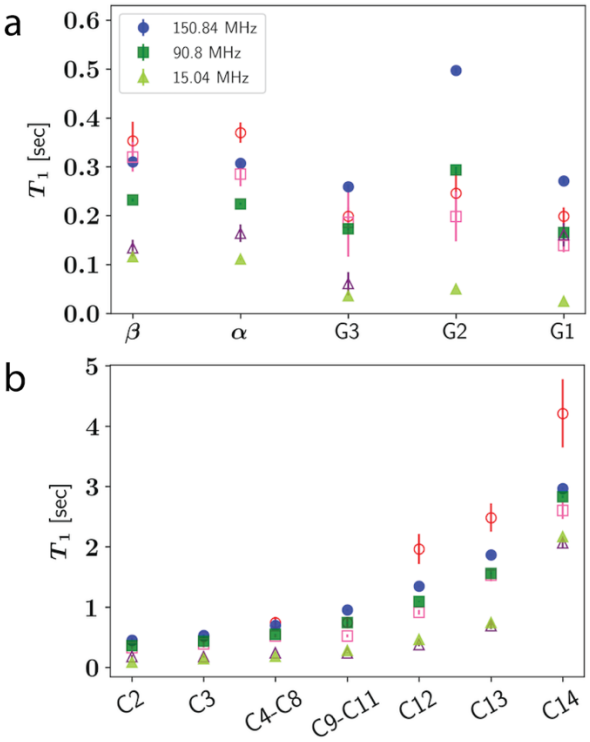

Figure 13. T_1 for DMPC at 303.15 K from experiment [174] (open symbols) and simulation (closed symbols). (a) Headgroup carbons; (b) Tail carbons.

To further study the frequency dependence for the tail region, T_1 of DPPC tail carbons were computed and compared to both experiments and C36 (Figure 14). There are two sets of experiments. One is from Brown et al.[175] measured with small vesicles, the other is from Klauda et al.[134] measured with large multilamellar liposomes (multilayers) at relatively higher frequencies. It should be noted that experimental $1/NT_1$ from multilamellar liposomes are, on average, 13% higher than those from small vesicles at 75.4 MHz, and 28% higher at 125.7 MHz. Possible causes are the waters remaining between the bilayers in the multilamellar liposomes and the different curvatures as discussed by Klauda et al.[134] In general, $1/NT_1$ for C36/LJ-PME agree well with the experiments, though slightly higher values are observed for <C4-C13> at lower frequencies compared to the vesicle data. There is also a trend that $1/NT_1$ for C36/LJ-PME at lower frequencies are higher than C36. These trends can be understood following Szabo's analysis of the frequency dependence of vesicle data[176] using his a "model-free" formalism.[149] Specifically, the spectral density for the i^{th} carbon in a chain, $J_i(\omega)$ is written

$$J_i(\omega) = (1 - A_i^2)\tau_i + \frac{A_i^2 \tau_s}{1+(\tau_s \omega)^2} \qquad (12)$$

where τ_i is the fast relaxation time associated with gauche-trans isomerization, A_i^2 is the generalized order parameter for each carbon, and τ_s is a slow relaxation common to all the acyl chain carbons. Since both the dihedral parameters (which are related to the fast motion) and the order parameter profiles for the acyl chain are identical for C36/LJ-PME and C36, the difference must arise from τ_s which here is the

slow relaxation associated with of wobble.[177] To verify this, the second rank reorientational correlation function $C_2(t)$ of the unit vector $\hat{\boldsymbol{\mu}}$ formed by C22 and C32 carbons (second carbons from the two acyl chains) were computed. $C_2(t)$ is defined as

$$C_2(t) = \langle P_2[\hat{\boldsymbol{\mu}}(0) \cdot \hat{\boldsymbol{\mu}}(t)] \rangle \tag{13}$$

where P_2 is the second Legendre polynomial. This correlation function is then fit with two exponentials in the form of Eq. 14 to extract the time constants corresponding to different modes of motion, and the slow relaxation time is assigned to wobble.[178]

$$C_2(t) = a_0 + \sum_{i=1}^{2} a_i e^{-\frac{t}{\tau_i}} \tag{14}$$

Fitting to Eq. 14 gives a slow relaxation time of 2.79 ± 0.12 ns for C36 and 4.15 ± 0.2 ns for C36/LJ-PME, which supports the idea that the larger $1/NT_1$ of C36/LJ-PME at lower frequencies are caused by slower wobble. In fact, the second term on the right-hand side of Eq 12 is a monotonically increasing function of τ_s within the frequency range studied here; besides, the lager τ_s of C36/LJ-PME makes the frequency dependence sharper than C36. Similar analysis can be conducted to explain the difference between C36 and C36/LJ-PME for DMPC T_1. Investigations into the chain order parameters show that C36/LJ-PME is less ordered near the lipid head compared to C36 but more ordered for the terminal carbons. For carbons near the head,

the influence of lower order parameter might be canceled by the slower wobble, thus T_1 for this region are similar between the two FFs. However, the relatively higher order parameters for the terminal carbons impose the same effect as the slower wobble, causing larger $J_i(\omega)$ and hence shorter T_1 for that region.

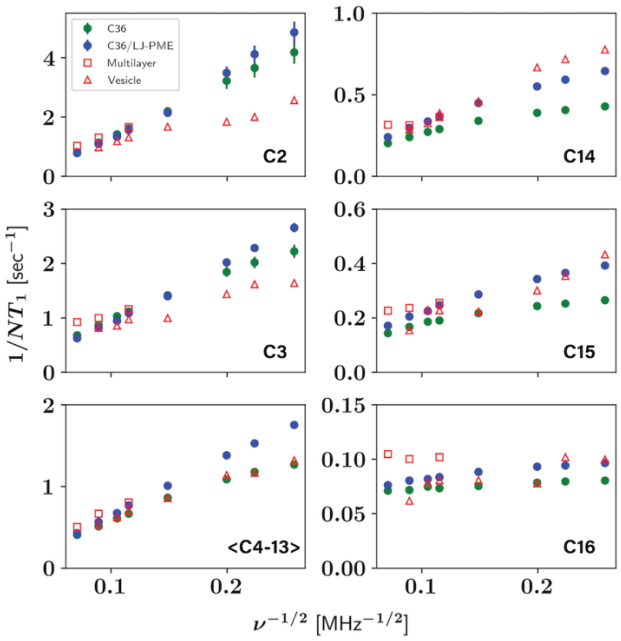

Figure 14. Frequency dependence of NMR T_1 for DPPC tail carbons. Vesical data from Brown et al.,[175] and multilayer data from Klauda et al.[134]

Another dynamic property studied is the diffusion constant of single lipids in a homogenous bilayer. The self-diffusivity of lipid has been a popular target for FF validation. However, obtaining meaningful diffusion constants from simulation to compare with experiments is not straightforward due to the strong effect of PBC on translational diffusion. Instead, extrapolation using knowledge of membrane shear viscosity (η_m) and interleaflet friction (b) is needed to get the diffusion constants to infinite system size. Based on the continuum hydrodynamic theory of Saffman-

Delbrück, Camley et al.[151, 152] developed the periodic Saffman-Delbrück (PSD) model to predict the diffusion constant of single leaflet spanning or monotonic cylinder in a membrane, which states

$$D^{\mathrm{PBC}} = \frac{k_{\mathrm{B}}T}{2L^2}\sum_{k\neq 0}\frac{A(k)}{A(k)^2 - B(k)^2}e^{-k^2\beta^2R^2/2} \tag{15a}$$

$$D^{\infty} = \frac{k_{\mathrm{B}}T}{2}\int\frac{d^2k}{(2\pi)^2}\frac{A(k)}{A(k)^2 - B(k)^2}e^{-k^2\beta^2R^2/2} \tag{15b}$$

$$A(k) = \eta_{\mathrm{mono}}k^2 + \eta_{\mathrm{f}}k\coth(2Hk) + b = \eta_{\mathrm{mono}}k^2 + \eta_{\mathrm{f}}k + b; \quad H\to\infty$$

$$B(k) = b + \eta_{\mathrm{f}}k\,\mathrm{csch}(2Hk) = b; \quad H\to\infty$$

where D^{PBC} is the diffusion constant for a periodic simulation box which measures L in *x*/*y*, and D^{∞} is the diffusion constant for infinite system size, β is a constant, R is the (effective) radius of the cylinder, H is the average height of water layers above and below the membrane, $\eta_{\mathrm{mono}} = \eta_{\mathrm{m}}/2$ is the monolayer surface viscosity, η_{f} is the bulk viscosity of the surrounding fluid (water), and b is the inter-leaflet friction. D^{∞} can be computed for a specific set of R, η_{m}, and b (the other parameters can be precisely determined from the simulation) using https://diffusion.lobos.nih.gov. Using a R of 0.45 nm for DPPC and fixing b at 1×10^7 P/cm,[179] Venable et al.[152] found that the best η_{m} to match the three C36 systems of different sizes is 4.4×10^{-8} P·cm. In a recent study of Zgorski et al.,[153] the surface

viscosity η_m for DPPC is calculated to be $(12.26 \pm 0.50) \times 10^{-8}$ P·cm, which is significantly larger than the fitted value by Venable et al. By fixing η_m at this value, Zgorski et al. found the best *R* to match the simulation results is 0.15 nm, a number much smaller than the radius estimated from the average area occupied by a single lipid in bilayer (0.45 nm as mentioned above). One explanation for this inconsistency is that an effective hydrodynamic radius might be smaller than one consistent with the surface area since a lipid is not a solid cylinder.

Table 15. Diffusion constants from C36/LJ-PME simulations (D^{sim}) and fittings using the PSD model (D^{PBC} and D^{∞}) for DPPC and DOPC.

lipid, temperature (K)	*R* (nm)	η_m (10^{-8} P·cm)	# lipid	D^{sim} (10^{-7} cm^2/s)	D^{PBC} (10^{-7} cm^2/s)	D^{∞} (10^{-7} cm^2/s)
DPPC, 323.15	0.45	5.7	72	1.12 ± 0.05	1.01	3.89
			288	1.45 ± 0.04	1.44	
			648	1.59 ± 0.07	1.70	
	0.15	11.3	72	1.12 ± 0.05	1.18	2.87
			288	1.45 ± 0.04	1.42	
			648	1.59 ± 0.07	1.55	
DOPC, 303.15	0.47	10.0	72	0.60 ± 0.04	0.58	2.24
			288	0.87 ± 0.03	0.82	
			648	0.89 ± 0.03	0.96	

Here, we fit η_m to the simulated results of DPPC using both *R*=0.45 nm and *R*=0.15 nm but those of DOPC only using *R*=0.47 nm, a radius consistent with the surface area. Table 15 lists the diffusion constants of DPPC and DOPC calculated from C36/LJ-PME simulations and those from fitting. Using a *R* of 0.45 nm for DPPC, η_m is determined to be 5.7×10^{-8} P·cm, which is at least 30% larger than the value obtained by Venable et al.[152] The fitted η_m of DOPC is 10.0×10^{-8} P·cm, a value 12% larger than Venable et al. This trend of higher viscosity is expected because of the added long-

range dispersion in C36/LJ-PME and it is consistent with the slower diffusion. When changing R to 0.15 nm, η_m of DPPC becomes 11.3×10^{-8} P·cm, a value much larger than 5.7×10^{-8} P·cm but smaller than the surface viscosity calculated by Zgorski et al. using C36. This indicates that effective radius of DPPC in C36/LJ-PME might be different from C36, since a higher-than-C36 surface viscosity is expected. D^{∞} of DPPC from fitting are 3.89×10^{-7} cm^2/s for R=0.45 nm and 2.87×10^{-7} cm^2/s for R=0.15 nm, which are still substantially larger than the experiment (1.5×10^{-7} cm^2/s).[180] D^{∞} for DOPC is 2.24×10^{-7} cm^2/s, while the experimental value is 0.825×10^{-7} cm^2/s.[181]

2.4. Discussion

Long-range dispersion has been incorporated into the CHARMM36 lipid FF through the LJ-PME method, resulting in a new FF denoted C36/LJ-PME. C36/LJ-PME was validated against a wide range of structural and dynamic properties using various lipids. While the overall agreement with experiment is like C36, the explicit inclusion of the long-range dispersion eliminates the sensitivity of the FF to the truncation of the LJ. Proceeding beyond lipids, simulations[182] using the CHARMM36m protein parameter set show the presence of LJ-PME does not impact assorted properties of ubiquitin. This is consistent with the use of an isotropic long-range LJ correction in condensed phase simulations during optimization of the model compounds (e.g., N-methylacetamide, methanol, dimethyldisulfide) that are the foundation of the C36 force field. Hence, it is reasonable to add proteins to membrane simulations carried out with C36/LJ-PME, though some testing would be prudent. Similar testing is also

recommended for simulations of membranes and nucleic acids and other compounds with the CHARMM force fields. The TIP3P water model is used to keep the balance between solvent-solvent and solvent-solute interactions and the long-range dispersion would not impact this balance because the dominant interaction is the short-range hydrogen bonding.

Another significant improvement over C36 is that C36/LJ-PME achieves consistency between bilayer and monolayer isotherms because of the long-range dispersion and the re-parametrization targeting both bilayer and monolayer properties. Monolayer simulations with C36/LJ-PME can now be compared directly with experiment, thereby greatly expanding the kinds of lipid systems that can be examined. Specifically, monolayers are often stable at lipid concentrations, temperatures, pH and ionic strengths that bilayers are not. For example, the molar fraction of the highly charged lipid, phosphatidylinositol 4,5-bisphosphate (PIP_2), is typically less than 5% on average in cell membrane,[183] though concentrations in local clusters can be much higher.[184] In experiment, the limit for PIP2 incorporation into bilayer vesicles with PC lipids is approximately 20%.[185, 186] Cholesterol has a stabilizing effect on PIP2 vesicles, but this requires a minimum cholesterol fraction of 20 mol%.[186] In contrast, monolayers can be formed with 100% PIP_2.[55, 187, 188] Another example is phosphatidylethanolamine (PE), for which pure monolayers can be formed[189, 190] but bilayers only exist in a narrow range of hydration.[191] In fact, the large spontaneous curvature of PE promotes the inverse hexagonal (H_{II}) phase under more physiologically relevant conditions (in this case, full hydration).[191]

Since LJ-PME is based on the efficient particle-mesh Ewald (PME) algorithm, the simulation speed should be comparable to the electrostatic PME method when using a proper real-space cutoff. Using the 72-lipid DPPC bilayer as a benchmark, the speed on a single NVIDIA V100 GPU is 179 ns/day when using a real-space cutoff of 10 Å and a time step of 2 fs in OpenMM 7.4.1. This is 10 ns/day faster than the electrostatic PME combined with a force-switching function ranging from 8-12 Å for the LJ interactions (other settings remain the same). While reducing the real-space cutoff can further speed up the simulation, a cutoff smaller than 9 Å is not recommended because tests have shown that the surface area of bilayer starts to deviate from its optimized value when the real-space cutoff is smaller than 9 Å.

Glycerophospholipids (GPLs) and ether lipids were parametrized separately because there is no overlap between the parameters. PCs were used for the parametrization of GPLs, where only the glycerol and ester groups were changed with respect to C36. This topologically restricted parametrization maximizes the consistency with the rest of the CHARMM FF by keeping the phosphate unchanged. It also allows freedom for future optimization targeting head-specific groups like the serine group in phosphatidylserine (PS) and the glycerol group in PG. The ether linkage was parametrized by targeting the DHPC bilayer surface area and thickness. Both parametrizations only introduced minimal changes to the C36 parameters, demonstrating the high quality of the C36 lipid FF and the efficiency of the semi-automated optimization approach.

The validation set for GPLs consists of a total of nine lipid types covering PC, PG and PE heads and saturated/unsaturated chains. Surface areas for these nine lipids

agree very well with experiments, with the largest deviation being DLPC (+5.8%). Like the C36 lipid FF, there is an overhydration of bilayers compared to experiment, especially for PCs with relatively short chains. The overall thickness D_B, which is calculated based on the water distribution, deviates from experimentally fitted values by 4.8% on average. As discussed in the original publication of C36, such an overhydration facilitates the improved estimate in the surface area. As the only PE bilayer simulated in this study, POPE has excellent agreement with experimentally measured structural data. The test for PG is based on DMPG bilayer at three different temperatures (303.15 K, 323.15 K and 333.15 K) and POPG bilayer at 303.15 K. C36/LJ-PME can reproduce the experimental surface area and thickness, as well as the scattering data. However, concern is raised when comparing the DMPG (303.15 K) chain order parameter to experiment (Figure 15). The average over the 12 largest values from both chains is 0.24 for C36/LJ-PME, while the NMR experiment measures about 0.21.[192] This higher order parameter is consistent with the lower surface area compared to experiment (59.8 $Å^2$ versus 62.5 $Å^2$). In fact, the gel to fluid phase transition temperature of the DMPG bilayer is 297.15 K (close to the simulated temperature),[192] and C36/LJ-PME fails to accurately predict properties at 303.15 K where transient domain(s) containing extended hydrocarbon chains are formed (Figure 16). This indicates that further refinement of the PG head group is needed for a more accurate treatment of the phase transition temperature and lipid packing. This will be a focus of future updates to the C36/LJ-PME FF.

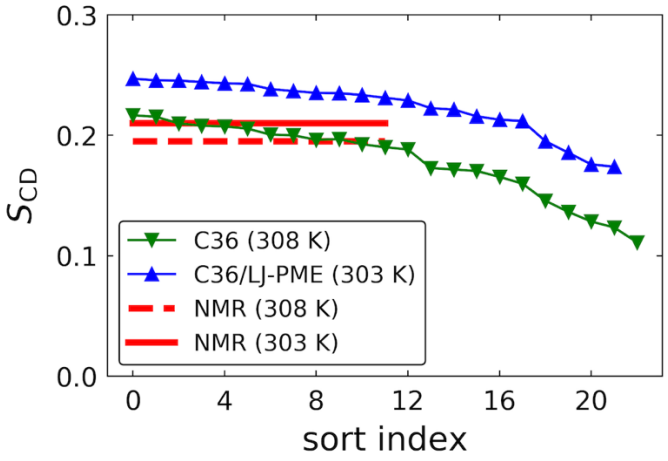

Figure 15. Sorted order parameters from both chains for DMPG bilayer. C36 data from Venable et al.[94] Experimental data from Loew et al.[192]

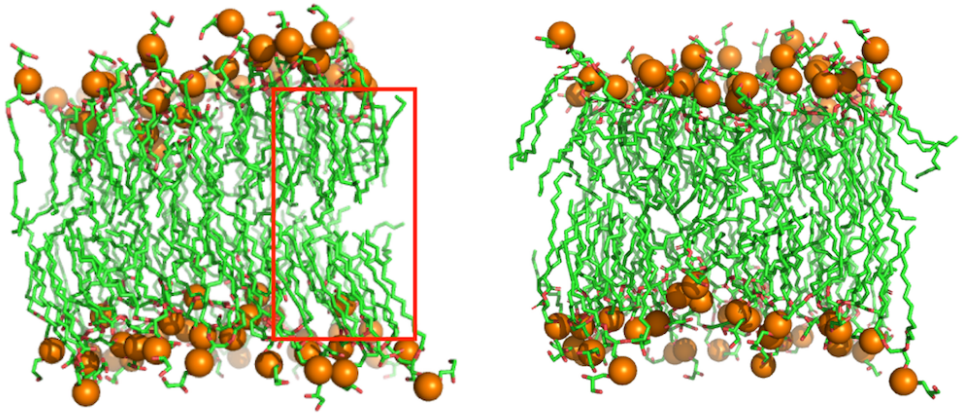

Figure 16. Snapshots from the DMPG bilayer (303.15 K) simulation. Left: snapshot at 120 ns; right: snapshot at 300 ns (end of simulation). Transient domain containing extended hydrocarbon chains is highlighted in red rectangular.

NMR spin lattice relaxation time T_1 of C-H bonds were compared to experiments. The head group region is well described by C36/LJ-PME. At high frequency (150.84 MHz), T_1 of carbon G2 is overestimated compared to the sonicated

vesicle data from Brown et al.[174] However, a more recent measurement[193] using multilamellar vesicles obtained by centrifuge is in favor of C36/LJ-PME and C36. At ^{13}C Larmor frequency of 125 MHz, Antila et al. detected a T_1 of 0.38 ± 0.04 s, a value right between the T_1 at 90.8 MHz and 150.84 MHz predicted by C36/LJ-PME. Considering the monotonic increasing of T_1 with respect to the Larmor frequency within the range studied here, this indicates that C36/LJ-PME predicts the T_1 of carbon G2 precisely. However, the same authors pointed out in another paper[194] that C36 underestimates the T_1 for the β and α segments due to the high weights of motion at the 0.1-1 ns timescale. An explanation for this could be the too fast diffusion of the TIP3P water considering these segments are close to the aqueous phase, though it is hard to quantify the influence. The frequency dependence of the chain is also well described by C36/LJ-PME as shown in Figures 13 and 14. The small difference with C36 can be explained by the slower wobble and rotational diffusion about the long axis of the lipid due to the increased viscosity at the presence of long-range dispersion.

Lipid diffusion calculated by C36/LJ-PME is slower than C36. Extrapolating the simulated results using the PSD model with an effective hydrodynamic radius of 0.15 nm generated a D^{∞} of 2.87×10^{-7} cm^2/s for DPPC, which is still 90% faster than experiment. The lower fluidity compared to C36 raises the concern that bilayers simulated with C36/LJ-PME will remain in the fluid phase above the transition temperature. To examine this problem, microsecond-long simulations were conducted for the 72-lipid DPPC bilayer at 323.15 K and 317.15 K (approximately 3 °C above the ripple (P_β) to liquid-crystalline (L_α) phase transition temperature).[195] The areas per

lipid for the two simulations are reported in Figure 17. The bilayer remained in the liquid-crystalline phase during the microsecond-long simulations.

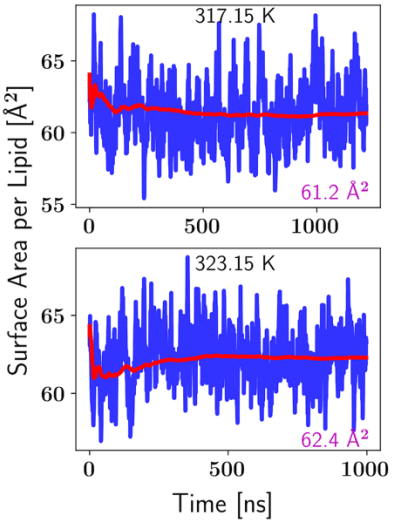

Figure 17. Surface area per lipid of DPPC bilayer at 317.15 K and 323.15 K during microsecond simulation using C36/LJ-PME. Average area after 200 ns shown on the bottom right.

While successful in many aspects, C36/LJ-PME has its limitations as an additive FF. For example, the electrostatic potential drop from the center of bilayer to water is about twice the experimental measures (Figure 18)[196] suggesting polarization is needed. Other situations a polarizable FF include water permeation, solvation free energy of apolar molecules in polar solutions or vice versa. In conclusion, C36/LJ-PME represents a major update to the C36 lipid FF by resolving the inconsistency between bilayer and monolayer and eliminating the "cutoff" dependence of the FF thereby increasing the range of systems that can be studied using the CHARMM FF.

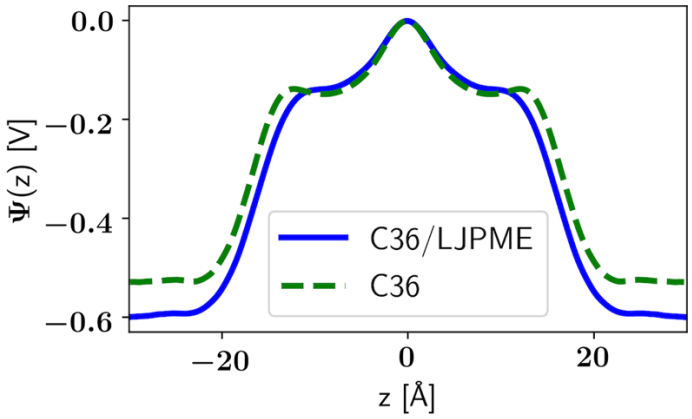

Figure 18. Electrostatic potential drop for DPPC bilayer at 323.15 K.

Chapter 3: Parameterization of the Drude polarizable lipid force field

3.1. Introduction

While fixed charge models like C36 and C36/LJ-PME account for induced polarization in a mean-field average way, they are still a poor approximation in regions where dielectric response is dominated by electronic polarization such as the hydrophobic core of a lipid bilayer. Taking the transfer free energies of water and ethane in a water/hexadecane system as an example, Venable et al.[46] have shown that C36 overestimates the water transfer free energy from water to hexadecane ($\Delta G_{\mathrm{w}\to\mathrm{h}}$) by 1.0 kcal/mol, while $\Delta G_{\mathrm{w}\to\mathrm{h}}$ for ethane is too negative by 0.4 kcal/mol. Electronic polarization is particularly important for membrane systems. First, as indicated by the water/hexadecane system, a fixed-charge model might not be able to precisely describe the transfer free energy of permeants from the aqueous phase to the membrane core, leading to related problems such as inaccurate prediction of permeability of drug-like molecules through a membrane. Second, the component breakdown to the membrane dipole potential shows significant difference between the polarizable Drude and nonpolarizable CHARMM models.[82] The induced polarization from the headgroups and water molecules generate a net residual electric field directed toward the water phase. The dipole potential from the hydrophobic core in is ~0.5 V in the polarizable model but is almost 0 V in the nonpolarizable model. This is directly associated with the overestimation of monolayer dipole potential using the nonpolarizable model.[82]

There are different methods to model induced dipoles as noted in Chapter 1. The Drude force field models electronic polarization based on the classical Drude

oscillator model. It is more computationally efficient compared to other existing methods and can be efficiently parallelized and implemented in high-performing MD simulation programs because of its particle-based structure and using an extended Lagrangian with a dual thermostat.[197-200] The FF can be used with a time step of 1 fs so that microsecond trajectories can be obtained.[53, 201]

The Drude lipid FF was first introduced by Chowdhary et al.[27] and a reparameterization for the headgroup dihedrals was attempted by Li et al.[28] to better match the experimental deuterium order parameters of the PC headgroup and to extend the FF to a wider range of lipids including PEs. Like the remainder of the Drude FF, this FF uses the SWM4-NDP water model[202] which has numerous advantages over the TIP3P model include but are not limited to viscosity, permanent dipole, and surface tension. It worth mentioning that both versions of the FF were validated using NAMD[203] with an analytic long-range correction to the energy and pressure from the LJ potential.[204] The latest update in 2017[28] used the unsigned deuterium order parameter as the primary fitting target for the headgroup dihedral parameters, which could have led to unrealistic dihedral distributions. Moreover, simulations performed by Li et al.[28] uniformly overestimate K_A for PC and PE lipids. A recent refinement of the phosphate nonbonded parameters was performed by Kognole et al.[205] and improved the water interaction energy surfaces of DMP when compared to QM calculations. Hydration free energies of molecular anions were also checked and an overall good agreement with experiments were seen. This refinement also unified the phosphate nonbonded parameters for lipids and nucleic acids. For these reasons, it would be beneficial to investigate the influence of these new parameters on lipid bilayers and

systematically parameterize lipid-specific nonbonded parameters and associated dihedrals if necessary.

This chapter introduces the newest results on the re-parameterization of the Drude lipid FF using a method modified from what was used for C36/LJ-PME. The parameterization also utilized the LJ-PME method for the long-range dispersions. While further refinement could be attempted, the current model sees a significant improvement over the 2017 version of the FF. K_A's generated by this model are in excellent agreement with experiments. Dihedral parameters were carefully adjusted based on multiple targets including the QM data, C36 potential of mean force (PMF), the unsigned S_{CD}, and the signed $^{13}C-^{1}H$ dipolar order parameters (S_{CH}). Validations based on lipids containing saturated chains of varying lengths also showed excellent agreement with experimental A_l, S_{CD}, monolayer isotherm, and monolayer dipole potential. Water permeabilities through lipid bilayers were also calculated and a huge improvement over the additive model was observed.

By way of outline, Section 3.2 describes the parameterization method, the training targets, and computational details. Section 3.3 delivers the results. Section 3.4 summarizes the work and discusses future directions.

3.2. Methods

In this parameterization, the FFLiP program described in Chapter 2 was modified to add support for the additional nonbonded categories in the Drude FF. The training set is limited to lipids with saturated tails because of a concurrent parameterization of the double bond parameters in the MacKerell group. This section begins with the training targets and associated membrane systems. The parameters to

optimize will be introduced next, followed by the modifications to the FFLiP program and the parameterization procedure. Computational details regarding simulations and property calculations are described last.

3.2.1. Training set

The training targets were selected to cover lipids with various saturated tails, namely DLPC, DMPC, DPPC, and 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC). Table 16 summarizes the systems studied in this chapter and their uses. The inclusion of multiple temperatures for each bilayer is meant for a better description of the temperature dependence of the FF.

Table 16. Systems simulated in this re-parameterization.

System	T (K)	Ensemble	N_{lipid}	$N_{water/lipid}$	Use
DSPC bilayer	333.15	NPT	72		optimization, validation
DPPC bilayer	323.15	NPT	72	30.4	optimization, validation
	333.15				optimization, validation
DPPC monolayer	321	NPAT (54 Å^2)	72	30.4	validation
		NPAT (64 Å^2)			validation
		NPAT (80 Å^2)			validation
DMPC bilayer	293.15	NPT	72	25.7	optimization, validation
	303.15				optimization, validation
	323.15				optimization, validation
DLPC bilayer	293.15	NPT	72	30.4	optimization, validation
	303.15				optimization, validation
	323.15				optimization, validation

Membrane properties used for the nonbonded parameterization are the surface areas per lipid of all bilayers listed in Table 16 and the order parameters of the DPPC headgroup. The experimental values can be found in Table 17. The surface area data is

from Kučerka et al.[131] When the past versions of the CHARMM/Drude lipid FFs were developed, the deuterium order parameters, S_{CD}, were used. According to Eq. 10, S_{CD} does not bear a sign. In fact, the sign can be determined by measuring the ^{13}C-^{1}H dipolar coupling in an NMR experiment,[206, 207] which gives the dipolar order parameter

$$S_{CH} = \frac{1}{2}\langle 3\cos^2\theta - 1\rangle \qquad (16)$$

where θ the angle between the C-H vector and the bilayer normal, the angular bracket denotes the time and ensemble average. We discovered that the S_{CH}'s for several carbons of the PC headgroup in the Drude 2017 lipid FF have opposite signs compared to experiments.[206, 207] In this work, signs of the order parameters were determined based on the S_{CH}'s reported in two publications,[206, 207] and values are based on the deuterium order parameters.[132, 134, 167] It worth mentioning order parameters are only used in the parameterization of the dihedral parameters so that scaling factors and weight factors (see Section 2.2.7 for definitions) are only needed for the A_l. Table 18 also lists the membrane-associated properties included in the validation of the FF.

Table 17. Membrane-associated targets, along with their scaling factors and weight factors.

property	system, temperature (K)	target value	scaling factor	Weight factor
A_l	DPPC bilayer, 323.15	65.0 (Å^2)[131]	60 Å^2	10
S_{CH}, C12		-0.04	N/A	
S_{CD}, C11		0.046		
S_{CD}, C1 (G3)		-0.213, -0.228		
S_{CD}, C2 (G2)		-0.196		
S_{CD}, C3 (G1)		-0.01, -0.133		
S_{CD}, C22		-0.095, -0.151		
S_{CD}, C32		-0.214		
S_{CD}, C23		-0.206		
S_{CD}, C33		-0.181		
A_l	DPPC bilayer, 333.15	63.1 (Å^2)[131]	60 Å^2	10
A_l	DSPC bilayer, 333.15	63.8 (Å^2)[131]	60 Å^2	10
A_l	DMPC bilayer, 303.15	59.9 (Å^2)[131]	60 Å^2	6
A_l	DMPC bilayer, 323.15	63.3 (Å^2)[131]	60 Å^2	6
A_l	DLPC bilayer, 293.15	59.6 (Å^2)[131]	60 Å^2	5
A_l	DLPC bilayer, 303.15	60.8 (Å^2)[131]	60 Å^2	5
A_l	DLPC bilayer, 323.15	64.8 (Å^2)[131]	60 Å^2	5

Table 18. Membrane-associated properties (validation only).

property	system, temperature (K)
X-ray/neutron form factors	DPPC bilayer, 323.15
Electrostatic potential profile	
Water permeability	
Area compressibility modulus	
X-ray/neutron form factors	DSPC bilayer, 333.15
Water permeability	
X-ray/neutron form factors	DMPC bilayer, 303.15
Area compressibility modulus	
Water permeability	
X-ray/neutron form factors	DLPC bilayer, 303.15
Water permeability	
Surface tension	DPPC monolayer (A_l = 54 Å^2), 321.15
Surface tension	DPPC monolayer (A_l = 64 Å^2), 321.15
Electrostatic potential profile	
Surface tension	DPPC monolayer (A_l = 80 Å^2), 321.15

3.2.2. Parameters to optimize

The 2017 version of the Drude lipid FF generates reasonable A_l for all the lipid bilayers tested, though overall slightly (~ 1 Å^2) lower than the experimental values. The parameterization of the C36/LJ-PME suggested that optimizing only the nonbonded parameters of the ester linkage region could give rise to a force field of same quality compared to optimizing the whole headgroup. Given these reasons, the parameterization of the Drude lipid FF is focused on the ester linkage region and the terminal carbons connected to the phosphate group. Figure 19 visualizes the atoms subject to the nonbonded parameterization. For consistency, CHARMM nomenclatures are used. In this parameterization, partial charges, atomistic polarizabilities, and Thole screening factors were allowed to change while the LJ parameters were fixed.

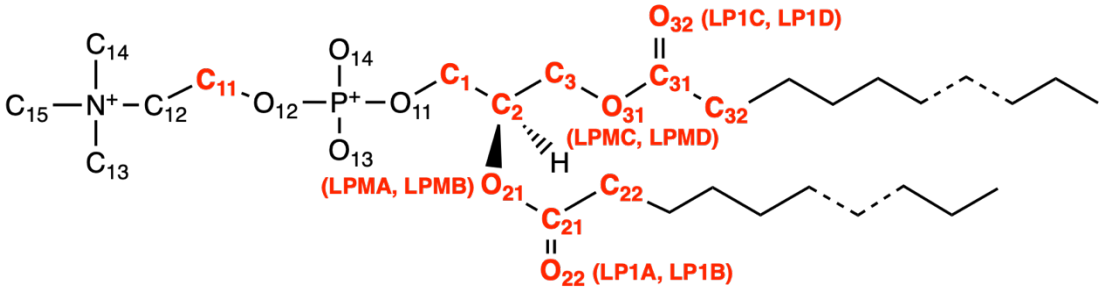

Figure 19. Chemical structure of PC lipids with saturated chains. Atoms subject to the nonbonded parameterization were labeled in red. Lone pairs of O21/O31/O22/O32 shown in parenthesis.

Dihedrals optimized in this work are listed in Table 19. The associated model compounds used for some of these dihedrals are plotted in Figure 20. There are three fitting protocols used for the dihedral parameters. The first is fitting to the QM PES of

model compounds. This starts with the smaller model compounds including DMP, ethyl-acetate (EAS), and methyl-pentanoate (MPEN). Larger model compounds including an esterified glycerol compound (GLYC) and a phospho-glycerol compound (GLYP) were used for dihedrals connecting the smaller model compounds. The second is to fit to the C36 PMF, which is performed using FFLiP. The last is reweighting to the experimental S_{CD}'s. These methods will be explained in Section 3.2.3.

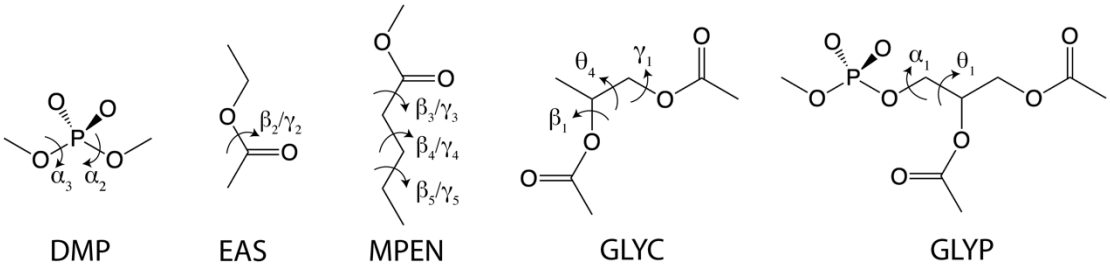

Figure 20. Model compounds used for the dihedral parameterization and the associated dihedrals.

Table 19. Dihedrals fitted and the corresponding fitting protocols. Model compounds (MCs) are shown in Figure 20. Symbols for the dihedrals are consistent with C36 and are used henceforth. In the "Fitting Protocol" column, "MC" stands for fitting to the QM PES of model compounds, "C36 PMF" means fitting to the C36 PMF from a DPPC bilayer simulation at 323.15 K. "S_{CD}" means reweighting to the experimental S_{CD}'s based a DPPC bilayer simulation at 323.15 K.

Dihedral	Symbol	Fitting Protocol	Model Compound
P-O11-C1-C2	α_1	MC, C36 PMF, S_{CD}	GLYP
O12-P-O11-C1	α_2	MC, S_{CD}	DMP
C11-O12-P-O11	α_3	MC, S_{CD}	DMP
C12-C11-O12-P	α_4	C36 PMF, S_{CD}	None
N-C12-C11-O12	α_5	C36 PMF, S_{CD}	None
O11-C1-C2-C3	θ_1	MC, C36 PMF, S_{CD}	GLYP
O21-C2-C3-O31	θ_4	MC, C36 PMF, S_{CD}	GLYC
C1-C2-O21-C21	β_1	MC, C36 PMF, S_{CD}	GLYC
C2-O21-C21-C22	β_2	MC, C36 PMF, S_{CD}	EAS
O21-C21-C22-C23	β_3	MC, C36 PMF, S_{CD}	MPEN
C21-C22-C23-C24	β_4	MC, C36 PMF, S_{CD}	MPEN
C22-C23-C24-C25	β_5	MC, C36 PMF, S_{CD}	MPEN
C2-C3-O31-C31	γ_1	MC, C36 PMF, S_{CD}	GLYC
C3-O31-C31-C32	γ_2	MC, C36 PMF, S_{CD}	EAS
O31-C31-C32-C33	γ_3	MC, C36 PMF, S_{CD}	MPEN
C31-C32-C33-C34	γ_4	MC, C36 PMF, S_{CD}	MPEN
C32-C33-C34-C35	γ_5	MC, C36 PMF, S_{CD}	MPEN

3.2.3. Parameterization procedure

The overall parameterization procedure is described in Figure 21. The parameterization starts with bilayer simulations using the 2017 parameter set with the updated phosphate nonbonded parameters by Kognole et al.[205] Upon completion of the simulations, the prediction for the best new parameter set is solved by minimizing

$$K_1 \cdot \boldsymbol{W}(\boldsymbol{S} \cdot \Delta \boldsymbol{P} - \boldsymbol{F})^2 + K_2 \cdot \sum_{\mathrm{MC}} \sum_{i} (U^{MM}_{\lambda+\Delta P} - U^{QM})^2 \qquad (17)$$

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_{prop} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{W}_{param} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} w_1 & & \\ & \ddots & \\ & & w_{N_{prop}} \end{pmatrix} & \boldsymbol{0} \\ \boldsymbol{0} & \begin{pmatrix} w'_1 & & \\ & \ddots & \\ & & w'_{N_{param}} \end{pmatrix} \end{pmatrix}$$

$$\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{prop} \\ \boldsymbol{S}_{param} \end{pmatrix} = \begin{pmatrix} \begin{pmatrix} s_{1,1} & \cdots & s_{1,N_{param}} \\ \vdots & \ddots & \vdots \\ s_{N_{prop},1} & \cdots & s_{N_{prop},N_{param}} \end{pmatrix} \\ \boldsymbol{I}_{N_{param}} \end{pmatrix}$$

$$\boldsymbol{F} = \begin{pmatrix} \boldsymbol{F}_{prop} \\ \boldsymbol{F}_{param} \end{pmatrix} = \begin{pmatrix} \boldsymbol{f}^{exp} - \langle \boldsymbol{f}^{sim} \rangle_{\boldsymbol{\lambda}} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $\boldsymbol{W}$ $\boldsymbol{S}$, $\Delta\boldsymbol{P}$ and $\boldsymbol{F}$ have the same definitions as Chapter 2. The second part of the objective function contains additional restraints from QM conformational energies for model compounds GLYC and GLYP. The two parts are balanced by factors K_1 and K_2. The FFLiP program was also modified to support parameter perturbations on the Drude

specific nonbonded categories, namely the Thole screening factor and the atomistic polarizability.

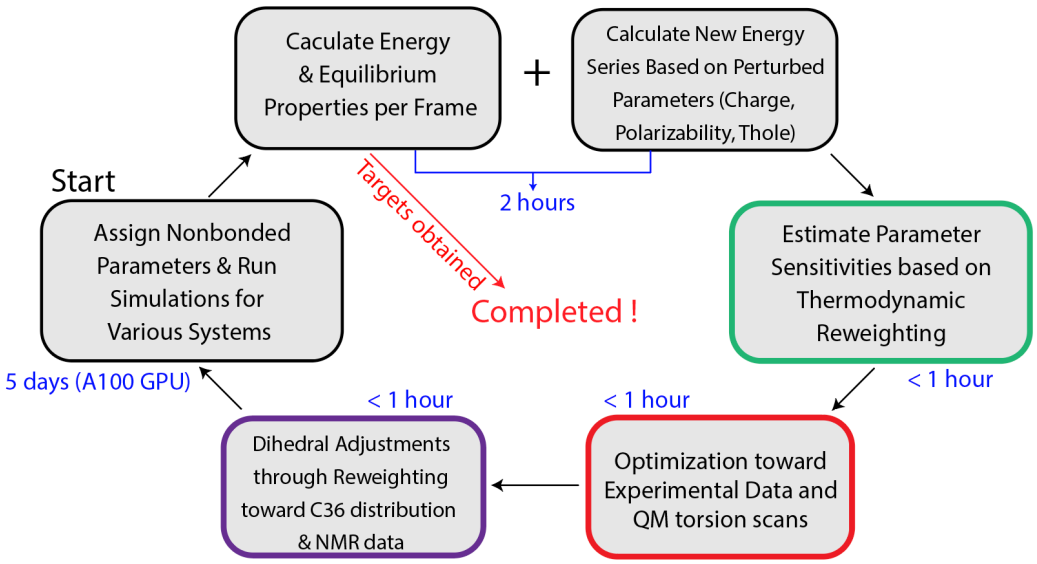

Figure 21. The parameterization procedure and approximate time per step.

Dihedral parameters were adjusted based on the predicted nonbonded parameters. The first step is to fit to QM PES of model compounds (Table 19) in the gas phase to generate a reasonable starting point. All QM scans were performed from -180° to 180° at an interval of 10°. For dihedrals fitted using EAS, MPEN, the remaining dihedrals were kept *trans*. When scanning α_2/α_3 using DMP, the other dihedral was placed at *trans* and *gauche* sequentially to include a wider range of conformers. The dihedral parameters were optimized to match the minimized conformer energies from QM. The PES generated by both QM and the force field were shifted such that the global minimum energy configuration had zero energy. Boltzmann weights based on the QM PES were used to weight different conformers. For larger

model compounds GLYC and GLYP, single molecular (gas phase) simulations were performed using the old FF to identify the most populated conformers, which were then minimized in PSI4[208] to determine the lowest energy conformers. For GLYC, 27 conformers were selected, among which the lowest 6 were used to perform the PES in PSI4 with the geometric plugin.[209] All the dihedrals except the one being scanned were fixed at the values of the minimized configuration. Dihedral parameters for β_1, θ_4, and γ_1 were fitted simultaneously using the 27 conformers and the PES based on the 6 lowest energy conformers. For GLYP, the same procedure was used, where 48 conformers were selected, among which the lowest 6 were used to perform the PES. Dihedral parameters for θ_1, and α_1 were fitted simultaneously. Boltzmann weights with an energy cutoff at 6 kcal/mol were used in the GLYC and GLYP fits.

Since these gas phase calculations do not perfectly represent the PMFs in the condensed phase, further adjustments toward the C36 PMFs are needed. This is based on DPPC bilayer simulations using both the C36 FF and the polarizable model with the updated dihedral parameters from the first step. The dihedral parameters were optimized to minimize $\sum_i (F_{C36} - F_{Drude} - V_0)^2$, where i runs over all the dihedral bins at a resolution of 1°. F_{C36} is the PMF from C36 and F_{Drude} is the PMF of the Drude model taking account the changes in the dihedral parameters. V_0 is a constant to ensure the minimum of PMF is at zero. The C36 FF was used because its dihedral parameters were already fine-tunned based on the experimental deuterium order parameters. After these steps, a 200 ns new DPPC bilayer simulation at 323.15 K was run with the newly fitted dihedrals, the last 150 ns of the trajectory was reweighted to reproduce the experimental S_{CH} by minimizing

$$K_1 \cdot \sum_{\text{carbon}} \left(\frac{\langle {S_{\text{CH}}}^{sim} \cdot e^{-\beta(U_B - U_A)} \rangle_A}{\langle e^{-\beta(U_B - U_A)} \rangle_A} - {S_{\text{CH}}}^{exp} \right)^2 + K_2$$

$$\cdot \left(\sum_i \sum_m \left({k^{i,m}}_A - {k^{i,m}}_B \right)^2 \right) \qquad (18)$$

This object function consists of two parts balanced by force constants K_1 and K_2. The first part measures the difference between the simulated and experimental order parameters. The second part are restraints with respect to the dihedral parameters from the two previous steps. ${S_{\text{CH}}}^{sim}$ is the order parameter from the simulation. ${S_{\text{CH}}}^{exp}$ is the experimental order parameter. U_A is the potential energy for each frame calculated using the simulated parameter set. U_B is the corresponding energy calculated with the new dihedral parameters. $\langle \cdot \rangle_A$ denotes the average over trajectory, which is also the ensemble average of parameter set $\boldsymbol{A}$. In the second part, the summation runs over all the optimized dihedral parameters (i) and all multiplicities (m). The BFGS algorithm in scipy[142] was used for this optimization. A new simulation was run with the optimized dihedral parameters and additional cycles of reweighting using the same method were performed until satisfactory agreement with the experimental order parameters.

A set of simulations using the bilayers listed in Table 17 were performed using the updated nonbonded and dihedral parameters. If needed, additional cycles could be performed according to Figure 21 until the FF is optimized.

3.2.4. Computational details

Initial structures of bilayers and monolayers were converted from the C36/LJ-PME end-of-simulation coordinates[210] using CHARMM-GUI[92] Drude prepper.[211] For DPPC monolayers, lipid headgroups from different leaflets were separated by a water slab, and the tails were separated by vacuum which measures about 180 Å in the monolayer normal direction. For all systems, the SWM4-NDP water model[202] was used. Bilayers were simulated in tetragonal boundary conditions in the NPT ensemble. Monolayers were simulated in the NPAT ensemble.

For the optimization, bilayers were simulated for 300 ns in OpenMM 7.7 using the Langevin Integrator with a time step of 1 fs. Friction coefficients were set to 5/ps for real atoms and 20/ps for Drude particles. Pressure was controlled by the Monte Carlo membrane barostat at 1 atmosphere. The real-space cutoff was set to be 10 Å for LJ-PME. The simulation method used for validation was the same as the optimization except an extra set of bilayers (DLPC, DMPC, and DPPC) used to study water permeation, for which the Nosé-Hoover Integrator was applied. The relaxation time of the thermostat was set to 0.1 ps for real atoms.

Monolayer initial structures were equilibrated and simulated for 150 ns using OpenMM 7.7 at desired surface areas (Table 16) with the Nosé-Hoover Integrator. The NVT ensemble was used. Pressure was controlled by the Monte Carlo membrane barostat at 1 atmosphere. 400-ps-long CHARMM[98] simulations were then started using the coordinates and velocities taken from the last 50 ns of the OpenMM simulations at an interval of 1 ns. The Nosé-Hoover thermostat was used to maintain system

temperature, and the relaxation time was specified to be 0.1 ps for real atoms. The last 200 ps of the CHARMM simulations were used to compute the average surface tension.

Properties calculated in this chapter include A_l, K_A, order parameters, membrane dipole potential, monolayer surface tension, and water permeability. Methods for calculating A_l, K_A, order parameters have been described in Chapter 2.

Monolayer surface tension was computed based on the pressure tensor reported in the CHARMM simulation as

$$\gamma = 0.5\langle L_z\left[P_{zz} - 0.5\left(P_{xx} + P_{yy}\right)\right]\rangle \qquad (19)$$

where L_z is the size of the simulation box normal to the alkane-vacuum interface.[212] P_{zz} is the normal component of the pressure tensor; P_{xx} and P_{yy} are the tangential components. The factor 0.5 accounts for the fact that there are two interfaces in the system. One-dimensional Membrane potential profiles were obtained as a solution to the Poisson equation from the average system charge density along the bilayer normal. The charge density in a certain snapshot was computed based on the positions of all the real atoms and Drude particles in the system and their partial charges, and a bin size of 0.1 Å was used. Water permeability through lipid bilayer was estimated using the flux-based counting method in Krämer et al.[213] Since this method relies heavily on the number of crossing events for statistical stability, three replicas of 300-ns-long simulation were used.

3.3. Results

In this section, results are presented for the best parameter set obtained, though further adjustments could be considered beyond this dissertation for final publication and release.

3.3.1. Nonbonded parameters.

One optimization cycle in Figure 21 was performed to obtain the final nonbonded parameters. An extra cycle of optimization was tried, and no improvement of membrane properties was observed. Table 20 compares the optimized nonbonded parameters to the ones at the beginning. Changes in partial atomic charge are all within 0.05 elementary charge due to the restraints to the old parameters. Changes in atomic polarizability and Thole screening factors were restricted to be less than 5% to ensure the fundamental physics were not deteriorated. The two carbons adjacent to the phosphate group, C11 and C1, were allowed to exchange partial charge with each other. This was found to be important to the PES of dihedral α_1 (see Section 3.3.2 for more details). The partial charges on O21 and O22 were kept as zero, and changes were only made to the associated lone pairs.

Table 20-a. Partial atomic charges (e) for optimized atoms and lone pairs in the PC headgroup.

atom name	Begin	Optimized
C11	0.228	0.268
C1	0.228	0.188
C2	0.202	0.156
O21	0.000	0.000
C21	0.697	0.651
O22	0.000	0.000
C22	-0.206	-0.238
LP1A	-0.349	-0.321
LP1B	-0.258	-0.230
LPMA	-0.170	-0.136
LPMB	-0.170	-0.136
C3	0.086	0.040
O31	0.000	0.000
C31	0.697	0.651
O32	0.000	0.000
C32	-0.206	-0.238
LP1C	-0.349	-0.321
LP1D	-0.258	-0.230
LPMC	-0.170	-0.136
LPMD	-0.170	-0.136

Table 20-b. Atomic polarizability for optimized atoms in the PC headgroup.

atom name	Begin	Optimized
C11	-1.642	-1.560
C1	-1.642	-1.724
C2	-1.797	-1.850
O21	-0.732	-0.761
C21	-1.370	-1.439
O22	-0.904	-0.859
C22	-1.993	-1.893
C3	-1.797	-1.850
O31	-0.732	-0.761
C31	-1.370	-1.439
O32	-0.904	-0.859
C32	-1.993	-1.839

Table 20-c. Atomic polarizability for optimized atoms in the PC headgroup.

atom name	Begin	Optimized
C11	0.862	0.819
C1	0.862	0.819
C2	0.410	0.431
O21	0.601	0.631
C21	1.747	1.660
O22	0.565	0.537
C22	0.410	0.389
C3	0.410	0.431
O31	0.601	0.631
C31	1.747	1.660
O32	0.565	0.537
C32	0.410	0.389

3.3.2. Dihedral parameters and S_{CD}

Dihedral parameters were fit through three steps. The first step, fitting to the QM PES, was found to be effective for smaller model compounds including EAS, DMP, and MPEN. Fitting to larger model compounds like GLYC and GLYP can be challenging because the energy surface has a much higher dimension, thus a multi-dimensional fit is required. Moreover, the nonbonded parameters and other dihedral parameters not being fit may not be in their optimal values, adding to the complexity of the fit. To increase the quality of the fit, a large set of rotamers were included and multiple starting conformations were used for the PES. This strategy was found to be effective for GLYC as shown in Figure 22, where PES for all conformers were in good agreement with the QM.

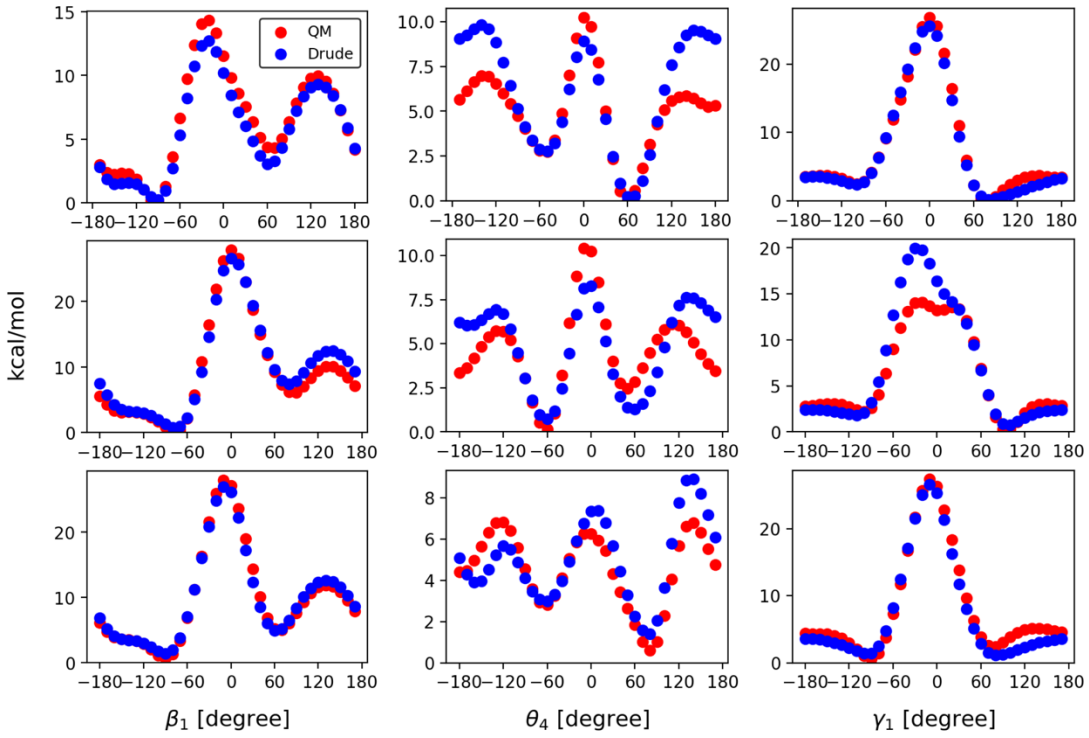

Figure 22. 1-D potential energy scan of dihedrals β_1, θ_4, and γ_1 in GLYC. Each row represents a rotamer. Energies are offset to ensure a zero global minimum.

When it comes to GLYP, however, the multi-dimensional fit is less satisfactory without using QM PES as additional restraints in the nonbonded parameterization (Figure 23), which demonstrates the essential role of these restraints. Adding these restraints led to an agreeable fit where all the local minima were evidently presented. The ultimate goal is to find a set of dihedral parameters that generate the right bond orientations, this was achieved by fitting to the PMF of the C36 lipid FF and further reweighting toward the experimental order parameters of the PC headgroup. Two iterations of reweighting were performed until a satisfactory S_{CD} profile was obtained (Figure 24). The optimized model is comparable to C36 and C36/LJ-PME, with improvements in the choline group but less forking of the C2 carbon of the *sn*-2 chain.

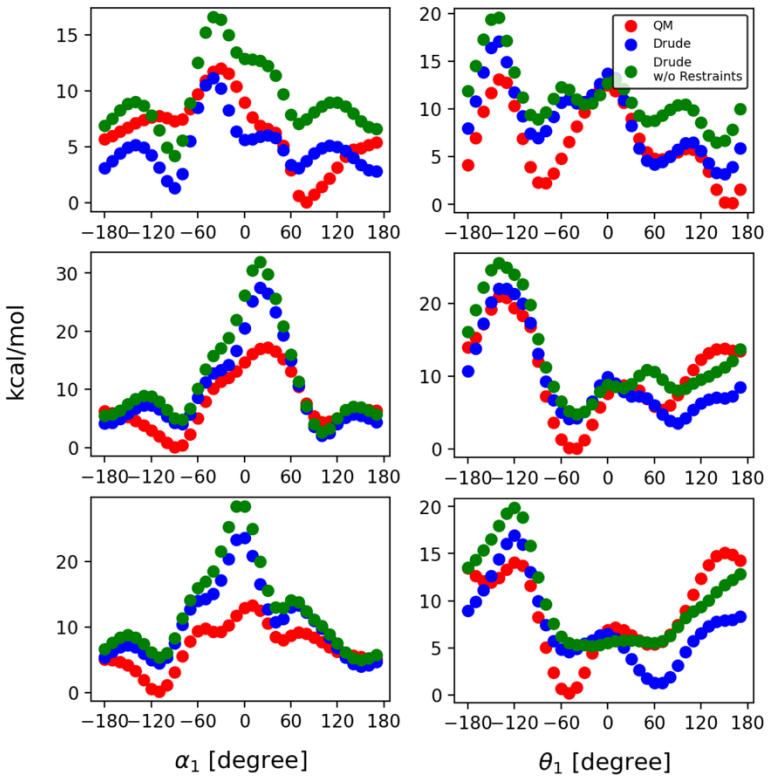

Figure 23. 1-D potential energy scan of dihedrals α_1 and θ_1, and γ_1 in GLYP. Each row represents a rotamer. Energies are offset to ensure a zero global minimum.

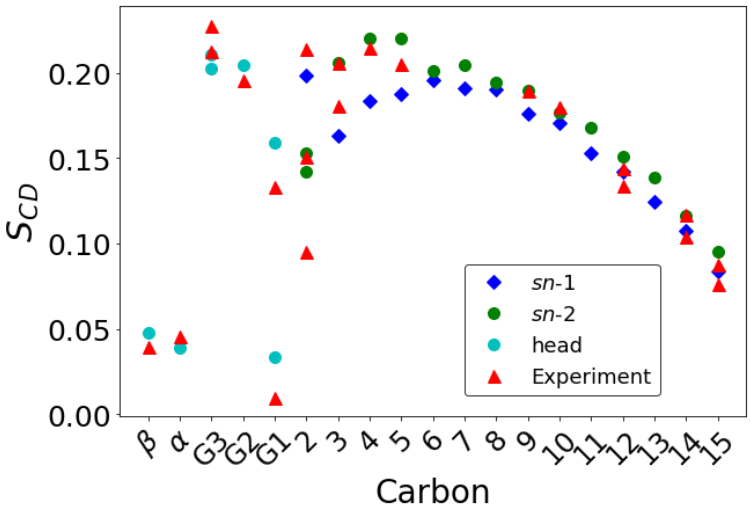

Figure 24. Deuterium order parameters for DPPC bilayer at 323.15 K.

3.3.3. Membrane properties

Surface areas per lipid are reported in Table 21. The overall quality is comparable to the 2017 version and the additive C36 FF. X-ray and neutron form factors calculated using the SIMtoEXP program[148] are shown in Figure 25 and agree well with experiments. A_l is slightly lower than the experiment for DPPC but higher for DMPC and DLPC. The chain-length dependence is more evident compared to the Drude 2017 FF, although more evidence is needed to support this (simulating the DSPC bilayer at 333.15 and other bilayers with the Drude 2017 FF at more temperatures would give more clues).

Table 21. A_l (Å^2/lipid) for bilayers. Available Drude 2017, C36, and experimental data included for comparison. Standard errors are given in parentheses.

Lipid	Temperature (K)	Drude	Drude 2017	C36	Experiment
DSPC	333.15	64.9 (0.5)			63.8 (1.3) [156]
DPPC	323.15	62.3 (0.2)	61.8 (0.5) [28]	62.9 (0.1) [108]	63.1 (1.3) [156]
	333.15	64.0 (0.2)		63.4 (0.4)	65.0 (1.3) [156]
DMPC	303.15	60.8 (0.2)	61.1 (1.9) [28]	61.5 (0.1) [108]	59.9 (1.2) [156]
	323.15	63.7 (0.3)		63.1 (0.3) [109]	63.3 (1.2) [156]
DLPC	293.15	61.0 (0.3)			
	303.15	63.1 (0.2)	61.6 (0.4) [28]	63.1 (0.3) [109]	60.8 (1.2) [156]
	323.15	65.9 (0.3)			

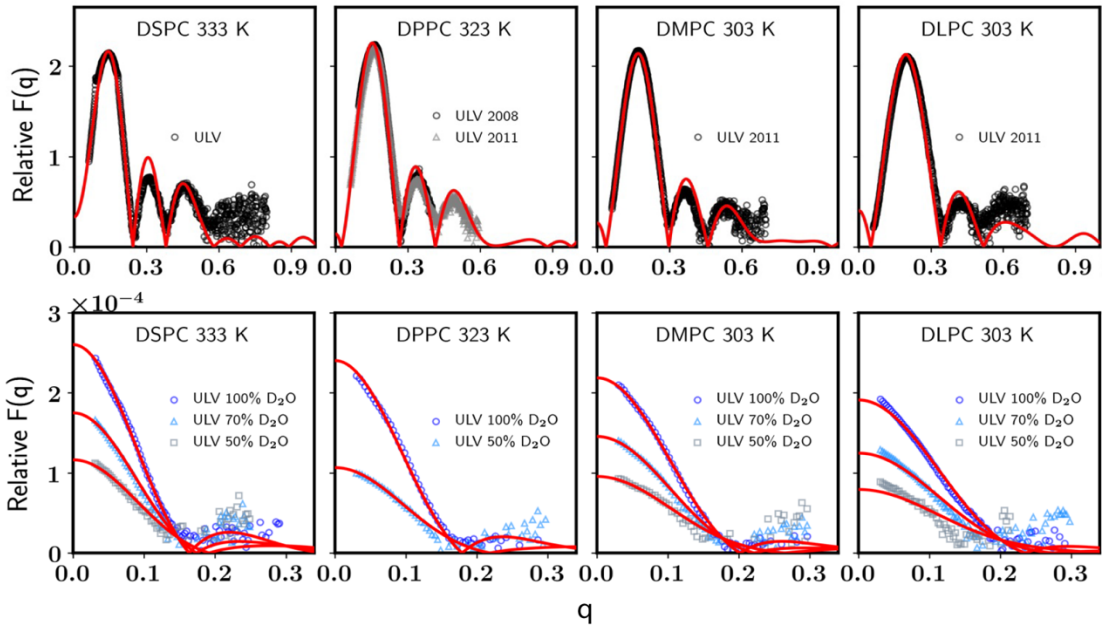

Figure 25. (Top) X-ray and (bottom) neutron scattering factors at selected temperatures.

The DPPC monolayer isotherm (surface tension vs. A_l) was also compared with experiment. As discussed in Chapter 2, the original C36 FF cannot reproduce this isotherm due to lack of the long-range dispersions and the C36/LJ-PME corrected this issue. Here, the Drude FF is parameterized with LJ-PME, so a good agreement with the experimental isotherm was expected. Figure 26 compares the three FFs, especially Drude and C36/LJ-PME. Both Drude and C36/LJ-PME accurately predict the monolayer isotherm. However, the DPPC monolayers simulated with the two FFs might be slightly too rigid based on the plot.

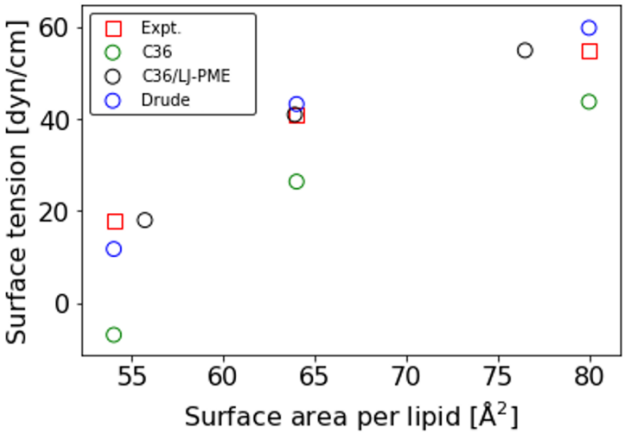

Figure 26. Isotherm for DPPC monolayer at 321 K.

K_A is an important mechanical property of membranes. The Drude 2017 FF systematically overestimates the K_A for bilayers. The updated phosphate nonbonded parameters[205] and the current re-parameterization give rise to a uniformly better estimation of K_A. Table 22 compares the current model (Drude) to the Drude 2017 FF and the C36 FF.

Table 22. K_A (dyn/cm) for bilayers. Available Drude 2017, C36, and experimental data included for comparison. Available standard errors are given in parentheses.

Lipid	Temperature (K)	Drude	Drude 2017	C36	Experiment
DSPC	333.15	210			
DPPC	323.15	230	393 (11)	230 (20)	231
	333.15	210			
DMPC	303.15	330	550 (65)	210 (30)	234
	323.15	300			
DLPC	293.15				
	303.15	270	288 (17)	260 [a]	
	323.15				

[a] C36/LJ-PME value from Yu et al.[210] is used.

Membrane electric potential profiles were calculated for a DPPC bilayer at 323.15 K simulated with the NPT ensemble and a DPPC monolayer at 321 K simulated with the NVT ensemble (A=64Å^2). The results are shown in Figure 27. In the bilayer, the profile has the same shape as the Drude 2017 model but has lower shoulders at the headgroup/water interface (0.65 V compared to 0.85 V in the 2017 model).[28] The dipole potential drop measured from the water phase to the hydrophobic core (0.4 V) is also lower than the 2017 model (0.55 V). The consensus experimental range for this quantity is 0.225-0.250 V,[196] although remain controversial. However, the experimental value for the DPPC monolayer dipole potential is unambiguous since it can be measured directly with electrodes and the shift in the interfacial potential before ($V_{\text{water-air}}$) and after ($V_{\text{mono-air}}$) spreading a PC monolayer on a water-air interface in experiment ranges between 0.3 and 0.4 V.[214] Here, $V_{\text{mono-air}}$ is about 0.85 V, and $V_{\text{water-air}}$ is around 0.5 V for the SWM4-NDP water model,[82] which means the shift is bracketed by the experiment values.

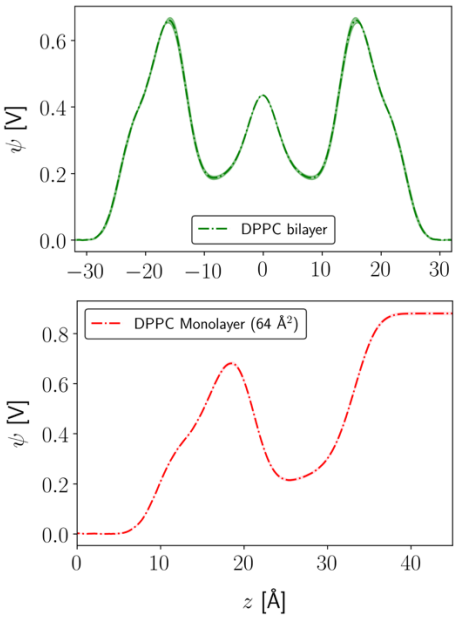

Figure 27. Membrane electric potential profiles for a DPPC bilayer at 323.15 K (NPT) and a DPPC monolayer at 321 K (NPAT with A=64 Å^2).

Water permeability (*P*) in lipid bilayers were computed for DLPC at 303.15 K, DMPC at 303.15 K and DPPC at 323.15 K (Table 23). The overall improvement over the additive C36 FF is impressive. DMPC is lower compared to the experiment by 42 % and DPPC is lower by 40 %. DLPC is almost identical to the experiment. Analysis of water PMF in the DPPC bilayer indicates a barrier of ~5.5 kcal/mol at the bilayer center with respect to the aqueous phase. This is at least1 kcal/mol lower compared to the C36 FF. The lower *P* compared to experiment indicates the diffusion of water within the bilayer, especially the hydrophobic core, might be too slow, though more evidence will be collected.

Table 23. Water permeabilities from counting for Drude and C36.[46] Standard errors given in parenthesis.

Lipid bilayer	Drude	C36	Expt.
DLPC (303.15 K)	9.3 (0.3)	3.7 (0.4)	10.4 ± 0.5 [215]
DMPC (303.15 K)	4.8 (0.2)	2.0 (0.4)	8.3 ± 0.76 [215]
DPPC (323.15 K)	16.3 (0.6)	6.7 (0.4)	27 ± 4 [137]

3.4. Discussion

In this chapter, the Drude FF has been re-parameterized with LJ-PME and a recent update to the phosphate nonbonded parameters. Membrane structures are accurately modeled by this FF. Bilayer compressibility moduli are greatly improved over the last version of the FF. The membrane dipole potential drop is also improved and is much better compared to the additive model, reassuring the usefulness of a polarizable model in modeling electrostatic properties. Water permeabilities through lipid bilayers are on average 2.5 times higher than those from the C36 FF, though still underestimate experiment.

The inclusion of the long-range dispersions using LJ-PME ensures a consistent modeling of bilayers and monolayers. Since the SWM4-NDP water model has an accurate surface tension (67 ± 4 dyn/cm at 298.15 K) compared to experiment (72.0 dyn/cm at 298.15 K),[202] a direct comparison to the experimental surface pressure of monolayer is now possible. The surface tension (γ) and the surface pressure (Π), is related through

$$\Pi = \gamma_0 - \gamma \qquad (20)$$

where γ_0 is the surface tension of a water-air interface. γ_0 at 321 K for the current Drude model was calculated to be 65 ± 1 dyn/cm using the CHARMM program with LJ-PME, leading to Π = 53, 22, and 5 dyn/cm at A_l = 54, 64, and 80 Å^2, respectively. The experimental values[162] are 50, 27, and 13 dyn/cm at these surface areas. Another experiment[216] at 323 K is consistent with these values. Although C36/LJ-PME tends to have the same monolayer surface tensions at these areas, it fails to give the right surface pressures due to the poor agreement with the experimental γ_0 (the water/air surface tension is 52.7 dyn/cm at 298.15 K for TIP3P water, the model used in C36/LJ-PME).

Order parameters were optimized using the dihedral parameters through reweighting in this work. Fitting to the experimental S_{CD} without checking the sign of S_{CH} is dangerous. In fact, re-investigation of the Drude 2017 FF indicates S_{CH} for several carbons are opposite compared to the experiment. To avoid this issue, the sign of S_{CH} was used in the fitting process. The initial fits to the QM conformational energies and the C36 PMF, and the restraints to these fitted values in the reweighting step ensured the final dihedral parameters are within a reasonable range.

This new Drude lipid FF uses the same phosphate nonbonded parameters as the nucleic acid component of the Drude FF, which is another step forward as this will give balanced interaction energies when simulating complex systems involving both lipids and nucleic acids. Using the 72-lipid DPPC bilayer (24481 atoms) as a benchmark, the speed on a single NVIDIA A100 GPU is 120 ns/day when using a real-space cutoff of 10 Å and a time step of 1 fs in OpenMM 7.7.

Chapter 4: Update to the CHARMM36 united atom chain model

This chapter was edited from the following publication:

> Yu, Y.; Klauda, J. B. Update of the CHARMM36 United Atom Chain Model for Hydrocarbons and Phospholipids. *Journal of Physical Chemistry B* **2020**.

4.1. Introduction

The parametrization of a lipid force field can be done at different levels other than all-atom. An UA lipid FF would lump nonpolar hydrogens into the attached heavy atom. Coarse gained FFs go one step further and group 3 to 4 heavy atoms into a single interaction site. The nonbonded parameters of the heavy atoms which the nonpolar hydrogens attach to are optimized to include the steric effect of hydrogens, making the system smaller in total number of particles, and hence improving the efficiency of the simulation. Polar hydrogens are explicitly modeled with their own nonbonded parameters since they may form hydrogen bonds and participate in other polar interactions. If parametrized well, UA FFs can be of similar accuracy to the AA ones.[217] Among currently available UA lipid FFs, GROMOS[20, 39] offers the greatest variety and is the most commonly used. It has several branches originated from different parameter sets but all use the GROMOS potential energy form. The most popular branch was initially developed by Chiu et al.[218] and now has been updated to include saturated phospholipids, monounsaturated phospholipids and sphingomyelins (SM).[20] This set has a good overall agreement with experimental A_l for all tested lipid bilayers and can be used to model mixed bilayers where the A_l for individual lipid type is critical.

However, the UA models from the GROMOS family are not compatible with CHARMM.

The CHARMM community has its own UA lipid FF. It uses the parameters from the CHARMM AA lipid FFs for the headgroup atoms and adopts the UA approach for the acyl chains.[29, 219] The early parametrization was done by Hénin el al.[219] and was based on the C27r lipids,[220] which does not allow simulations in the isothermal–isobaric (NPT) ensemble. A later update was attempted by Lee et al.[29] in 2013 where the C36[19] parameters were used for the headgroups, so that the FF was named the CHARMM36 United-Atom chain model (C36UA). The Lennard-Jones (LJ) parameters for the acyl chain were chosen by comparing several existing UA force fields on pure alkane systems, the OPLS-UA hydrocarbon parameter set[221] was selected because it gave the best overall agreement with experimental densities and heats of vaporization for heptane and pentadecane. This update reintroduced the hydrogens to the first methylene group of each lipid tail, which explicitly presented the known splitting of the order parameters of carbon-2 on the *sn*-2 chain.[132, 222] While the model was able to match the experimental structural parameters well for the four pure lipid bilayers tested, subsequent tests (unpublished work) demonstrated that the trend of A_l for saturated lipids with different chain lengths was imperfect and further optimization of the hydrocarbon parameters was needed.

In this chapter, we present the recent development of the C36 UA lipid FF. Most of the author's efforts were put on the parametrization of the methyl, methylene and methine groups, but parameters for polyunsaturated fatty acids (PUFA) were also developed. The new parameter set is denoted as C36UAr. C36UAr achieves great

accuracy and retains compatibility with the CHARMM FF family. In Section 4.2, we will first describe the methods used to optimize the parameters, to run and analyze the MD simulations of bulk hydrocarbons and lipid bilayers. Results will be presented in Section 4.3, starting with the dihedral fittings, followed by benchmarks on a variety of pure hydrocarbon properties. Results from pure lipid bilayer simulations will also be presented. For most of these results, we compare them with available experimental data as well as C36. Comparisons to the hydrogen mass repartitioning (HMR) method are also available for selected properties. In Section 4.4, we will discuss these results in terms of accuracy and circumstances where the model can or cannot be used.

4.2. Methods

An automated optimization procedure was used for the parametrization of the LJ potential parameters and associated dihedral parameters. Figure 28 provides an overview of the whole process. In the remaining of this section, details of the optimization process and the validation methods will be presented.

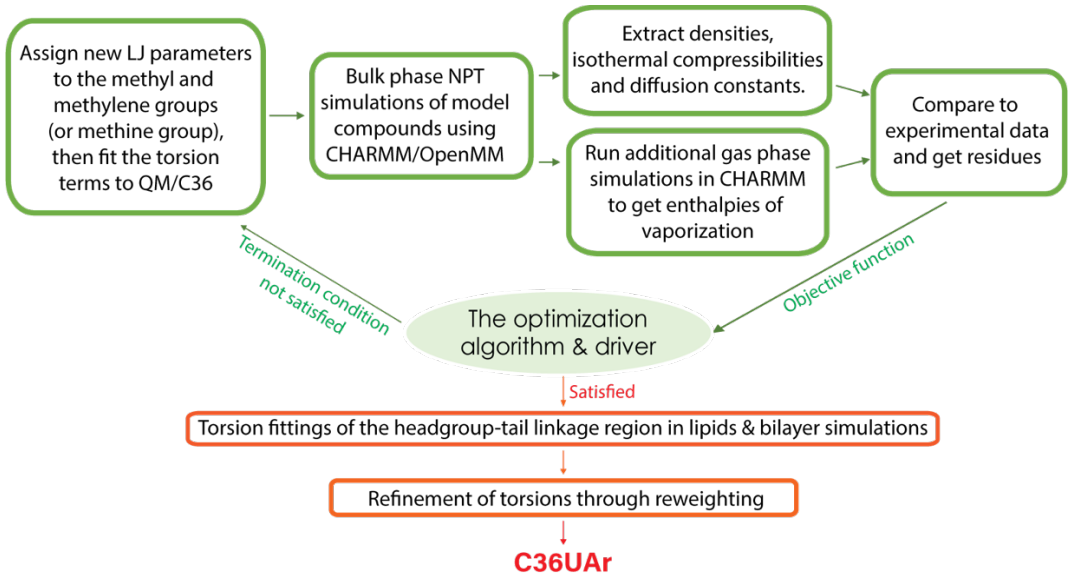

Figure 28. The optimization procedure to get C36UAr. Green blocks are hydrocarbon parametrization; red blocks are parametrization based on lipid bilayers.

4.2.1. Optimization procedure

The Sbplx algorithm[223] from the NLopt package[224] was used for the automated search in the parameter space. The original C36UA parameters were used as initial guesses. The loss function $F(\boldsymbol{p})$ used for the optimization can be written as

$$F(\boldsymbol{p}) = \sum_i W_i \cdot \left(\frac{f_i^{sim}(\boldsymbol{p}) - f_i^{exp}}{f_i^{exp}} \right)^2 \qquad (21)$$

where $\boldsymbol{p}$ is the parameter set in a particular optimization step. The summation is over all properties used as targets. W_i is the weight factor for a particular property, which can be found in Table 16. f_i^{exp} is the experimental value of the property and $f_i^{sim}(\boldsymbol{p})$ is the value calculated from simulation.

Table 24. Fitting targets for hydrocarbon parametrization and their weights.

Optimization Phase	Phase I - Alkanes					Phase II - Alkenes			
Molecule	*n*-heptane		*n*-hexadecane			*cis*-2-hexene		*cis*-5-undecene	
Temperature (K)	303.15	312.15	288	298	312.15	293	298	293	348
Density	1	1	1	3	3	1	1	1	0
Heat of vaporization	1	1	0	1	1	1	1	0	1
Compressibility	1	1	0	0	0	0	0	0	0
Self-diffusivity	1	1	0	0	0	0	0	0	0

As shown in Figure 28, the LJ parameters were fit to various bulk liquid properties of hydrocarbons. It should be noted that the methyl (named CH3E in C36UAr) and methylene (CH2E) groups were optimized first and the methine group (CH1E) was optimized with the optimal CH2E and CH3E parameters. To accurately model the temperature dependence of the properties and to model individual groups correctly, hydrocarbons of different lengths were simulated at different temperatures in each optimization step to obtain various physical properties. Table 24 lists all properties used as targets and their weight factors in Eq. 21; details of calculating these properties can be found in Section 4.2.2. LJ parameters of CH3E and CH2E were fit to experimental data simultaneously using *n*-heptane and *n*-hexadecane, while CH1E was fitted using *cis*-2-hexene and *cis*-5-undecene. Initial coordinates of systems were built from monomers using PACKMOL.[225] The *n*-heptane system contained 64 molecules; the 2-hexene system contained 128 molecules; the *n*-hexadecane system and the 5-undecene system contained 256 molecules. In each optimization step, energy of each system was first minimized using the adopted basis Newton Raphson method (500 steps) followed by the steepest descent method (500 steps) in CHARMM to avoid

unfavorable van der Waals contacts. After that, a 2-ns equilibration in OpenMM was conducted to relax the dihedrals. The final NPT simulation of 4 ns also run in OpenMM and was used to obtain the density, isothermal compressibility, and the diffusion constant. The same simulation was also used for calculating the heat of vaporization, for which a detailed explanation is given in Section 4.2.3. In both equilibration and production runs, long-range LJ interactions were included through LJ-PME with a 10 Å real space cutoff. Langevin Integrator with a 2-fs time step was used for the dynamics and Monte Carlo Barostat was used for pressure control at 1 bar. Friction coefficient for the Langevin Integrator was chosen carefully to be 0.02 ps^{-1} to reproduce the diffusion constant calculated using CHARMM (Figure 29). The inclusion of the long-range dispersion for hydrocarbons in this work was motivated by two facts. The first is that these interactions are crucial to the properties of pure alkanes and interfacial systems including alkanes as discussed in Chapter 2. Second, the all-atom C36 lipid FF produced closer-to-experiment properties for tested linear alkanes when including the long-range LJ interactions compared to excluding them.[101] To keep consistent with C36, the parametrization of hydrocarbons was done with LJ-PME, while the lipid bilayer simulations were performed with the same nonbonded method (PME) as C36.

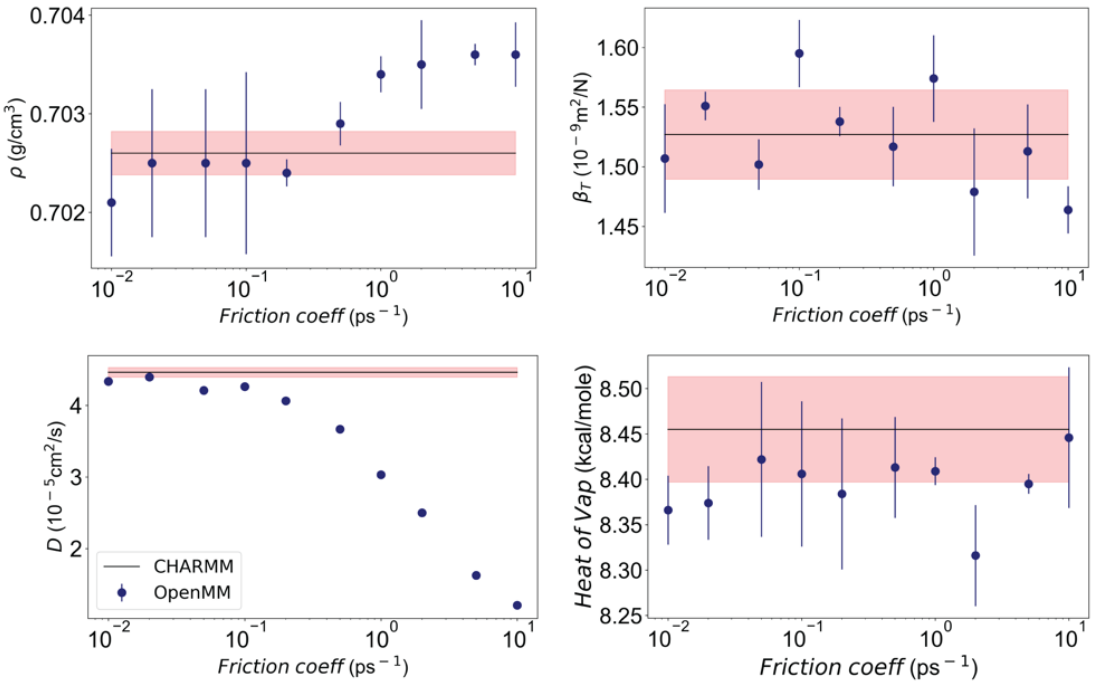

Figure 29. Comparisons between OpenMM and CHARMM for density (upper left), isothermal compressibility (upper right), diffusion constant (lower left) and heat of vaporization (lower right) using heptane at 312.15 K. Standard errors of CHARMM shown in red.

4.2.2. Model validation using hydrocarbons

For validation of the model, 4 replicas of 10-ns NPT simulations in OpenMM were conducted using the same set-up as the optimization except that the Nosé-Hoover chain velocity Verlet integrator[226] was used for the constant temperature control. The NPT simulations are used to compute the density, heat of vaporization (see Section 4.2.3 for details), isothermal compressibility and self-diffusivity. Besides, CHARMM LJ-PME simulations were conducted to get alkane/vacuum interface surface tensions and shear viscosities. In the 50-ns canonical (NVT) simulations to calculate surface tensions, box sizes in the z direction were kept at 120 Å, while sizes in the x and y directions were kept at 42 Å for systems containing 288 heptane molecules and 50 Å

for systems containing 256 hexadecane molecules. Alkane slabs were put at the center of the simulation box and the distances between image slabs were at least 70 Å. The last 30 ns of these simulations were used for the surface tension calculations. For viscosity calculations, cubic simulation boxes were used to run another set of 150 ns NVT simulations, where box sizes were calculated using the densities determined from the NPT simulations. The last 120 ns of these simulations were used for calculating the viscosities and standard errors were estimated using a block size of 40 ns. For all CHARMM simulations, the Nosé-Hoover thermostat was used to maintain constant temperature.

The surface tensions of the alkane-vacuum interfaces were calculated based on Eq. 19.

Shear viscosities η were evaluated from the Green-Kubo formula[204] as

$$\eta = \frac{V}{k_B T} \int_0^\infty \langle P_{\alpha\beta}(t) \cdot P_{\alpha\beta}(0) \rangle \, \mathrm{d}t \qquad (22)$$

where k_B is the Boltzmann's constant, T is the temperature, $P_{\alpha\beta}(t)$ is the off-diagonal elements of the pressure tensor at time t, and V is the volume of the simulation box.

Isothermal compressibilities were calculated from

$$\beta_T = \frac{\langle \delta V^2 \rangle}{\langle V \rangle k_B T} \qquad (23)$$

where $\langle V \rangle$ is the average volume of the simulation box, $\langle \delta V^2 \rangle$ is the fluctuation in that volume, k_B is the Boltzmann's constant, and T is the temperature.

The slope of the mean squared displacement (MSD) versus time was used to determine the diffusion constant D^{PBC} using a weighted least squared fit with weights obtained from averages of 8 subgroups of molecules. This number was then corrected by the following equation to account for the system-size effect brought by the periodic boundary conditions[150] to get the self-diffusivity D_s:

$$D_s = D^{PBC} + \frac{k_B T \xi}{6\pi\eta L} \tag{24}$$

where k_B is the Boltzmann's constant, T is the temperature, L is the average box length, η is the viscosity, and $\xi = 2.837297$. While the most common strategy to calculate self-diffusivity is to use a canonical (NVT) ensemble, we argue that the NPT ensemble would yield similar results for the systems studied. A benchmark was done on heptane at 312.15 K using OpenMM with and without the Monte Carlo Barostat, and good agreement between the two ensembles was observed (see published paper[30] for more details).

For heat of vaporization, extra 200-ps long gas phase simulations of single molecules were performed in CHARMM with nonbonded cutoff larger than the molecule size but smaller than half of the periodic box size. 100-ps equilibrations were used to relax the conformations got from the last frames in the liquid phase simulations. The heat of vaporization can be calculated from

$$\Delta H^{vap} = \langle U_g \rangle - \frac{\langle U_l \rangle}{N} + RT \tag{25}$$

where R is the gas constant, $\langle U_l \rangle$ is the average potential energy over time of the liquid state, N is the total number of molecules, and $\langle U_g \rangle$ is the average gas phase potential energy obtained from N gas phase simulations.

4.2.3. Dihedral parameter fitting

Torsional terms were fitted using different strategies for saturated hydrocarbon chains and unsaturated chains. For saturated chains, dihedral parameters were fitted to PMF calculations from a liquid phase simulation of hexadecane using the C36 FF. It should be mentioned that direct fits to QM were attempted at first but led to significant overestimates of A_l and underestimates of S_{CD} in lipid bilayer simulations (will be discussed in the Section 4.3). For monounsaturated and polyunsaturated chains, QM scans of the two-dimensional conformational energies for *cis*-2-hexene and 2,5-heptadiene were taken from previous works.[19, 77] For *cis*-2-hexene, the two consecutive dihedrals scanned were CH1E-CH1E-CH2E-CH2E and CH1E-CH2E-CH2E-CH3E. For *cis*-2,5-heptadiene, the scanned dihedrals were the two centered between the two double bonds (Figure 30). Since both nonbonded and dihedral parameters changed for the hydrocarbons, additional adjustments to the dihedral parameters of the headgroup-tail linkage regions in lipids were required. Consequently, dihedral parameters of the linkages were fitted to reproduce the C36 PES. For all dihedral fittings, a program from MacKerell's group[143] was used; it uses a simulated annealing method which allows for a fast and steady search in the parameter space.

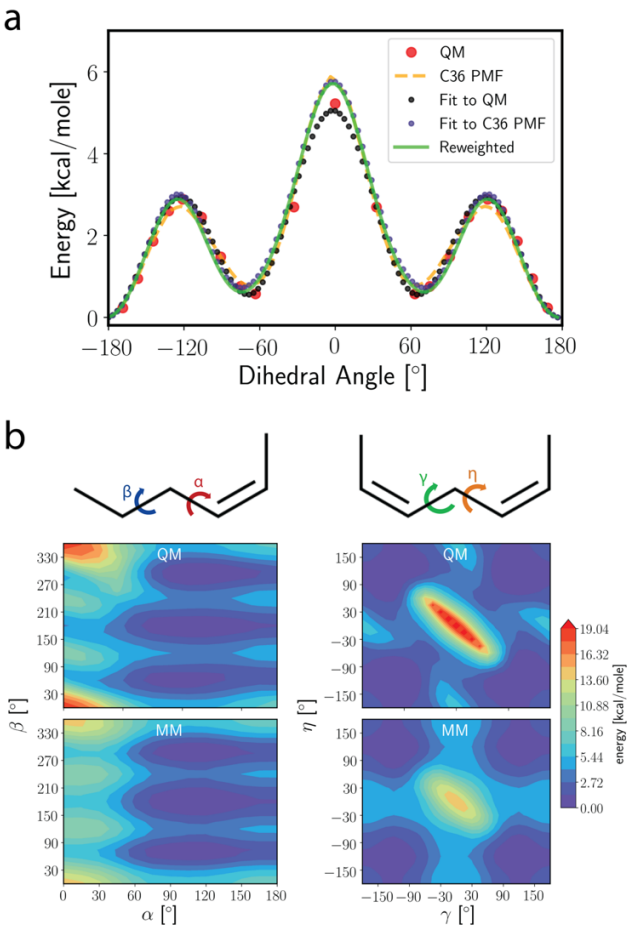

Figure 30. Dihedral fitting results. (a) potential energy scans for CH2E-CH2E-CH2E-CH2E; (b) Two-dimensional (2D) fitting of one double bond using *cis*-2-hexene and 2D fitting of two consecutive C=C double bond for PUFA lipids using *cis*-2,5-heptadiene. The chemical structures and fitted dihedrals were shown in the top panel.

4.2.4. Reweighting on the torsional parameters

After the torsional fits to C36 PMF for saturated chains, test simulations of 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) bilayers at different temperatures indicated that adjustments in the torsional parameters might still be needed to reproduce the experimental surface areas. Since only small changes were expected, reweighting was used (see Eq. 5). In each iteration, A_l, *gauche* and *trans* ratio of the C-C-C-C

dihedral of saturated tails were calculated. Energy series were computed using OpenMM and perturbations in the torsional force constants were set to 0.01 kcal/mol. The best parameter set was solved by assuming a linear relationship between the change in the force constant and the property since parameter changes were expected to be small, hence the sensitivity would remain almost constant. Quadratic restraints were set to avoid overfits and large deviations from the starting parameters. The new parameter set was subject to tests using the same set of target properties and further refinement until satisfactory results were obtained.

4.2.5. Bilayer simulations

In this study, model validations were performed on DLPC, DMPC, DPPC, DSPC, POPC, DOPC, 1,2-dilauroyl-sn-glycero-3-phosphoethanolamine (DLPE), POPE, POPG, 1-stearoyl-2-docosahexaenoyl-sn-glycero-3-phosphocholine (SDPC) and 1,2-diarachidonyl-phosphatidylcholine (DAPC). The modified TIP3P water model was used to keep consistent with the CHARMM family. Simulations of lipid bilayers were performed using NAMD[227] and OpenMM. 100 to 200 ns short simulations were run in NAMD to get equilibrium properties including A_l, S_{CD} and EDPs, while 1 μs OpenMM simulations were used to study diffusion processes in bilayers. For both simulation tools, a force-based switching function was used to switch the van der Waals forces to zero over a range of 10-12 Å. Long range (>12 Å) electrostatic interactions were treated with the PME method. All the bilayer simulations were conducted under the NPT ensemble with the pressure set to be 1 atmosphere and with the semi-isotropic cell, where x was constrained to be equal to y but allowed to vary independently with respect to z. In NAMD, Langevin dynamics was used to maintain constant

temperatures, while the Nosé−Hoover Langevin-piston algorithm[228, 229] was used to maintain the constant pressure. In OpenMM, the Nosé-Hoover chain velocity Verlet integrator was used for the constant temperature control and the Monte Carlo Membrane Barostat was used for the constant pressure control. 72 lipids were used for all the NAMD simulations, as this is large enough to get the equilibrium properties studied in this work.[154, 230] However, the diffusion of individual lipids in bilayer has a strong dependence on the size of the simulation box[152] and large artifacts can be introduced by the periodic boundary conditions (PBC) when simulating with small number of lipids, so that DPPC and DOPC bilayers with both 72 and 288 lipids were built and simulated using OpenMM to extract the diffusion constant of individual lipid.

Studies[231, 232] have shown that hydrogen mass repartitioning (HMR) can be a decent way to accelerate MD simulations and it is expected to produce the correct equilibrium thermodynamic averages of observables. Here, we are particularly curious about how it compares to C36UAr. DMPC, DPPC, DOPC, DLPE and POPE bilayers containing C36HMR lipids were built and simulated, where the mass of hydrogens was scaled by a factor of 3 and the extra mass was subtracted from the bonded heavy atoms. The simulations were used to study A_l, K_A, phase transition, hydrogen bonding and dynamic properties of single lipid bilayers. A C36 DOPC bilayer (72 lipids) was also simulated to study water permeation. Details about these comparisons can be found in Section 4.3. Initial structures of the 72-lipid bilayers (C36HMR and C36 systems) were generated by CHARMM-GUI and then equilibrated in OpenMM. For all UA systems, corresponding AA systems were built in CHARMM-GUI and equilibrated following the standard six-step equilibrium strategy provided by CHARMM-GUI before

converting to the UA topology using CHARMM. In all cases, the coordinates of the 72-lipid systems after 100 ns of simulations were used to build systems containing 288 lipids.

4.2.6. Calculations of bilayer properties

Methods to calculate A_l, K_A, S_{CD}, membrane thicknesses, and lipid self-diffusion were the same as those used in Chapter 2 and will not be repeated. Standard errors of lipid-diffusion were estimated based on 3 replicas for each system, and each replica contained 4 blocks of 200 ns trajectory.

4.3. Results

4.3.1. Dihedral fitting

As mentioned in the previous section, there were three sets of dihedrals to fit in this study. The fitting results can be found in Figure 30. For the saturated dihedrals, fits to QM were satisfactory for the hydrocarbon properties, but led to higher A_l and lower S_{CD} compared to experiments in bilayer simulations. As a result, fitting to C36 PMFs and further reweighting targeting at the experimental A_l and the C36 *trans*/*gauche* ratio were needed. For monounsaturated chains, the 2D fit to QM generated accurate deuterium order parameters near the double bonds in the lipid bilayer simulations. For PUFA chains, the transition barriers between the minima given by the fitted parameters were lower than QM, and the shape of the wells were slightly different. However, the A_l and chain order parameters calculated for the PUFA lipid bilayers were comparable to those of C36p[77] (the update on the original C36 set for PUFA tails).

4.3.2. Hydrocarbons

Results from LJ-PME simulations for hydrocarbons are shown in Table 25 and Table 26. Densities and heats of vaporization are accurately modeled, while other properties are also in decent agreement with experiments. Densities of hexadecane at tested temperatures are slightly better than heptane, while heptane is better modeled for isothermal compressibility (only 3.0% above experiments on average) and heat of vaporization (only 1.2% below experiments on average). For alkenes, C36UAr generated lower densities compared to experiments. This might be limited by the number of parameters we used to do the fit, because the optimization of CH1E was based on the optimized CH2E and CH3E parameters. However, an average error of 3.3% for the tested systems is acceptable. The surface tension is hard to model with UA model and an average error of ~20% for all tested systems was observed. Previous work indicated that the long-range LJ forces play an important role in determining surface tension of the liquid/vapor interface, and the inclusion of more long-range LJ forces can change the density profiles of alkanes dramatically.[96] Since we conducted all the hydrocarbon simulations using the LJ-PME method, the way of treating nonbonded interactions should not be a concern. Instead, the large deviation is likely caused by the UA model itself. Therefore, one should be careful when using this model for lipid monolayer simulations.

Table 25. Bulk Properties of Hydrocarbon from C36UAr and experiments (Expt.). If ±standard error is not given, then the standard error of C36UAr less than last digit.

Molecule	Temperature (K)	Density (g/cm^3)		Compressibility (10^{-10} m^2/N)		Heat of vaporization (kcal/mol)		Liquid/vapor surface tension (dyn/cm)	
		C36UAr	Expt.[233-237]	C36UAr	Expt.[235]	C36UAr	Expt.[237-240]	C36UAr	Expt.[235]
n-heptane	303.15	0.663	0.675	15.47 ± 0.16	15.03	8.6	8.7	22.8	19.5
	312.15	0.656	0.666	16.75 ± 0.19	16.25	8.4	8.5	21.8	18.5 [c]
Avg. Deviation		-1.6%		3.0%		-1.2%		17.3%	
n-hexadecane	298	0.767	0.770	8.14 ± 0.03		20.2	19.5	33.9	27.6 [b]
	303.15	0.764	0.767	8.16 ± 0.06	8.90	20.1	19.4	33.1	26.7
	310.15	0.759	0.762 [a]	8.48 ± 0.04	9.3 [a]	19.9	19.1[a]	32.7	26.1
	312.15	0.758	0.761 [a]	8.85 ± 0.06		19.8	19.0[a]	32.6	26.2 [c]
	323.15	0.751	0.753	9.40 ± 0.07	10.06	19.5	18.0	31.4	25.0
Avg. Deviation		-0.4%		-7.9%		4.8%		24.4%	
cis-2-undecene	293	0.742	0.754	9.32 ± 0.03		13.9			
	348	0.705		13.07 ± 0.23		13.0	12.3		
Avg. Deviation		-1.6%				5.7%			
cis-2-hexene	293	0.658	0.687	16.53 ± 0.08		7.4	7.7		
	298	0.654	0.683	16.50 ± 0.28		7.3	7.5		
Avg. Deviation		-4.2%				-3.3%			

[a] Values at 310.15 K and 312.15 K interpolated with a polynomial fit. [b] Value at 298 K interpolated with a polynomial fit. [c] Experimental data is for 313 K.

Viscosities and self-diffusivities of heptane and hexadecane were also calculated (Table 26). Self-diffusivity was overestimated by a noticeable amount, though it was included in the fitting targets. Previous work has shown that C36 was able to predict the self-diffusivity of hexadecane precisely with an error of 3.0% when

using LJ-PME, indicating that the UA model failed to catch some of the dynamics in the liquid alkane systems. It could be that the UA representations of methyl and methylene groups causing more collisions between molecules (larger van der Waals radius), and that the energy surfaces of the UA molecules are much smoother than AA ones. The viscosity for heptane was overestimated despite a higher-than-experiment self-diffusivity, indicating the conformation of the short alkanes were not well represented and the terminal methyl groups may not be perfectly modeled by the UA approach or that the properties of this terminal methyl change with increasing alkane length. Since the focus here is on lipids, this disagreement for alkanes did not require further parameterization.

Table 26. Dynamic properties of alkanes from C36UAr and experiments (Expt.). Errors for simulations are standard errors from 4 replicas.

Molecule	Temperature (K)	Viscosity (cP)		Self-diffusivity (10^{-6} cm^2/s)	
		Sim	Expt.[235, 241-244]	Sim	Expt.[245-247]
n-heptane	303.15	0.52 ± 0.07	0.37	41.7 ± 0.6	32.2
	312.15	0.52 ± 0.04	0.33	46.5 ± 0.5	37.2
Avg. deviation		49%		27%	
n-hexadecane	298	2.19 ± 0.26	3.02	4.97 ± 0.08	
	303.15	2.12 ± 0.32	2.77	5.94 ± 0.05	4.50
	310.15	1.69 ± 0.15	2.39 [a]	6.69 ± 0.07	
	312.15	1.49 ± 0.20	2.30	7.09 ± 0.07	
	323.15	1.33 ± 0.11	1.85	8.52 ± 0.12	6.32
Avg. deviation		-29%		33%	

[a] Values at 310.15 K interpolated with a polynomial fit.

4.3.3. Lipid bilayers

A_l is an important metric for testing the quality of a lipid FF. C36 is known to reproduce A_l within the experimental errors for various lipid heads and tails. Since the model being parameterized keeps the full resolution of the C36 headgroups, presumably it can generate the same accurate A_l once the tails are parametrized properly. It has been shown in Section 4.3.2 alkane densities were in good agreement with experiments, and the errors were comparable to C36. However, when using these

hydrocarbon parameters directly on the lipid bilayers with a force-based LJ switching function over 10 to 12 Å, the A_l for DPPC was significantly overestimated. This was found to be related to the dihedral angle distributions in the bilayer simulations. While the alkane dihedrals for the UA model were fit to the same QM data as C36, they did not generate the same distributions for the CH2-CH2-CH2-CH2 dihedrals in DPPC bilayer simulations. PMF calculations were performed for these dihedrals (Figure 31). The energy gap between the *gauche* and *trans* rotamers for CH2-CH2-CH2-CH2 was about 1.05 kcal/mol for C36 but only 0.72 kcal/mol for the UA model from NPT simulations. A NPAT simulation was also run, and no significant improvement was observed, indicating that a re-fit for the dihedral was needed. This problem was solved by fitting the dihedral parameters to the PMFs calculated from a C36 hexadecane simulation (Figure 30) and further reweighting targeting the experimental A_l and the C36 *trans/gauche* ratio of DPPC and DMPC bilayers. It took only two iterations to fix the A_l for both lipids. Though the *trans*/*gauche* ratio was not in perfect agreement with C36 at the end of the reweighting (1.76 vs. 2.44), calculated S_{CD} profiles for selected bilayers agreed well with experiments.

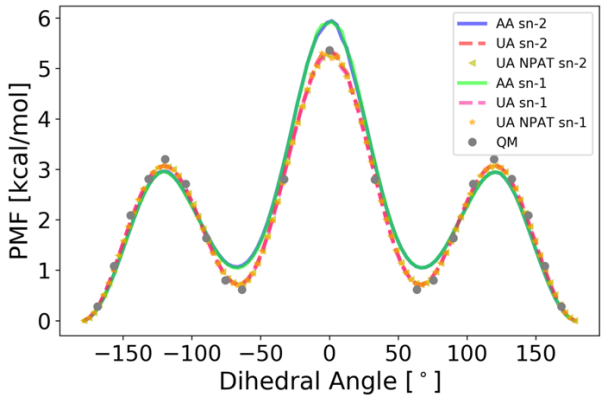

Figure 31. PMFs of CH2-CH2-CH2-CH2 for DPPC, the constant area used in the NPAT simulation was 63.0 Å^2. The UA model shown in this plot uses the dihedral parameters fitted to the QM data.

After circumventing this obstacle, the model was tested against a wide range of lipid headgroups and tails. Tested systems and results for A_l and K_A can be found in Table 27. Overall, MD simulations with C36UAr resulted in accurate A_l. For lipids containing only saturated tails (DSPC, DPPC, DMPC, DLPC and DLPE), C36UAr is slightly better than C36. This is not surprising because dihedral parameters were fine-tunned in the last reweighting step to match the A_l of DPPC and DMPC. For lipids containing monounsaturated tails (DOPC, POPC, POPG and POPE), C36UAr achieved the same level of accuracy as C36. For PUFA lipids (SDPC and DAPC), the only surface area data available is SDPC bilayer at 297 K, for which C36UAr and C36 are comparable. It should be noted that the last version of C36UA failed to model the chain-length dependence of A_l accurately - the A_l for DPPC and DMPC were about the same (61.7 ± 0.3 Å^2 and 61.6 ± 0.3 Å^2, respectively). K_A's estimated by C36UAr were lower than C36 and the experiments, which is consistent with the lower isothermal compressibilities observed for hexadecane at different temperatures. Besides, A_l and K_A for selected C36HMR lipid bilayers can be found in the published paper. While the

A_l for C36HMR does not differ from C36, K_A's for DLPE and POPE bilayers are higher than C36, which means less fluctuation in the time series of A_l. A recent study by Balusek et al.[232] also observed similar A_l of DPPC between C36 and C36HMR.

Table 27. Surface area per lipid and lateral compressibility (K_A) for tested systems. A_l data for POPG came from two publications[71,72]. Standard errors are shown in parenthesis.

Lipid, Temperature (K)	A_l (Å^2)			K_A (dyn/cm)		
	C36UAr	Expt.[71–75]	C36[38,76,77]	C36UAr	Expt.[78–82]	C36[38,77]
DSPC, 333.15	63.7 (0.2)	63.8	61.8 (0.3)	220 (20)		
DPPC, 323.15	63.3 (0.2)	63.0	62.9 (0.1)	170 (20)	231	230 (20), 210 (10)
DMPC, 303.15	60.8 (0.3)	60.6	61.5 (0.1)	180 (20)	234	210 (40), 210 (30)
DLPC, 303.15	62.1 (0.2)	60.8	63.1 (0.3)	260 (20)		
DOPC, 303.15	67.2 (0.2)	67.4	68.9 (0.1)	240 (10)	300, 260	280 (10), 290 (20)
POPC, 303.15	63.8 (0.3)	64.4	66.0 (0.1)	250 (10)	180-330	240 (10), 280 (30)
POPG, 303.15	67.8 (0.2)	64.3, 66.0	67.5 (0.1)	180 (20)		220 (20), 180 (30)
POPE, 308.15	56.2 (0.2)	58.0	58.7 (0.1)	280 (20)	233	280 (20), 260 (20)
DLPE, 308.15	53.7 (0.1)	51.7	54.4 (0.3)	280 (20)		300 (10)
SDPC, 297	67.7 (0.2)	68.2	69.7 (0.1)	210 (20)		285 (40)
SDPC, 303	68.8 (0.2)		70.8 (0.2)	270 (20)		220 (10)
DAPC, 303	74.2 (0.2)		76.7 (0.2)	230 (20)		240 (10)
DAPC, 323	75.3 (0.2)		78.2 (0.2)	220 (20)		240 (10)

It should be noted that error of A_l of lipid bilayer from experiments is usually in the range of 1 to 2 Å^2, and more direct comparisons with experiments are made possible by calculating the X-ray and neutron form factors. Selected results are presented in Figure 32 while data for more systems can be found in the published paper.[30] For saturated lipids, C36UAr agreed well with experimental X-ray scattering data for all tested lipids. The first two lobes of the X-ray form factors for DLPC from simulation are slightly smaller than experiment, indicating that the actual bilayer could be more expanded in the *x*-*y* plane. For lipids with monounsaturated tails, the results were also impressive. POPC and POPG were in excellent agreement with both X-ray and neutron experiments. The lobes of the simulated X-ray form factors for DOPC are a little bit too slim but the neutron form factors agree well with the experiment. The simulated form factors for POPE deviate most from experiments, indicating that the surface area might be underestimated. For PUFA lipids, X-ray data for SDPC at 297 K is available, but the comparison between the experiment and the simulation is not very informative due to the limited number of data points from experiments. However, with an average root-mean error (RMSE) of 0.26, the result from C36UAr is comparable to C36p (RMSE = 0.18) and better than the original C36 FF (RMSE = 0.35). It should be stressed that the quality of the form factors is uniform for different headgroups.

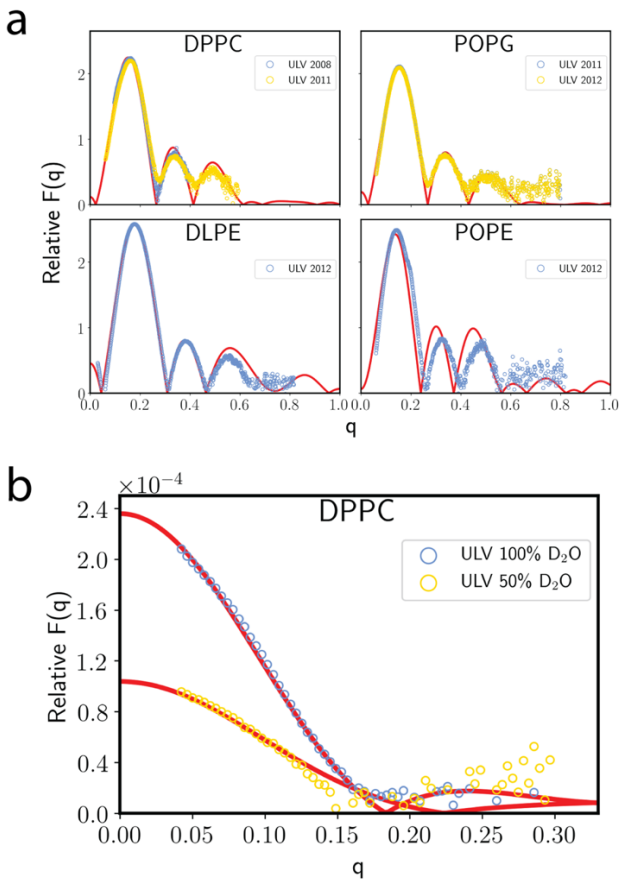

Figure 32. Form factors from C36UAr and experiments.[131, 157, 159, 160, 163] Years in legends are for the experimental collection dates but may differ from the publications. (a) X-ray form factors for DPPC at 323.15 K, POPG at 303.15 K, DLPE and POPE at 308.15 K; (b) Neutron form factors for DPPC at 323.15 K.

Table 28 also compares the membrane thicknesses with experiments. The simulated D_B agreed well with experiments for all the three headgroups tested, though D_B of PC and PG lipids were slightly underestimated while those of PE lipids were overestimated. D_{HH} generated by C36UAr were higher than experiments except POPG, which was statistically the same as experiment. Out of the three types of thicknesses, $2D_C$ was most accurately modeled by C36UAr. The deviations from experiments were within the standard errors for almost all lipids tested, which indicates that the

parametrization of the hydrocarbons was successful. Comparisons of membrane thicknesses between C36UAr and C36 are available in the published paper,[30] and the previous version of C36UA is also included in those comparisons.

Table 28. Overall thickness (D_B), headgroup-headgroup distance (D_{HH}) and hydrophobic thickness ($2D_C$) for each lipid bilayer from C36UAr and experiments (Expt.). The experimental values with uncertainties for saturated PC lipids and POPC from 2011, Kučerka et al.[131] DOPC experimental values from earlier (2008) work of same group.[157] PE experimental values from their later (2015) work focusing on PE.[159] POPG from their 2014 paper.[161] SDPC obtained by Eldho et al.[248] Standard errors for C36UAr are less than 0.1 Å.

Lipid	Temperature (K)	D_B (Å)		D_{HH} (Å)		$2D_C$ (Å)	
		C36UAr	Expt.	C36UAr	Expt.	C36UAr	Expt.
DLPC	303.15	31.6	32.6 ± 0.7	31.0	29.8	21.7	21.7 ± 0.4
DMPC	303.15	35.9	36.7 ± 0.7	35.6	35.3	25.7	25.7 ± 0.5
DPPC	323.15	38.7	39.0 ± 0.8	38.0	38.4	28.6	28.5 ± 0.6
DSPC	333.15	40.9	42.2 ± 0.9	41.2	43.3	31.7	31.9 ± 0.6
POPC	303.15	38.3	39.1 ± 0.8	38.4	36.5	29.0	28.8 ± 0.6
DOPC	303.15	38.1	38.7	39.2	37.0	29.4	28.8
SDPC	297	39.3		40.2	37.9 ± 0.1	30.8	
Avg. Deviation (PC)		-2.1%		2.3%		0.4%	
DLPE	308.15	35.4	34.9 ± 0.7	34.8	33.5	24.8	25.4 ± 0.5
POPE	308.15	43.4	40.5 ± 0.8	42.8	38.3	33.3	32.1 ± 0.6
Avg. Deviation (PE)		4.3%		7.8%		0.7%	
POPG	303.15	36.2	37.6 ± 0.8	36.4	36.6	27.8	28.5 ± 0.5
Avg. Deviation (PG)		-3.7%		-0.6%		-2.4%	
Avg. Absolute Deviation		2.8%		3.2%		1.3%	

While S_{CD} for the headgroup can be easily obtained, we only report the chain S_{CD} for selected lipids from C36UAr because the headgroup S_{CD}'s are expected to be similar as C36. There are models developed to relate experimental acyl chain order parameters and surface area, and one of them is the model from Nagle[249]

$$Area\ per\ chain = \frac{2V_{CH2}}{(1+2S)b_{CC}} \tag{26}$$

where S is the average S_{CD} of the "plateau" region (usually defined as carbon 4-8 of saturated chains), V_{CH2} is the volume of one CH2 group, and b_{CC} is the projected length of the C-C bond along the bilayer normal for an all-trans chain. It should be noted that simulations done by Venable and co-workers[108] supported this model as the saturated lipids they tested agreed well with Eq. 26. It follows directly from the equation that higher S_{CD} of the "plateau" region means lower A_l, so that S_{CD} is also valuable data to test whether the surface area in a simulation is good or not. Figure 33 compares the acyl chain order parameters of three characteristic lipids with experiments. DPPC has two saturated tails with 16 carbons, and C36UAr successfully reproduced the order parameters for nearly all the carbons, except that the splitting of the C2 is not as accurate as C36. POPC has a double bond at the carbon-9 position of the *sn*-2 chain. Perfect matches between the simulated and experimental S_{CD}'s around that region were observed, while S_{CD}'s for other carbons were also good. SDPC is the only PUFA lipid for which experimental S_{CD}'s are available. C36UAr was able to reproduce the experimental S_{CD}'s for the *sn*-1 chain while higher-than-experiment S_{CD}'s were observed for the *sn*-2 chain. It should be mentioned that the order parameters for the

sn-2 chain from C36UAr was closer to C36 but not C36p, while the surface area was closer to C36p and agreed best with experimental estimate.

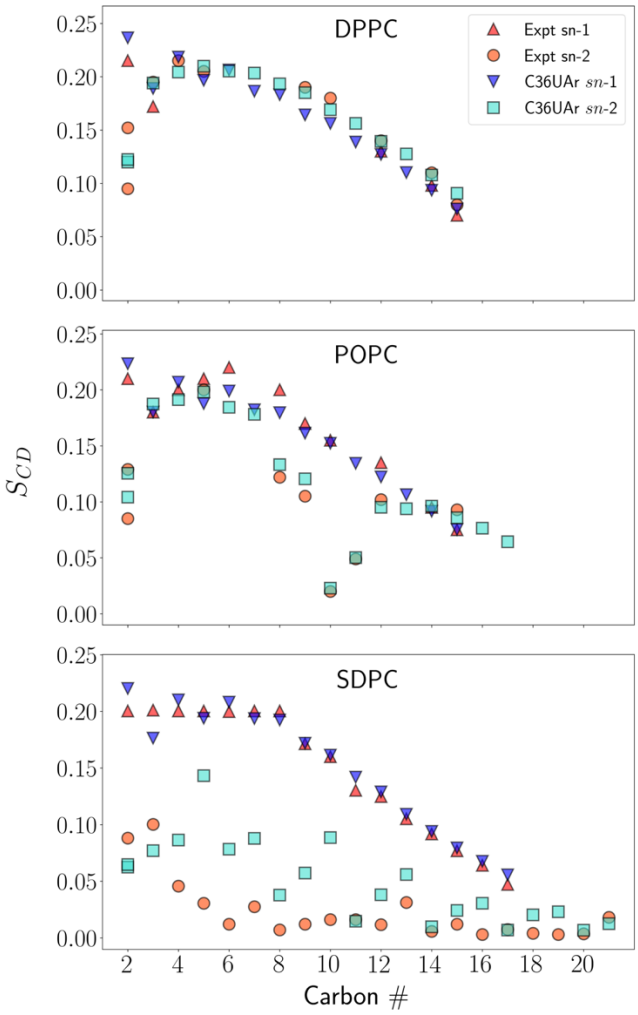

Figure 33. Acyl chain order parameters for DPPC, DOPC and SDPC. Temperature for DPPC is 323.15 K and the experimental data was obtained at the same temperature;[132, 134, 167] Simulation temperature for POPC is 303.15 K while the experiment was done at 300 K;[222] Temperature for SDPC is 303 K, temperature used in the experiment was 298 K.[248]

4.3.4. Phase transition and hydrogen bonding

This subsection focuses on the phase transition and hydrogen bonding, and comparison with C36UAr and C36HMR, while Section 4.3.5 provides more details for the dynamic properties.

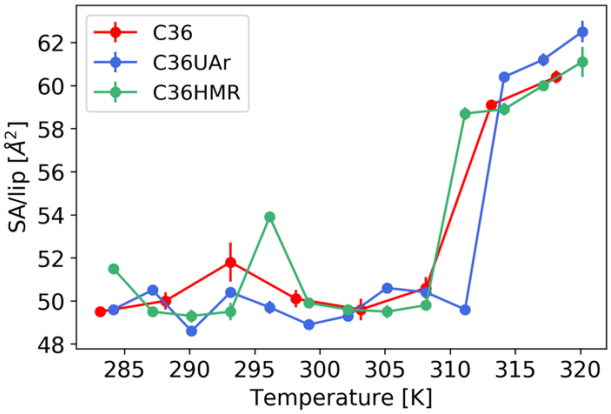

Figure 34. A_l for DPPC bilayer at different temperatures. C36 data from Khakbaz et al.[250]

A previous study by Khakbaz et al.[250] has shown that C36 is able to predict the main transition temperatures from the liquid-crystalline (L_α) phase to ripple (P_β) or gel (L_β) phases for DMPC and DPPC bilayers, though slightly underestimated when compared with experimental results. In this work, we seeded simulations from 283.15 K to 320.15 K every 3 Kelvins to study the phase transition of DPPC bilayer for both C36UAr and C36HMR. The initial coordinates were equilibrated under 323.15 K and NPT ensembles were used to conduct all simulations of 300 ns. All simulated systems reached equilibrium after 200 ns so that the last 100 ns from all simulations were used to calculate membrane properties. It should be noted that the system size used here (72 lipids) was not large enough to predict the wavelength of the P_β phase. The average A_l for all temperatures is in Figure 34; previous results for C36 are also included. It is

clear from Figure 34 that C36UAr predicted the main transition temperature of DPPC (experimentally determined to be 314.15 K[195]) most accurately among the three. Average surface areas per lipid from the gel phase replicas were 49.7 ± 0.1 Å^2 for C36UAr, 51.8 ± 0.2 Å^2 for C36HMR and 49.9 ± 0.3 Å^2 for C36. The gel phase A_l for DPPC determined by Nagle and co-workers was 47.2 ± 0.5 Å^2 at 292 K.[251] In their measurement, the inversely proportional relationship between the cosine of the average tilt angle of lipid tails and the surface area was used. While the gravimetric/hydration method was more commonly used at that time, the authors argued that defects in the multilamellar vesicles (MLV) samples used in the gravimetric/hydration method would yield higher number of hydration numbers, leading to overestimates of A_l for bilayers in the gel phase. Assuming Nagle's result is more accurate, C36UAr is closer to the experimental result compared with C36HMR. In addition, C36HMR has a much higher chance to get trapped in the sub-gel phase compared to C36UAr; among the 10 temperatures below the experimental main transition point, 5 for C36HMR were trapped. C36 simulations performed by Khakbaz et al. also saw the sub-gel phase formed but the proportion is much lower (~17%). The proportion for C36UAr was calculated to be 20%. Melting of headgroups and interdigitation of leaflets were seen at 284.15 K and 287.15 K for C36UAr, indicating forming of the P_β phase. For C36HMR, P_β phase were formed in the 287.15 K and 296.15 K simulations, with the 296.15 K replica yielding a much higher A_l compared with other temperatures. This higher A_l was studied by looking at the leaflets interdigitation, which was defined by Das et al.[252] as

$$\lambda_{\mathrm{ov}} = \int_{-L}^{+L} 4 \frac{\rho_t(z) \times \rho_b(z)}{(\rho_t(z) + \rho_b(z))^2} \mathrm{d}z \qquad (27)$$

where λ_{ov} is the parameter measuring the degree of interdigitation, $\rho_{\mathrm{t}}(z)$ and $\rho_{\mathrm{b}}(z)$ are the electron densities of the top and bottom leaflets at distance z from the bilayer center, [-L, L] is the region along the z axis in which $\rho_{\mathrm{t}}(z)$ or $\rho_{\mathrm{b}}(z)$ is not zero. The calculated interdigitations were 13.2 Å for the 296.15 K replica and 11.0 Å for the 287.15 K replica. For reference, the interdigitation for the 302.15 K replica (gel phase) was also calculated, and the result was 5.8 Å. The higher interdigitation in the 296.15 K replica can lead to a thinner bilayer and thus a larger surface area if the total volume occupied by the bilayer is fixed. When the P_β phase was formed, decrease in the average tilt angle of tails was observed (Table 29). It should be mentioned that the average tilt angle in the gel phase for C36UAr (31.8 ± 0.8°) also agreed well with the experimental result (32.0 ± 0.5°);[251] the average tilt was calculated to be 29.2 ± 1.8° for C36HMR and 36.1 ± 0.6° for C36.

Table 29. Average tilt angle of DPPC tails and corresponding phase identified based on simulation snapshots.

Temperature (K)	C36UAr (°), phase	C36HMR (°), phase
284.15	20.3, ripple	27.7, sub-gel
287.15	20.0, ripple	18.6, ripple
290.15	33.0, gel	27.9, gel
293.15	29.1, sub-gel	32.9, sub-gel
296.15	31.3, gel	18.1, ripple
299.15	31.1, gel	33.7, sub-gel
302.15	30.4, gel	32.9, gel
305.15	30.0, gel	31.9, sub-gel
308.15	35.0, gel	30.8, sub-gel
311.15	29.3, sub-gel	26.9, gel

Inter-lipid hydrogen bonding of PE lipids was investigated using CHARMM. Table 30 lists the probabilities to form inter-lipid hydrogen bonds and the lifetimes of those hydrogen bonds. All PE bilayers were simulated at 308.15 K; H-acceptor distance was restricted to be less than 2.4 Å and the donor-hydrogen acceptor angle was set to be more than 150°. The lifetime was calculated using the "AVHBLF" substitution parameter in CHARMM, which counts and averages the number of consecutive frames

a particular hydrogen bond exists. Based on Table 30, both the hydrogen bonding pattern and lifetime were not influenced by the UA model and the HMR method.

Table 30. Inter-lipids H-bond probabilities and H-bond lifetimes. Probabilities calculated on a per lipid basis. Standard errors calculated from 3 replicas of 80-ns simulations and shown in parenthesis.

Lipid	DLPE			POPE		
Method	C36UAr	C36HMR	C36	C36UAr	C36HMR	C36
N-H:::O-P/O=P	0.641 (0.004)	0.630 (0.005)	0.627 (0.005)	0.596 (0.004)	0.591 (0.006)	0.573 (0.005)
Lifetime (ps)	3.68 (0.01)	3.64 (0.02)	3.66 (0.01)	3.65 (0.02)	3.59 (0.01)	3.60 (0.02)
N-H:::O=C	0.071 (0.002)	0.061 (0.002)	0.062 (0.002)	0.066 (0.002)	0.062 (0.001)	0.057 (0.002)
Lifetime (ps)	2.56 (0.02)	2.57 (0.02)	2.62 (0.02)	2.58 (0.01)	2.60 (0.03)	2.53 (0.01)

4.3.5. Dynamics of lipid bilayers

Diffusion constants of individual lipids in a bilayer are available in Table 31. Bayesian analysis was conducted following the protocol described in Section 2.3.7 to extrapolate diffusion constants for infinite system size (D^{∞}). Only the 288 systems were used as inputs. D^{∞} are reported in Table 31.

Table 31. Diffusion constants for DPPC and DOPC calculated by C36UAr and C36HMR. $D^{bi\text{-}}$ stands for diffusion constant after removing COM of whole bilayer; $D^{mono\text{-}}$ is diffusion constant after removing COM of individual leaflet. D^{∞} is the diffusion constant for infinite system size extrapolated using Bayesian analysis based on $D^{mono\text{-}}$ of the 288 system. 95% trust regions are used for D^{∞}.

Lipid, Temperature (K)	System size (# of lipids)	$D^{bi\text{-}}$ (10^{-7} cm^2/s)	$D^{mono\text{-}}$ (10^{-7} cm^2/s)	D^{∞} (10^{-7} cm^2/s)	$D^{expt.}$ (10^{-7} cm^2/s)
DPPC-C36UAr, 323.15	72	3.28 ± 0.02	1.57 ± 0.02	(6.20, 7.48)	1.5[180]
	288	2.52 ± 0.03	2.52 ± 0.02		
DPPC-C36HMR, 323.15	72	2.74 ± 0.01	1.70 ± 0.02	(6.20, 7.50)	
	288	2.51 ± 0.03	2.51 ± 0.02		
DOPC-C36UAr, 303.15	72	1.37 ± 0.01	0.68 ± 0.01	(2.05, 2.59)	0.825[181]
	288	1.04 ± 0.02	1.04 ± 0.01		
DOPC-C36HMR, 303.15	72	1.19 ± 0.02	0.91 ± 0.01	(3.41, 4.22)	
	288	1.25 ± 0.02	1.25 ± 0.01		

Results from two different methods of removing the COM motion are also compared in the Table 31. The 288 systems are barely influenced by the different methods, while the 72 systems see significant differences caused by the sliding of the

two leaflets relative to each other, and the same trend has been observed in the previous study.[154] Difference between C36UAr and C36HMR is not statistically significant for DPPC. However, C36UAr shows significant differences with C36HMR for DOPC and is in better agreement with experiment. C36 generated $D^{\infty} = (4.8 - 6.6) \times 10^{-7}$ cm^2/s for DPPC at 323.15 K and $D^{\infty} = (2.0 - 2.7) \times 10^{-7}$ cm^2/s for DOPC at 303.15 K,[152] the later one is close to simulations using C36UAr. These results are consistent with the faster diffusion of alkanes observed for C36UAr. The diffusion constants calculated by Balusek et al.[232] are systematically smaller than what we got even with larger DPPC bilayers (240 lipids/leaflet). Possible reasons include the use of Langevin dynamics, and the lower A_l observed in their study.

4.3.5. Limitations of C36UAr

Although C36UAr can reproduce many observables for a wide range of lipids, the implicit treatment of chain hydrogens can lead to problems associated with both the relatively smooth surface and diminished local dipole between the carbons and hydrogens.

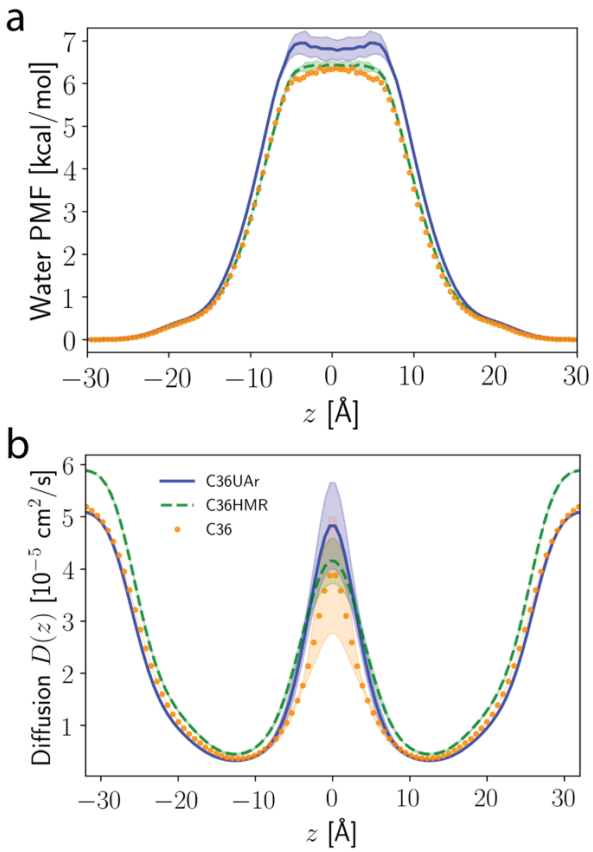

Figure 35. (a) Symmetrized PMFs of water along the DOPC bilayer normal. Errors for C36UAr and HMR from 3 replicas (900 ns used for each replica). C36 PMF taken from Krämer et al.[47] (b) One dimensional diffusion profile of water along the DOPC bilayer normal. Errors for C36UAr and C36HMR from two different sets of optimizing basis functions and 3 replicas (300 ns used for each replica). Errors for C36 from two different sets of optimizing basis functions and 3 blocks of a single trajectory (each block is 120 ns).

One good example is the water permeability through lipid bilayer. Krämer et al.[47] have shown in their recent study that the TIP3P water model may underestimate the permeability of water through the bilayer in MD simulations using additive FFs, primarily due to the lack of polarizability of the additive force field. Their counting method was used to calculate the permeability of water through a DOPC bilayer. The

calculated permeability at 298 K is $(1.1 \pm 0.2) \times 10^{-3}$ cm/s for C36UAr and $(2.5 \pm 0.2) \times 10^{-3}$ cm/s for C36HMR, the errors are from three replicas of 900-ns simulation. Same calculation was performed for C36 but using only one replica (400 ns), which generated a permeability of 1.8×10^{-3} cm/s, this number is bracketed by $(1.5 \pm 0.4) \times 10^{-3}$ cm/s obtained by Krämer et al. To help understand these results, water PMFs and diffusion profiles were estimated (Figure 35). The PMFs were generated directly using the water distribution along the bilayer normal, while the one-dimensional diffusion profiles along the bilayer normal were generated by a recent method developed by Krämer et al.,[253] which uses the Covariance Matrix Adaptation-Evolution Strategy. The PMFs in Figure 35 suggest that the UA model has a transition barrier higher than C36 and C36HMR. This is not surprising since the polarizations of the methylene and methyl groups are further decreased to zero in C36UAr. Even though the overall diffusion rate within the bilayer (~ -20 Å < z < 20 Å) for C36UAr is faster than C36, the exponential dependency of permeability on the negative relative free energy suggested by the inhomogeneous solubility diffusion (ISD) model[23, 254, 255] may dominate the process and the net effect is C36UAr having the lowest permeability. The higher permeability of C36HMR compared with C36 can be explained by the faster diffusion throughout the membrane, but it should be noted that the estimated diffusion profiles come with large uncertainties due to the poor sampling near the bilayer center.

The compressibility modulus of various bilayers with C36UAr were overall underestimated compared to both the experiments and the all-atom FF simulations, which is probably caused by the smoother surface and higher fluidity of the UA tails. Diffusion constant is less of a concern because the result for DOPC is close to C36,

though DPPC is a little faster. However, there must be other reasons for the too fast diffusion for the CHARMM lipid FF family; a discussion on this issue is available in the paper by Venable et al.[152]

4.4. Discussion

In this chapter, new potential energy parameters of united-atom hydrocarbons were developed for the CHARMM FF. The parametrization was initially targeted at bulk properties of linear alkanes and alkenes, and further modification was made possible by comparing to experimentally based surface areas of lipid bilayers and C36 dihedral distributions. The final parameters were tested against a wide range of targets including bulk liquid properties of hydrocarbons and properties of single lipid bilayers. The dynamics of hydrocarbons and lipid bilayers were also studied. The overall agreement of C36UAr with experiment and C36 was great, and it can be used for a wider range of lipid headgroups and tails compared to the original C36UA parameter set. Since C36UAr uses the additive scheme for the nonbonded interactions and the functional forms of the potential energy are consistent with the CHARMM community of FFs, it is compatible with the rest of the C36 additive FF.

C36UAr was successful in reproducing the bulk properties for various alkanes and alkenes over a considerable range of temperatures. The original C36UA is a hybrid of the C36 headgroup parameters and OPLS UA hydrocarbon parameters. While the OPLS parameters were tested to be the best out of the 4 existing UA parameter sets when using the force switch method in the original publication of C36UA, the agreement to experiment was not as good as the new set presented in this paper. The density of heptane at 312.15 K was tested to be 0.64 g/cm^3 for C36UA while the

experimental value is 0.67 g/cm^3. Densities of longer linear alkanes were probably in better agreement with experiments since the density of pentadecane at 312.15 K was tested to be 0.77 g/cm^3, only 0.01 g/cm^3 higher than the experimental value.[234] The lower density for heptane and higher density for pentadecane indicated that the chain-length dependence is not perfectly described by C36UA and it could be difficult to obtain the correct bilayer surface areas for lipids with large variations in different tail lengths (beyond the range tested in this work). Heats of vaporization for C36UA were also lower than experiments by significant amounts. The errors for heptane and pentadecane at 312.15 K were 15.1% and 7.3%, while the newly developed set lowered down the error to less than 5% for both. The parametrization of C36UAr also considered self-diffusivity of bulk alkanes, although this property was not significantly improved. Compared to C36, C36UAr achieved similar accuracy for densities and compressibilities. The average error of hexadecane density was calculated to be -0.40% for C36 at selected temperatures when explicitly including the long range LJ interactions,[101] which is exactly the same as the average error calculated for C36UAr in this study. The isothermal compressibilities of hexadecane deviated from experiments by 13% for C36 on average, while the average deviation for C36UAr is only -7.9%. However, C36UAr is not as good as C36 in terms of diffusion; C36 was only 3.0% off for hexadecane while C36UAr overestimated the diffusion constant by about 30% for both heptane and hexadecane. Moreover, the inaccurate surface tensions for alkane/vapor systems indicated that C36UAr might not be suitable for monolayer simulations.

The ability of C36UAr to accurately model bulk properties of hydrocarbons makes it possible to simulate various lipid bilayers. One of the most important features of C36UAr is that it reproduces the correct surface areas of varying tail lengths. Form factors were also compared to experiments and the overall quality was high. By comparing various membrane thicknesses to experiment based values and C36, the accuracy of C36UAr was further confirmed since the overall agreement of C36UAr with experiment is comparable to C36 and much better than C36UA (Figure 36), and C36UAr is slightly better than C36 for PC lipids. C36UAr also included PUFA lipids by fitting relevant dihedrals to a 2D QM PES. With the new parametrization, C36UAr was tested to be accurate for the PG and PE benchmarks, though further validations should be done for more tails to prove it is a reliable FF for those headgroups.

C36UAr and C36HMR were compared at different levels, and major differences occur in the phase transition behavior and dynamics. In terms of phase transition, C36UAr better estimated the gel phase surface areas and tilt angles of lipid tails; the prediction of the main transition temperature was also slightly better. For lipid diffusion, C36UAr is similarly to C36. The permeation of water molecules through lipid bilayer was also studied, and the calculated water permeability was higher for C36HMR compared with both C36UAr and C36, probably due to the faster diffusion of water. It should be mentioned that TIP3P water model has a faster-than-experiment diffusion,[130] and this overestimation can be magnified by C36HMR. Last, a set of benchmarks was done to compare the speeds of the two methods, using bilayers composed of 72 POPC lipids and 2188 water molecules. All the systems were run on a single Titan-XP GPU with one CPU core. The actual simulation speeds were 275

ns/day for C36HMR and 230 ns/day for C36UAr, and the speed for C36 was 170 ns/day. The boost of speed for C36HMR came from the longer integration step (4 fs) being used (2 fs step size was used for C36UAr and C36). Another way to accelerate simulations of hydrogen-rich systems is the virtual-sites technique,[256, 257] in which the hydrogen degrees of freedom are removed, and the positions of hydrogen atoms are reconstructed after every integration step to calculate the force acting on the heavy atoms. While this method would allow a timestep of 5 fs, membrane properties are influenced by the method,[258] hence it should be used with care.

In conclusion, C36UAr improved the accuracy of C36UA significantly. The new FF can predict both bulk hydrocarbon properties and bilayer properties more accurately, and the spectrum of lipid types that can be simulated with the FF has been extended to include PE and PG headgroups and PUFA tails. Although it is still expected to be a bit slower than the GROMOS FFs, the more detailed headgroups in C36UAr may be desired in some cases and allows a force filed that is compatible with CHARMM. When simulating large systems for extended simulation times, C36UAr can be an alternative to C36 if the user does not care too much about the acyl chain polarization and other defects mentioned above.

Chapter 5: Modeling realistic bio-membranes

This chapter was edited from publications:

> Yu, Y.; Klauda, J. B. Modeling Pseudomonas aeruginosa inner plasma membrane in planktonic and biofilm modes. *The Journal of Chemical Physics* **2018.**
>
> Yu, Y.; Klauda, J. B. Symmetric and Asymmetric Models for the *Arabidopsis thaliana* Plasma Membrane: A Simulation Study. *Journal of Physical Chemistry B* **2021**.

This chapter focuses on two applications of the CHARMM36 lipid force field on realistic bio-membranes. The first is modeling *Pseudomonas aeruginosa* (*P. aeruginosa*) inner plasma membrane; the second is modeling *Arabidopsis thaliana* (*A. thaliana*) plasma membrane. Besides their interesting physical chemistry properties, lipids are top players in protein functions.[1] Experimentally, the interactions between lipids and proteins can be studied at many different levels, and methods can be classified into two groups: *in vivo* and *in vitro*.[259] While being successful in many cases, *in vitro* methods suffers from the fact that they are usually conducted under artificial conditions such as non-physiological concentrations,[260] hence the physiological relevance of the interactions needs to be investigated through orthogonal and more physiological methods.[259] *In vivo* methods, on the other hand, can reduce such artifacts greatly. They usually involve the perturbation of a membrane component or a protein segment followed by the measurement of the effects of these perturbations based on phenotypic readouts.[261-267] However, due to the complex nature of many physiological pathways, interactions cannot always be measured directly and metabolized products

of the mutated lipids may disturb the measurements. To compensate for these shortcomings, MD simulations can be used. Not only can MD simulation model the lipid diversity, but it also provides an atomistic view of the lipid-lipid and lipid-protein interactions at timescales relevant to biology.[268-270] Apart from these benefits, asymmetric membranes can be modeled in MD simulations,[95, 271] though they are not commonly seen in *in vitro* experiments.

5.1. Models for *Pseudomonas aeruginosa* inner plasma membrane

5.1.1. Introduction

The gram-negative bacterium *Pseudomonas aeruginosa* is an opportunistic pathogen that mostly attacks immunocompromised and vulnerable patients, especially those with burn wounds, cystic fibrosis, acute leukemia, organ transplants, and intravenous-drug addiction.[272, 273] It has become a major cause of nosocomial infections worldwide and is highly resistant to a wide range of antibiotics, partly due to the overuse and misuse of antibiotics.[274] The biofilm form of *P. aeruginosa* is drastically more resistant to antimicrobials compared with its planktonic form,[275, 276] and it has been suggested that this resistance is due to restricted antimicrobial penetration, physiological gradients, formation of persister cells and quorum sensing.[277, 278] Although efforts have been made in studying genomic, transcriptomic, proteomic and metabolomic details of *P. aeruginosa* for improving antibiotics,[279-282] there are few studies focusing on its membranes. However, the role of the bacteria's membrane system in the defense process is definitely important because certain defense mechanisms, e.g., quorum sensing, involves release, delivery and uptake of small molecules,[283] these sophisticated functions can be achieved due to the unique physical

and chemical properties of the membrane and also the versatile proteins associated with it.

The inner membrane of gram-negative bacteria is mainly composed of glycerophospholipids. Although wet lab experimentation on cell membranes has provided structural and functional information on cellular membranes,[174] molecular dynamics (MD) offers an effective way of studying lipid membrane biophysics with atomistic resolution. In terms of gram-negative bacteria, the inner membrane is symmetric and contains several different phospholipids while the outer membrane is asymmetric with outer leaflet composed almost exclusively of lipopolysaccharide (LPS) and inner leaflet composed of phospholipids.[284] In this chapter, MD simulations are used to investigate the physical properties of *P. aeruginosa* inner membrane at the atomic level with complex models containing more than 5 lipid types. According to experiments, the lipid composition of *P. aeruginosa* inner plasma membrane changes significantly according to the growth modes, namely planktonic and biofilm,[285, 286] thus it is necessary to model membranes in both modes of growth. After 400 ns of all-atom simulations, properties of the model membranes are calculated and some of them are compared to experimental data or simulation results of other microbial membranes.

5.1.2. Method

An electrospray ionization mass spectroscopy experiment[286] measured the lipid composition of the inner PM of *P. aeruginosa* PAO1 strain in both planktonic and biofilm forms. In the experiment, inner membrane enriched samples were extracted following the spheroplast protocol,[287] only predominant PE and PG were reported. One more step was taken to select the 5 most abundant lipid types for planktonic and 8 for

biofilm since all other lipids together share less than 20 percent of the total lipids. The detailed compositions are shown in Table 32, where the Expt. (%) column is normalized to exclude the ~20% of lipids excluded from the model. For simplicity, we exclude the branched chain lipids because they only appear for a short period and mainly in the outer membrane according to the experiment. As a result, the cyclic-containing lipids are slightly overestimated but can reflect the average level of these lipids during the aging of the biofilm. The amount of phosphatidic acid, cardiolipin and ornithine lipids are very low according to the experiment, thus they are also neglected. In addition to the two model membranes mentioned above, four controlled simulations were run to study the effects of the complexity in the composition and the influence of the cyclopropane-containing lipids. Two of them are single-component membranes containing either 1-palmitoyl-2-cis-11,12-methylene-stearic-acid-sn-glycero-3-phosphoethanolamine (PMSPE) or 1-palmitoyl-2-cis-11,12-methylene-stearic-acid-sn-glycero-3-phosphoglycerol (PMSPG), the other two are named Half-PMSPE and Half-PMSPG whose compositions can be found in Table 32. 45 and 50 lipids per leaflet were used for planktonic and biofilm membranes, respectively, 36 lipids per leaflet were used for the four controlled simulations, other simulation details can be found in 26. All bilayers are fully hydrated with a hydration number of 50 waters per leaflet. Potassium ions were used to neutralize the negative charges introduced by PG lipids.

Table 32. Lipid composition of simulated membranes, single-component membranes are not shown in the table. YOPE: 1-palmitoleoyl-2-oleoyl-*sn*-glycero-3-phosphatidylethanolamine, YOPG: 1-palmitoleoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol, DPPE: 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine, DPPG: 1,2-dipalmitoyl-*sn*-glycero-3-phosphoglycerol.

Model		Planktonic		Biofilm		Half-PMSPE	Half-PMSPE
Lipid	*sn*-1/*sn*-2	Exp (%)	Leaflet	Exp (%)	Leaflet	Leaflet	Leaflet
POPE	16:0-18:1	36.0	16	31.6	15	12	12
POPG	16:0-18:1	17.6	8	15.3	7	6	6
YOPE	16:1-18:1	N/A	0	12.3	6	0	0
PMSPG	16:0-19:0 cyc (11,12)	25.4	11	12.4	6	0	18
PMSPE	16:0-19:0 cyc (11,12)	14.1	6	8.1	4	18	0
DPPE	16:0-16:0	6.9	4	7.3	4	0	0
YOPG	16:1-18:1	N/A	0	6.6	4	0	0
DPPG	16:0-16:0	N/A	0	6.6	4	0	0

Table 33. System size and parameters for each membrane.

Model	Number of lipids	number of H_2O	number of K+	Total atom number
Planktonic	90	4500	38	24934
Biofilm	100	5000	42	27582
Pure PMSPE	72	3600	0	20016
Pure PMSPG	72	3600	72	20232
Half-PMSPE	72	3600	12	19944
Half-PMSPG	72	3600	48	20052

Initial coordinates for these bilayers were built using the CHARMM-GUI *Membrane Builder*.[288-290] However, YOPG, PMSPG and PMSPE were not available in the library of CHARMM-GUI. Atom coordinates for membranes containing these lipids were modified from membranes of similar compositions where these three lipids were replaced by DOPG, DSPG and DSPE, respectively in CHARMM.[98] The additional carbon in the cyclic moiety is counted as carbon 19 (Figure 36).

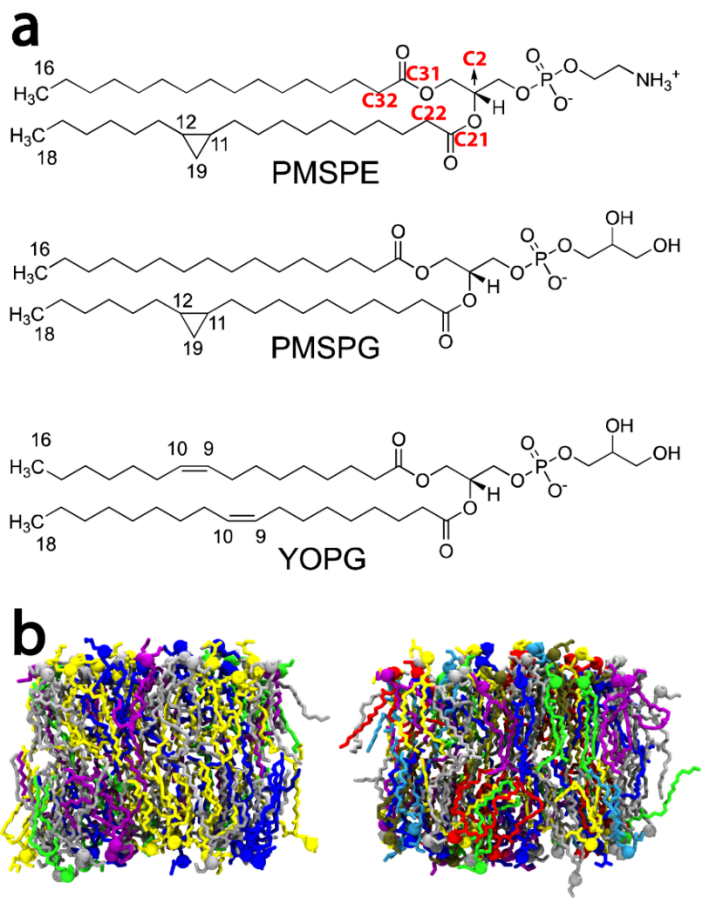

Figure 36. (a) Chemical structures of PMSPE, PMSPG, and YOPG. Labeled carbons are terminal carbons, carbons in cyclopropane, carbons in double bond and carbons used in component area calculation. (b) Snapshot of planktonic (left) / biofilm (right) membrane at the end of the simulation. POPE in silver, POPG in blue, PMSPE in purple, PMSPG in yellow, YOPE in red, YOPG in cyan, DPPE in green, and DPPG in tan. Larger rendering is used for phosphates.

NAMD[291] is used to perform the equilibrium and production runs of the bilayers with the C36 lipid force field[292] and the modified TIP3P water model.[58, 293] Parameters for the cyclic moiety in PMSPG and PMSPE were developed by Pandit et al.[9] After the standard six-step equilibrium using input files generated by CHARMM-GUI, production runs were continued for 400 ns for planktonic and biofilm membranes and

200 ns for the controlled systems with a time step of 2 fs. Three replicas were simulated for each membrane. The NPT ensemble with a constant temperature of 303.15 K and a constant pressure of 1 atmosphere was used. Constant temperature was maintained through Langevin dynamics and constant pressure was kept using the Nose-Hoover-Langevin piston,[228, 229] which allowed the cell box size to change semi-isotropically ($x = y \neq z$). The SHAKE algorithm[294] was used to constrain hydrogen atoms. For long range electrostatics, PME was used.[295] Lennard-Jones potential with a force-switching function[296] from 10 to 12 Å was used to describe the van der Waals interactions.

Analysis of the planktonic and biofilm model membranes was based on the last 150 ns of the production runs, while the last 100 ns was used for the controlled simulations. All systems were confirmed to be at equilibrium from the beginning time of analysis by examining the time series of A_l. Properties computed and analyzed in this chapter include overall A_l, component surface area, K_A, S_{CD}, bilayer thicknesses, and hydrogen bonding. Methods for calculating these properties were described in the preceding chapters except the component surface area, which is computed based on atoms that can stably generate the area at the lipid-water interface. This includes one atom in the lipid headgroup and other two at the linkage region (C2, C21, C31, see Figure 36). The area for a single lipid was then estimated by summing over the areas of the convex polygons constructed around these representative atoms using Voronoi tessellation.[297]

5.1.2. Results

Results listed in this section allow comparisons between the planktonic and biofilm membranes. Further comparisons with experimental data and simulation results for other microbial membranes will be discussed in Section 5.1.3.

Table 34 lists the overall A_l and K_A for the planktonic and biofilm membranes. Pure and simple-mixture membranes are also included for comparison. In addition, Table 35 summarizes the component surface areas for each lipid type in the two model membranes. The overall A_l's for the two models are statistically identical. This can be explained by the close PE:PG ratios (26:19 for planktonic, 29:21 for biofilm) and similar component surface areas for each lipid type (Table 35). The compressibility moduli for the two models are statistically the same (P-value=0.11).

Table 34. A_l and K_A for planktonic and biofilm membranes. Errors are standard errors from the three replicas.

Model	A_l (Å^2)	K_A (dyn/cm)
Planktonic	61.19 ± 0.25	280 ± 20
Biofilm	61.20 ± 0.22	230 ± 20
Pure PMSPE	58.58 ± 0.29	350 ± 50
Pure PMSPG	68.58 ± 0.29	230 ± 30
Half-PMSPE	59.12 ± 0.12	270 ± 40
Half-PMSPG	64.58 ± 0.16	270 ± 50

Table 35. Component areas (Å^2) for each lipid type. Errors are standard errors from the three replicas.

Model	Planktonic	Biofilm
POPE	59.8 ± 0.6	60.7 ± 0.9
POPG	63.0 ± 1.1	61.8 ± 0.5
YOPE		61.0 ± 1.4
PMSPG	61.8 ± 1.1	62.6 ± 1.1
PMSPE	63.6 ± 1.7	60.8 ± 0.5
DPPE	60.1 ± 1.4	59.2 ± 0.1
YOPG		61.9 ± 1.3
DPPG		61.6 ± 1.0

The S_{CD} profiles follow the general trend, decreasing at the occurrence of double bonds and cyclopropanes. Figure 37 compares the S_{CD} of PMSPG in the two model membranes. S_{CD} profiles of the two model membranes are very close to each other, which means the influence of the composition is negligible. However, it is interesting that the order parameters around the cyclic region in PMSPG are quite different between planktonic and biofilm. Carbon-carbon bond orientations relative to the bilayer normal were calculated for this region and differences were observed between the two forms. Based on the 2D angle distributions in Figure 38(b), this region is more flexible in the biofilm membrane, which could potentially cause the lower order parameter for C19.

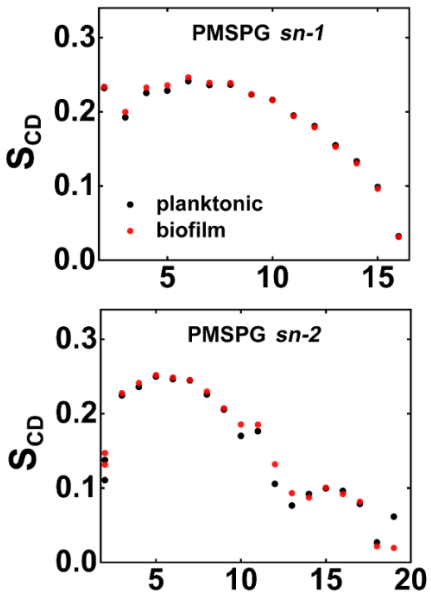

Figure 37. S_{CD} profiles for PMSPG, where *sn-1* and *sn-2* stand for the fatty acid chain linked to the first and second carbon of the glycerol, respectively. The standard errors from the three replicas are smaller than the symbol sizes and are not shown.

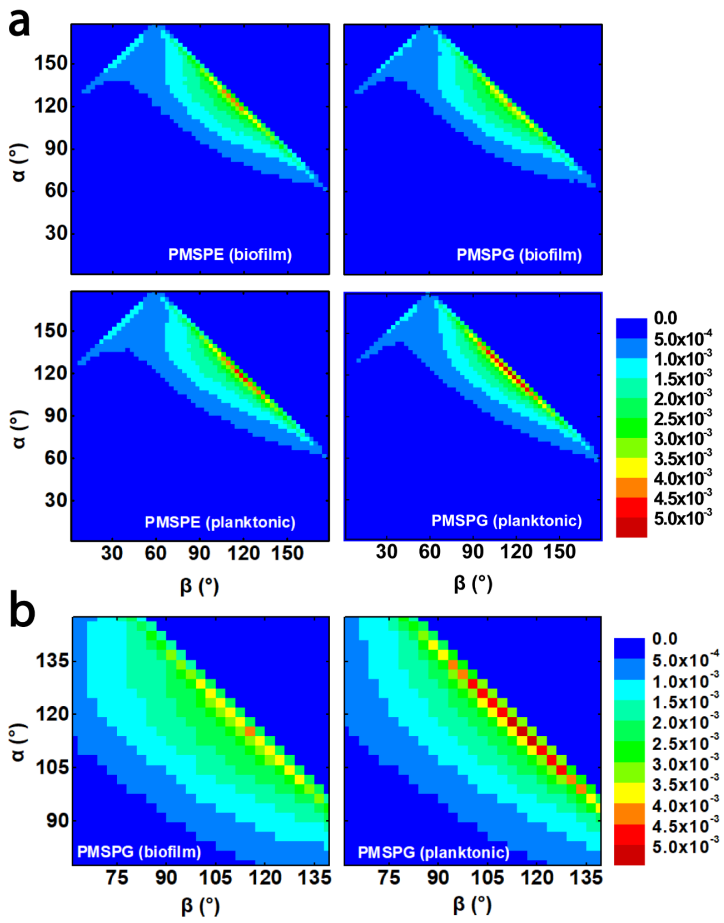

Figure 38. (a) 2D angle distributions for the bond C11-C19 (α) and C19-C12 (β). α and β are defined as the angle between the bond C11-C19 / C19-C12 and the bilayer normal. Distribution was calculated based on a bin size of 3°×3°. (b) The zoomed-in plots for PMSPG.

Previous work by Pandit et al. compared the simulation-based S_{CD} of PMPE (PE 16:0-17:0 cyc(9,10)) to experimental results of l-palmitoyl-2-di-hydrosterculoyl-*sn*-glycero-3-phosphoethanolamine (PDSPE, PE 16:0-19:0 cyc(9,10)),[9] and an 0.04 underestimation of S_{CD} at C10 was observed. Here, the S_{CD} profiles are compared to the same experiment in which the quadrupolar splittings of PDSPE (*sn*-2, carbon number 5,9, and 10) were measured at 303.15K.[298] C9 and C10 in PDSPE were shifted

to match the positions of the cyclic moieties (C11 and C12) in PMSPG and PMSPE. Results are shown in Figure 39 with controlled membranes. S_{CD}'s of C5 in the two model membranes perfectly match the experiment, S_{CD}'s of C11 are slightly overestimated. As for C12, there is a broader distribution, but the experimental value is bracketed. PMSPE in the pure PMSPE membrane and the Half-PMSPE membrane exhibit higher order parameters compared to the model membranes and PDSPE in the experiment due to a significantly smaller A_l. Considering the position of the cyclopropane and lipid impurity in the experiment, these results are satisfactory.

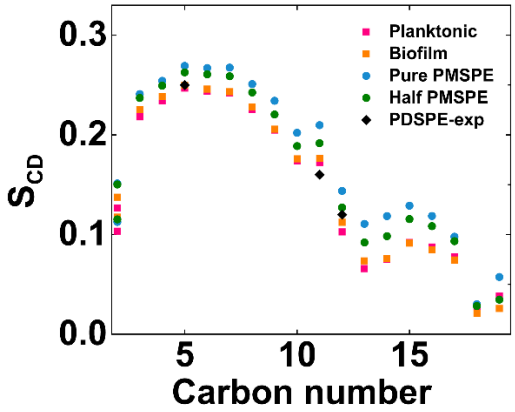

Figure 39. PMSPE *sn*-2 chain S_{CD} profiles. PMSPE in planktonic membrane (pink squares), PMSPE in biofilm membrane (orange squares), pure PMSPE (blue circles), PMSPE in Half-PMSPE membrane (green circles), and PDSPE from experiment at 303.15 K (black diamonds). C9 and C10 in PDSPE are shifted to C11 and C12 to match the position of the cyclopropane in PMSPE, C5 in PDSPE at its original site.

Table 36 lists the thicknesses computed based on the EDP. Planktonic membrane is thicker by about 1 Å compared with its biofilm counterpart. Pure PMSPE membrane is the thickest while pure PMSPG membrane is the thinnest, which is consistent with the A_l results in Table 34, assuming the volume of a lipid is fixed and

the sizes of the two lipids are close. The small difference between planktonic and biofilm might arise from the inclusion of the three additional lipid types in the biofilm membrane and the lower chain interdigitation in the planktonic membrane (see published paper[78] for details). For comparison, The average hydrophobic thickness of transmembrane proteins in *P. aeruginosa* inner membrane was calculated, based on available data from the OPM database.[299] The average turns out to be 30.4 ± 0.7 Å, which is in good agreement with the simulated results.

Table 36. Membrane thicknesses. Errors are standard errors from the three replicas.

Model	D_{HH} (Å)	D_B (Å)	$2D_C$ (Å)
Planktonic	41.1 ± 0.2	39.9 ± 0.2	30.4 ± 0.1
Biofilm	40.3 ± 0.1	38.9 ± 0.3	29.7 ± 0.3
Pure PMSPE	42.2 ± 0.1	41.7 ± 0.1	31.9 ± 0.1
Pure PMSPG	39.4 ± 0.1	36.3 ± 0.1	27.8 ± 0.1
Half-PMSPE	42.1 ± 0.2	41.1 ± 0.1	31.4 ± 0.1
Half-PMSPG	40.1 ± 0.1	37.9 ± 0.1	28.9 ± 0.1

Both PE and PG have huge potential to form hydrogen bonds due to the occurrence of both donors and acceptors in them. The probabilities of forming hydrogen bonds are listed in Table 37. In this calculation, H-acceptor distance was restricted to be less than 2.4 Å and the donor-hydrogen-acceptor angle was set to be more than 150°. It is obvious that PG forms more intra-lipid hydrogen bonds than PE, while PE forms more inter-lipid hydrogen bonds. In terms of intra-lipid hydrogen

bonds, PG has two options with its hydroxyl groups and can reach out better to the oxygens on the phosphate. However, PG has a net negative charge on its headgroup so its interaction with another PG tends to be repulsive, resulting in less inter-lipid hydrogen bonds (see Figure 40). Another reason for PE forming more inter-lipid hydrogen bonds is that its donor, amine, has a positive charge that can be attracted by the negatively charged phosphate on another lipid. It was reported in a previous work using the OPLS force field that PE formed 1.07 ± 0.02 intra-lipids hydrogen bonds per lipid in a POPE:POPG = 3:1 system,[91] which is fairly consistent with the current work, considering the lower concentration of PE in their simulation.

Table 37. Intra-lipid and inter-lipid hydrogen bond count per lipid. Errors are standard errors from the three replicas.

Lipid	intra-lipid		inter-lipid	
	planktonic	biofilm	planktonic	biofilm
POPE	0.085 ± 0.002	0.083 ± 0.003	0.99 ± 0.05	1.04 ± 0.07
POPG	0.461 ± 0.006	0.469 ±0.006	0.79 ± 0.05	0.81 ± 0.08
PMSPE	0.082 ± 0.004	0.081 ± 0.007	1.14 ± 0.09	1.23 ± 0.14
PMSPG	0.489 ± 0.005	0.489 ± 0.008	0.78 ± 0.04	0.79 ± 0.08
DPPE	0.086 ± 0.006	0.082 ± 0.005	1.17 ± 0.11	1.20 ± 0.15
DPPG	--	0.466 ± 0.009	--	0.82 ± 0.10
YOPE	--	0.080 ± 0.004	--	1.17 ± 0.12
YOPG	--	0.468 ± 0.007	--	0.83 ± 0.11

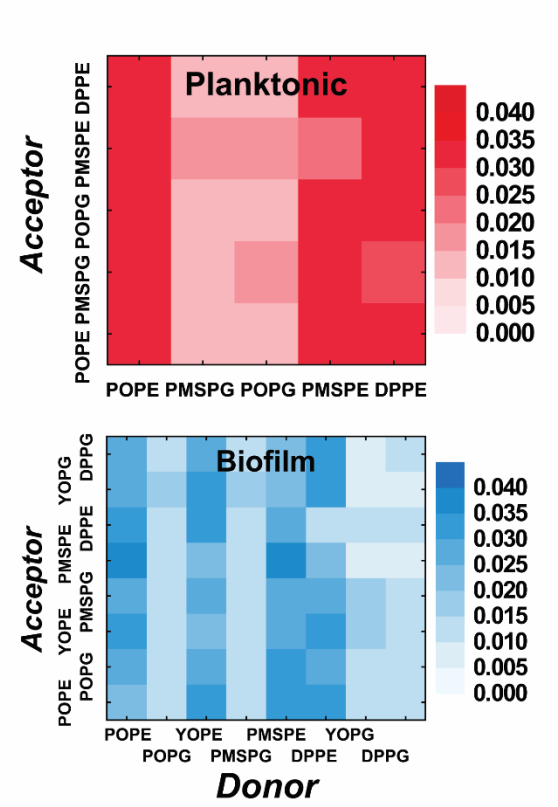

Figure 40. Probabilities of forming hydrogen bond in the *Donor-Acceptor* space, In each model, values in all the blocks sum up to 1.

5.1.4. Discussion

The two model membranes studied in chapter are for the inner PM of *P. aeruginosa* in planktonic and biofilm modes. An important feature of these membrane is the cyclic-containing lipids PMSPE and PMSPG. The overall surface areas per lipid for the two model membranes are statistically the same (61.19 ± 0.25 Å^2 for planktonic and 61.20 ± 0.22 Å^2 for biofilm). The overall A_l computed in a previous study[91] on a POPE:POPG = 3:1 lipid bilayer was 61.5 ± 0.2 Å^2. The *E. coli* membrane model by Pandit et al. has an overall A_l of 64 ± 1 Å^2, and the simulation study by Khakbaz et al.[10]

based on different phases of growing in *E. coli* have A_l ranging from 60.77 ± 0.02 Å^2 to 63.08 ± 0.17 Å^2. Overall, our results agree with these studies on different bacteria.

K_A was computed based on the full equilibrated trajectory and error was estimated through the three replicas for each model. The two-sample *t*-test indicates there are no statistical difference between the two models, although Khakbaz et al. found a consecutive increase in K_A from 0.28 ± 0.02 N/m to 0.33 ± 0.03 N/m for the *E. coli.* membrane as the concentration of cy17:0-containing chains increases.[10] Compared to *E. coli*, the inner PM of *P. aeruginosa* is softer. The K_A of *Chlamydia trachomatis* is even higher (0.58 ± 0.08 N/m for the elementary body cell and 0.46 ± 0.07 N/m for the reticular body cell). The difference between *E. coli* and *P. aeruginosa* can be explained by the higher PE:PG ratio in the *E. coli* membrane because PE can form more inter-lipid hydrogen bonds, which stabilize the membrane bilayer. The major reason for the significantly higher K_A for *Chlamydia trachomatis* is the inclusion of cholesterol since it can rigidify the bilayer. Since lower K_A means more fluidity and less energy cost for proteins insertion, the lower K_A of the *P. aeruginosa* membrane might be more favorable to protein turnover.

Another set of experimental data we can compare are the hydrophobic thicknesses ($2D_C$) of transmembrane proteins in the OPM database.[299] The hydrophobic thicknesses computed for our models are 30.4 ± 0.1 Å (planktonic) and 29.7 ± 0.3 Å (biofilm), which are statistically identical to the experimental average, 30.4 ± 0.7 Å. For comparison, previous simulations of gram-negative bacteria membranes are listed in Table 38. Compared with all the *E. coli* models,[10] the inner membrane of *P. aeruginosa* have higher thickness. Several factors may contribute to this difference.

The most obvious one is the length of the fatty acid chain. PMSPE and PMSPG have two more carbons in their *sn*-2 tails compared to PMPE (PE 16:0-17:0 cyc (9,10)) and PMPG (PG 16:0-17:0 cyc (9,10)), which consist of a large fraction in the *E. coli* membrane. Second, the *E. coli* membrane models have more unsaturated lipid tails containing sixteen carbons, while most unsaturated lipids in *P. aeruginosa* membrane models are POPG (16:0-18:1-PG) and POPE (16:0-18:1-PE). The PE: PG ratio also plays an important role in membrane thickness. As shown in Figure 41, in a specific bilayer system, PG lipids penetrate more into the aqueous phase, while remaining similar inter-leaflet interdigitation compared to PE lipids. Thus, PG lipids can increase the difference between the headgroup-to-headgroup distance and the overall thickness. This is further justified by the four controlled simulations in which the percentage of PG lipids increases from 0 to 100% (Table 36). As for Chlamydia, the ordering effect of cholesterol can significantly reduce the A_l,[300] causing a higher thickness even though the fatty acids are on average shorter than those in the *P. aeruginosa* inner PM.

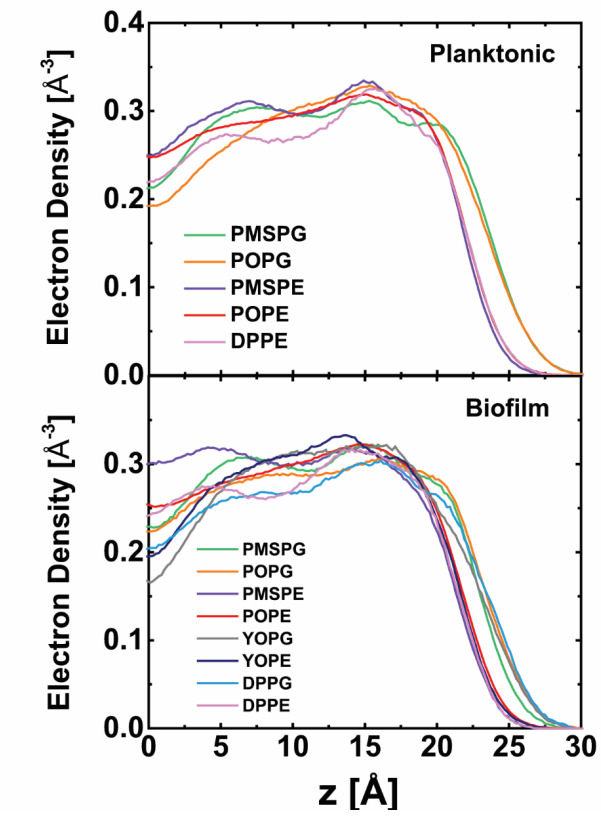

Figure 41. Electron Density Profile for each lipid type. The densities are normalized using lipid numbers so that EDP for each lipid type looks like the EDP for a single component bilayer.

Table 38. Thicknesses for various bacterial membrane models. D_{HH} is the headgroup-to-headgroup distance, D_B is the overall bilayer thickness, $2D_C$ is the hydrophobic thickness.

Model	D_{HH} (Å)	D_B (Å)	$2D_C$ (Å)
Planktonic	41.1 ± 0.2	39.9 ± 0.2	30.4 ± 0.1
Biofilm	40.3 ± 0.1	38.9 ± 0.3	29.7 ± 0.3
E. coli Early-log [10]	37.7 ± 0.1	38.3 ± 0.1	30.6 ±0.1
E. coli Mid-log [10]	38.0 ± 0.1	38.2 ± 0.1	30.4 ± 0.1
E. coli Stationary [10]	37.3 ± 0.1	37.1 ± 0.1	30.00 ± 0.03
E. coli Overnight [10]	37.0 ± 0.2	36.8 ± 0.1	30.1 ± 0.1
E. coli Top6[9]		37.3 ± 0.2	29.8 ± 0.1
Chlamydia EB [301]	44.1*		
Chlamydia RB [301]	42.3*		

* no errors provided

5.2. Models for *Arabidopsis thaliana* plasma membrane

5.1.1. Introduction

Arabidopsis thaliana, a small flowering plant, was widely used as a model organism in the past half-century to study plant biology and beyond, and the pioneers to study it can be traced back to the 1930s.[302] During the 1970s, suspicion about using it as a model organism was raised because it seemed to resist most initial attempts to proliferate and regenerate in culture and its small-sized chromosomes.[302] However, with the emergence of efficient tissue culture methods[303-305] and alternative protocols,[306] and the realization in the field of the advantages of its small-sized chromosomes,[307] it quickly regained the focus of biologists. Taking gene sequencing as an example, the latest annotation of the *Arabidopsis* genome identified 2655 protein-

coding and 5178 non-protein coding genes.[308] Moreover, comprehensive studies have been carried out by research groups from all over the world and results have been assembled into books[309] and a free-access database.[310] Studies using *Arabidopsis* have also played a leading role in basic biological discoveries,[308, 311] such as proteins identified as human orthologs in the innate immune system,[312] the role of auxin on the ubiquitin pathway conserved among eukaryotes,[313] and a light signaling component COP1 whose mammalian orthologs has a role in tumorigenesis.[314]

While most molecular biologists studying *Arabidopsis* have focused on its nucleic acids (chromosomes) and proteins, its membrane composition and chemical physics have not been explored thoroughly. In fact, the plasma membrane (PM), in particular, plays a critical role in plant physiological processes including growth and development, ion and metabolite transport, perception of environmental changes, and disease resistance.[315] PM of plants is composed of three major lipid types, which are sterols, glycerolipids and sphingolipids.[2] Sterols are essential for plant development, particularly embryogenesis, cell elongation, and vascular differentiation.[316] They can also act as structural lipids and induce the liquid-ordered phase in membrane.[2] The plant PM also has a high sphingolipid composition. In tobacco leaves, for example, study has shown that glycosylinositolphosphoceramides (GIPCs) may represent up to 30-40 mol % of the PM.[317] Sphingolipids are engaged in lipid raft formation in conjunction with sterols.[318] These rafts are sometimes referred to as detergent insoluble domains (DIM) by biologist and have putative functions in plant cell signaling in cooperation with their associated proteins.[318-320] Glycerolipids, primarily phospholipids, are another major lipid type in plant PM. Historically, they are classified

by abundance into structural lipids and signaling lipids, though more recent studies have shown that some "structural lipids" are also involved in signal-transducing.[2] The dominating phospholipids in plant PM are PCs and PEs, which represent up to 68–80% of the structural phospholipids. The remainder consists of PG, PI, PS, and phosphatidic acid PA.[2]

In terms of lipid distribution, the PM is asymmetric with the outer leaflet containing a fair portion of sphingolipids, while the inner leaflet consists nearly exclusively of sterols and phospholipids.[321] In this chapter, an asymmetric membrane model is built to represent the PM of *Arabidopsis thaliana*. The model uses five different phospholipids which are representative of the major types found in the PM of *Arabidopsis*. Two sphingolipids are considered and only used in the apoplastic (outer) leaflet of the membrane. Sitosterol is used as the only sterol in the model and sterol derivatives are not considered due to their minor amounts.[322] Since the component area of these lipids are not known beforehand, two symmetric models representing the cytosolic (inner) and outer leaflets are built and simulated until reaching equilibrium before they are merged to form the asymmetric model. The analysis of the simulations covers overall structure of the membrane (A_l, S_{CD}), mechanical properties of the membrane (K_A and pressure profile), and lipid-lipid interaction (clustering). Hidden Markov Modeling (HMM) is utilized to further study lipid correlation and compositional phase separation (potential raft formation).

The purpose of this study is to provide readily useful models of plant membrane for future study (for example, auxin partitioning). Although the membranes are modest

in size, they can be easily expanded to simulate large bilayers with proteins or other biologically relevant molecules.

5.1.2. Methods

The relative proportions of sphingolipids, glycerophospholipids and sterols (free and conjugated) in the PM of *Arabidopsis thaliana* leaves were determined by Uemura et al.[323] to be 7, 47, 46%, respectively. The same study was used to inform the overall glycerophospholipids composition, from which five major ones were selected (see Table 39). While the leaflet asymmetry has not been determined by any experiment at this point to the best of the author's knowledge, an asymmetric distribution of glycerophospholipids between the two leaflets that is similar to other eukaryotic cells was proposed, where the cytosolic leaflet is enriched in PS and PI.[1] Since PG is also negatively charged, it is reasonable to deduce it mainly resides in the cytosolic leaflet. To make the model representative of the real membrane and to facilitate computational analysis, PS, PI and PG lipids were excluded from the apoplastic leaflet and only incorporated into the cytosolic leaflet. The ratio between different glycerolipids were kept as close as possible to the original values reported by Uemura et al.[323] For all glycerophospholipids, only the predominant species were used based on Uemura et al.,[323] which can be found in Table 39.

Table 39. Lipid Definition Abbreviations for glycerophospholipids are consistent with naming in the CHARMM-GUI *Membrane Builder*.

lipid name	abbrev.	*sn*-1/ fatty acid chain	*sn*-2/ sphingosine chain
1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine	PLPC	16:0	18:2
1,2-dilinoleoyl-sn-glycero-3-phosphoethanolamine	DLiPE	18:2	18:2
1-palmitoyl-2-(α-linolenoyl)-sn-glycero-3-phospho-(1'-rac-glycerol)	PNPG	16:0	18:3
1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphoinositol	PLPI	16:0	18:2
1,2-dilinoleoyl-sn-glycero-3-phospho-L-serine	DLiPS	18:2	18:2
D-glucose-(α1→4)-D-glucuronic acid-(α1→2)-inositol-phosphoryl-(N-palmitoyl-4R-hydroxy-D-erythro-sphingosine)	GIPC	16:0	18:0
β-D-glucosyl-N-palmitoyl-4R-hydroxy-D-erythro-sphingosine	GluCer	16:0	18:1
sitosterol	SITO		

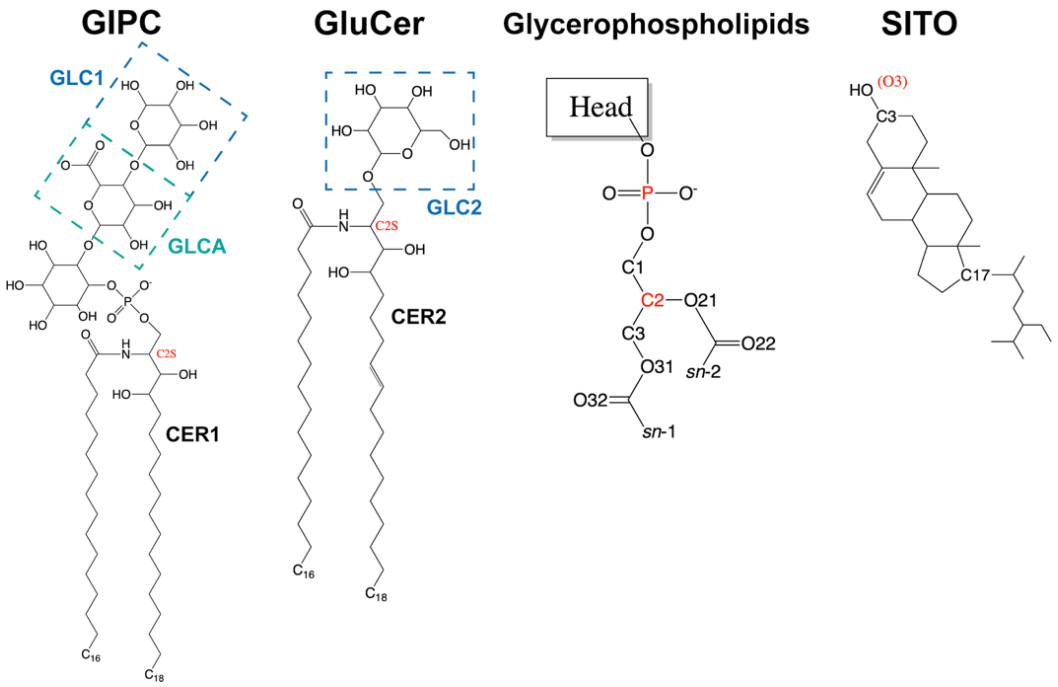

Figure 42. Chemical structures of lipid types involved in this study.

It is well known that the predominant sterols in plants are β-sitosterol (sitosterol hereafter), stigmasterol, and campesterol,[324] and Grosjean et al.[325] observed an almost even distribution of these species in tobacco suspension cells. Since these sterols are similar to each other in structure, only sitosterol was used in the model where it accounts for 50% of the outer leaflet and 42% of the inner leaflet.

Additional caution was imposed when determining the molecule species of sphingolipids and their precise structure. It has been reported by Markham et al.[326] that GIPC, glucosylceramide (GluCer), and ceramides accounts for 64, 34, and 2% of the total sphingolipids, so that ceramides were excluded from the model and the molar ratio of GluCer (1): GIPC (2) was used. The chemical structure of the predominant GIPC was also determined in the same study to be hexose-hexuronic-inositol-phosphoceramide for *Arabidopsis thaliana*. The linkage conformations between the six-membered rings determined in two other publications[327, 328] were used (see Table 39 for details). In terms of tails, long chain base (LCB) of eighteen carbon-long bearing two hydroxyl groups at carbon 3 and 4 has been identified as the predominant sphingosine chain, and a double bond may appear at carbon 8.[323, 326, 329] The fatty acid was determined to be mostly 16:0 by Uemura et al.[323] Within the PM of eukaryotic cells, sphingolipids are primarily located in the outer leaflet, so that they were excluded from the inner leaflet in the model. Table 40 lists the lipid composition for each model.

Table 40. Model Composition.

lipid	inner	outer	plant inner	plant1 outer	plant2 outer	plant3 outer
SITO	25	30	25	33	38	33
PLPC	3	17	3	18	15	19
DLiPE	12	5	12	5	5	6
PNPG	7		7			
PLPI	8		8			
DLiPS	5		5			
GluCer		3		3	3	4
GIPC		6		7	6	4
Total	60	61	60	66	67	66

The inner and outer models were constructed using the CHARMM-GUI *Membrane Builder*. For lipids not included in the CHARMM-GUI library, similar lipids were used as placeholders and mutated afterwards using CHARMM.[98] A hydration number of 35 water molecules per lipid was used for the inner model and 50 was used for the outer model. NAMD[330] was used to perform the equilibrium and production runs for the two symmetric models (3 replicas for each) with the C36 lipid force field[19] and the modified TIP3 water model.[58, 59] The total area of the simulation box tangential to the membrane surface (the *x*-*y* plane henceforce) and the lipid-wise (component) area for each lipid species were computed after the systems reached equilibrium. The asymmetric membranes were constructed by taking one leaflet (including the water molecules bound to that leaflet when the system is recentered around the center of the bilayer) from each model at the end of the simulation and joining them using an area-match method through CHARMM. Since the simulated total area of the inner model was larger than the outer model, the outer model was expanded in the *x*-*y* plane using periodic images and the excess lipids were selectively cutoff to make the sum of the total area of the primary cell and the lipid-wise areas from the

images as closest to the total area of the inner model as possible. The resulting lipid composition for each outer leaflet in the asymmetric model are listed in Table 40, while the inner leaflet has the same composition as the symmetric model. The different lipid compositions for the outer leaflet come from the randomness of lipid placement at the end of simulation for the outer model and they mimic the natural fluctuation of local lipid content in a realistic membrane.

The same software and force field was used to simulate the asymmetric membranes. The simulation settings were the same as those used for simulating the *P. aeruginosa* models. The production run was 400 ns for each replica of the inner model and 500 ns for the outer model. For each of the three asymmetric models, two replicas were simulated for 500 ns.

Analysis for all model membranes were based on the equilibrated portion of the simulation, which was identified through the time evolutions of surface area per lipid and sterol-sterol radial distribution function. Other properties presented include component surface area, K_A, S_{CD}, membrane thickness, 2D-RDF between lipids, lipid clustering, compositional phase separation, and pressure profile, while more analysis can be found in the published paper.[331] The 2D-RDF in the *x*-*y* plane was calculated based on the coordinates of representative atoms defined for each lipid species, which can be found in Table 41. Lateral lipid clustering was computed using the density-based spatial clustering of applications with noise (DBSCAN) algorithm implemented in scikit-learn.[332, 333] While the cutoff distance in DBSCAN is a uniform value specified by the user, lipid-lipid distances from the trajectory can be scaled to reflect the different sizes of various lipids. This scaling was made possible by analyzing the 2D-RDFs

between different lipid species and was shown to provide a more robust, and perhaps more importantly, an unbiased analysis of lipid clustering. More information about this will be provided in Section 5.2.3.

Table 41. Representative atoms used for the 2D-RDF calculation as noted in Figure 42.

lipid	representative atom
PLPC, DLiPE	P
GIPC, GluCer	C2S
SITO	O3

Compositional phase separation was analyzed using the method developed by Sodt et al.,[269] which utilized a hidden Markov model (HMM) to detect the "hidden" states for individual lipids. In this method, the local composition in the vicinity of each lipid determined by the six nearest lipids (including itself) is used as the emission signals. Assuming the two "hidden" states are the liquid-ordered (L_o) and the liquid-disordered (L_d) phases, it is reasonable to group the lipids into three major types, i.e., sitosterol, lipids with one or less double bonds (GIPC and GluCer), and lipids with two or more double bonds (all other lipids). The local composition is then encoded with the number of each lipid type, leading to 28 different local compositions. Since there are 2 transition probabilities and 28 emission probabilities for each state, the total number of independent probabilities are 56. These probabilities are optimized by the Baum−Welch algorithm,[334, 335] which is an iterative procedure to find a more-likely model based on some initial guesses of the probabilities. Once the probabilities are determined, the most likely state sequence is determined by the Viterbi algorithm.[336]

The sampling frequency for this analysis (0.2 ns) is much lower than the other analysis considering lipid diffusion is at the order of 10^{-7} cm^2/sec (thus it only covers an area of 2.4 $Å^2$ during 0.2 ns, on average), and this frequency is close to what was used by Sodt et al. (0.239 ns).

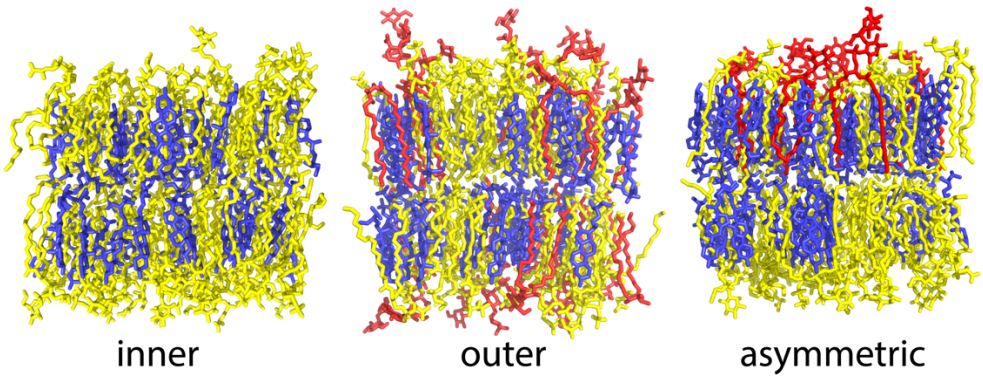

Figure 43. Example snapshots at the end of simulations. Yellow: glycerophospholipids; red: glycolipids; blue: sitosterols.

Pressure profiles were generated by restarting the simulations in CHARMM from checkpoints containing both coordinates and velocities. A total of 100 checkpoints (every 2 ns from the last 200 ns of the NAMD simulation) were used for each inner membrane replica, and a total of 125 checkpoints (every 2 ns from the last 250 ns of the NAMD simulation) were used for each outer membrane replica and each asymmetric membrane replica. A time step of 1 fs was used for these simulations and the simulation length is 0.1 ns for each checkpoint. The local pressure was calculated and recorded at a frequency of 100 timesteps. While the total surface tension is zero in an NPT simulation, the contribution from each leaflet can be calculated by

$$\gamma_{lower} = \int_{-\frac{L}{2}}^{0}(P_N - P_T(z))dz \quad \text{or} \quad \gamma_{upper} = \int_{0}^{\frac{L}{2}}(P_N - P_T(z))dz \qquad (28)$$

where P_N (1 atm) and P_T are the components of the pressure tensor normal and tangential to the bilayer surface. Previous work has shown that P_T exhibits huge oscillations between negative values at the water/hydrocarbon interface and positive values at the bilayer center.[108] One can also calculate the first derivative of the bending free energy per lipid at zero curvature for each leaflet, $\bar{F}'(0)$, by

$$\bar{F}'(0) = \frac{d\bar{F}(c)}{dc}|_{c=0} = -\int_0^{\frac{L}{2}} z[P_T(z) - P_N]\, dz = -K_c^m c_0 \qquad (29)$$

where c stands for the curvature, K_c^m is the monolayer (leaflet) bending constant, and c_0 is the spontaneous curvature of the monolayer (leaflet). This integral provides the curvature tendency of the leaflet, although the extraction of the exact value of c_0 needs K_c^m to be determined, which is beyond the scope of this work.

5.2.3. Results

In this section, results will be presented for all the membrane models. Since the three asymmetric models are slightly different from each other in terms of their outer leaflet composition, we will use “plant 1”, “plant 2” and “plant 3” to distinguish them in figures and tables. For all simulations, flip-flop of sitosterol was ruled out by checking the positions of the O3 atoms using the whole trajectory recentered around the bilayer center. When the average of the three asymmetric models is presented, the label “plant” will be used. The total surface area and the sitosterol 2D-RDF were used to estimate the starting points of equilibrium. The starting points of area equilibrium, sample plots of total area versus time, and sample plots of block averaged 2D-RDF for sitosterol can be found in the published paper.[331] Based on these results, the last 200 ns

of the simulation data of the inner membrane model, and the last 250 ns of the simulation data of the outer membrane model and asymmetric membrane models were used for the analyses.

A_l is reported in Table 42 and the A_l is 48.35 ± 0.01 Å^2 for the symmetric inner model and 43.79 ± 0.02 Å^2 for the symmetric outer model. This difference is majorly caused by the higher sitosterol concentration and the inclusion of GluCer and GIPC in the outer model. While the component areas (Table 43) for these lipids are lower than the glycerophospholipids, the higher concentration of sitosterol and saturated tails also produced an ordering effect, which slightly reduced the component areas for the other lipids in the system. In the asymmetric models, A_l for each leaflet roughly matches its corresponding symmetric model. Plant-2 has a lower A_l compared to the other two asymmetric models, especially for the outer leaflet. This likely originated from the higher number of sitosterols in the outer leaflet. It is interesting that the A_l for the inner leaflet (which corresponds to the inner model) in all the asymmetric models are significantly lower than the A_l for the symmetric inner model. While a possible origin is a small area mismatch between the two leaflets when they were joined together, a much lower A_l for the lower leaflet in plant-2 indicates the higher sterol concentration in the outer leaflet might also modulate the A_l of the opposite leaflet. A previous simulation study of soybean hypocotyl and root plasma membranes[81] generated A_l of 52.7 ± 0.2 Å^2 and 51.9 ± 0.1 Å^2, respectively, which are larger than the A_l for both leaflets of the *A. thaliana* plasma membrane. This is expected since the soybean membrane models contain more unsaturated tails and less sterols.

Table 42. A_l and Compressibility Modulus (K_A). Errors for the inner/outer symmetric models are standard errors based on three replicas. Average K_A based on all three asymmetric models is reported with standard error from all six replicas to reduce statistical error.

model	A_{lip} outer leaflet (Å^2)	A_{lip} inner leaflet (Å^2)	K_A (N/m)
inner	48.35 ± 0.01		0.68 ± 0.05
outer	43.79 ± 0.02		0.99 ± 0.05
plant 1	43.66 ± 0.03	48.03 ± 0.03	
plant 2	42.53 ± 0.04	47.50 ± 0.04	0.94 ± 0.04
plant 3	43.67 ± 0.04	48.03 ± 0.04	

K_A for all model membranes are listed in Table 42. K_A for the inner model is significantly lower than the outer model, although both are remarkably higher than a single-component glycerophospholipid bilayer,[108] reinforcing the role of sterols in the modulation of membrane stiffness.[15, 337] K_A for the asymmetric model is closer to the outer model, indicating the area compressibility is non-additive and biased to the leaflet with more compression resistance.

Table 43. Component area (Å^2).

lipid	inner	outer	plant inner	plant outer
SITO	31.3 ± 0.4	29.6 ± 0.3	31.3 ± 0.6	30.1 ± 0.3
PLPC	60.1 ± 1.4	58.6 ± 0.6	59.4 ± 2.4	59.3 ± 0.7
DLiPE	60.0 ± 0.9	58.4 ± 1.2	59.9 ± 1.1	59.4 ± 1.8
PNPG	61.9 ± 1.1		60.4 ± 1.6	
PLPI	60.1 ± 1.1		60.8 ± 1.9	
DLIPS	60.4 ± 1.6		60.8 ± 1.4	
GluCer		53.3 ± 1.1		53.1 ± 1.7
GIPC		54.1 ± 0.9		53.9 ± 1.3

Membrane thicknesses computed based on the EDPs are listed in Table 44. Here, only the averages of the asymmetric models are reported. D_{HH} for the three

different systems are statistically the same. The overall thickness, D_B, is highest in the outer model and lowest in the inner model. Normally D_B can be used to reflect the degree of water penetration and a higher D_B means less water penetration. However, in our case, the major reason for the higher D_B in the outer model is the glycolipids taking up extra space beyond the membrane surface. The hydrophobic thickness, $2D_C$, is also higher in the outer model, probably caused by the higher percentage of saturated tails and sitosterol. For the asymmetric model (avg.), D_B and $2D_C$ are approximately the average of the inner and outer model. Compared with the soybean model membranes,[81] the asymmetric membrane is slightly thicker, perhaps as a consequence of more saturated tails and higher sterol content.

Table 44. Membrane thicknesses for model membranes. Standard errors are based on replicas.

model	D_B (Å)	D_{HH} (Å)	$2D_C$ (Å)
inner	40.1 ± 0.1	43.5 ± 0.1	33.4 ± 0.1
outer	41.6 ± 0.4	42.8 ± 0.5	34.8 ± 0.2
asymmetric (avg.)	40.8 ± 0.2	43.6 ± 0.4	34.3 ± 0.1

Figure 44 plots the S_{CD} for DLiPE and PLPC. DLiPE has two tails with double bonds at carbons 9/10 and 12/13, so that the S_{CD} dropped twice around that region for both tails. PLPC has only one tail (*sn*-2) containing double bonds at carbons 9/10 and 12/13, thus a similar double-dip is observed. Lipid tails are more ordered in the outer model or the outer leaflet of the asymmetric model, especially for the *sn*-1 chain of PLPC. This is expected because there are more saturated tails and sitosterols in the outer model/leaflet. The ordering effect of GIPC and phytosterols on glycerophospholipids were studied in a recent experiment by Cassim et al.,[338] where

they measured the ^{2}H-NMR powder spectra of deuterated palmitoyl-oleoyl phosphatidylcholine containing 31 atoms of deuterium on the palmitoyl chain (POPC-(2)H_{31}) mixed with GIPC and/or phytosterols (sitosterol or stigmasterol). Adding phytosterols to the POPC-(2)H_{31} sample led the gel-to-fluid phase transition of the POPC bilayer at (−2.5 ± 2.4) °C to be abolished and highly ordered aliphatic chain was seen for all measured temperatures (-10 to 40 °C). By averaging the NMR data at 20 °C and 30 °C, a S_{CD} of 0.34 was obtained for the palmitoyl chain in a membrane composed of sitosterol and POPC-(2)H_{31} (1:1 mol). The ternary system composed of GIPC/sitosterol/POPC-(2)H_{31} (1:1:1 mol) also demonstrated a rigidifying effect of GIPC and sitosterol, with an estimated S_{CD} of 0.26 at 25 °C. Since the S_{CD} estimated from the NMR experiment should be interpreted as the average of carbon positions having significantly higher values compared to others, average S_{CD} of carbon 4 to carbon 11 of the palmitoyl (*sn*-1) chain (PLPC) in our models were computed. These averages are 0.30 for the inner model, 0.31 for the inner leaflet of the asymmetric model, and 0.36 for both the outer model and the outer leaflet of the asymmetric model, which compare favorably with the experiment considering the presence of other lipids and the different proportion of GIPC/sitosterol.

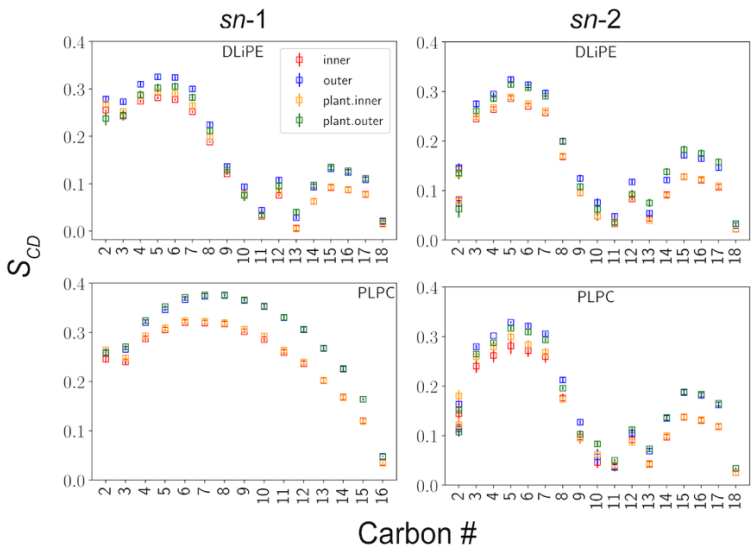

Figure 44. S_{CD} profiles for DLiPE and PLPC. Errors for the symmetric models are based on three replicas; errors for the symmetric model (plant inner and plant outer) are based on all three models listed in Table 40 (6 replicas in total).

Two-dimensional radial distribution functions (2D-RDFs) were generated using representative atoms (Figure 36 and Table 41). These are the phosphorus atoms of the glycerophospholipids, the center carbon atoms in the amino alcohol of the glycolipids, and the oxygen atoms of sitosterols. Figure 45 presents the interactions between two major lipid types – sitosterol and glycolipid. There is a strong peak for the sitosterol-sitosterol 2D-RDF in all the model systems, perhaps due to the high concentration and the rigid structure of sitosterol. Despite of the higher sitosterol concentration in the outer model or the outer leaflet of in the asymmetric model, the height of this peak is lower compared to the inner model or the inner leaflet. This can be explained by the presence of glycolipids. Sitosterol shows preferential association with GluCer and GIPC at short range, and this might have diminished the sitosterol-sitosterol association. Repulsion between glycolipids were noticed as shown in Figure

45. This is likely caused by the negative net charge on GIPC (-2 *e*) as well as the steric effect of its bulky head.

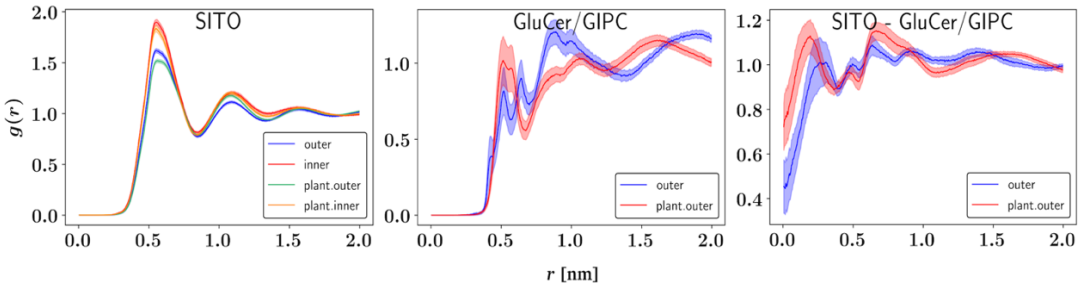

Figure 45. 2D-RDFs for selected lipid pairs. SITO: between any two sitosterols; GluCer/GIPC: between any two glycolipids; SITO – GluCer/GIPC: between sitosterols and glycolipids.

Analysis of lipid clustering and compositional phase separation can provide information about lateral lipid organization within a leaflet. Clustering analysis was made possible by the DBSCAN algorithm, which uses distances between points of interest to compute the clusters. Here, the points are the same representative atoms used to compute the 2D-RDFs. In this set of analysis, two distance cutoff schemes were utilized. In the first, a uniform cutoff distance (6 Å) between lipids was used for all lipid pairs, regardless of their sizes. In the second, the distance was scaled by the average distance of two neighboring lipids of specific types estimated from the 2D-RDFs. This led to different cutoffs for different lipid pairs. A lower cutoff means the two lipids need to be closer to form a cluster. Table 45 reports the tendency to form clusters for each lipid type. The different cutoff schemes did not change the tendency significantly for most lipids except sitosterol. This is because the modifications to the pairwise distance were mostly related to sitosterol and were all in the direction of

decreasing the cutoff. The tendency for DLiPE increases when the pairwise distances are used, because it is a potent hydrogen bond donor. Moving from the symmetric membranes to the asymmetric membrane did not change the tendency to form clusters for all lipids, which means there was little inter-leaflet coupling in terms of clustering.

Table 45. Lipid % in cluster using different D_{cut} scheme. The minimum number of lipids to form a cluster is set to be 4. Membrane composition (%) also provided for comparison.

lipid	SITO	PLPC	DLIPE	PNPG	PLPI	DLIPS	GIPC	GluCer
	inner							
composition	42	5	20	12	13	8		
uniform D_{cut}	49 ± 1	4 ± 1	19 ± 1	9 ± 1	12 ± 1	7 ± 1		
customized D_{cut}	42 ± 1	4 ± 1	24 ± 1	8 ± 1	13 ± 1	9 ± 1		
	outer							
composition	49	28	8				10	5
uniform D_{cut}	57 ± 2	22 ± 1	7 ± 1				9 ± 1	5 ± 1
customized D_{cut}	53 ± 2	24 ± 1	9 ± 1				9 ± 1	5 ± 2
	plant inner							
composition	42	5	20	12	13	8		
uniform D_{cut}	48 ± 1	4 ± 1	19 ± 1	9 ± 1	12 ± 1	7 ± 1		
customized D_{cut}	42 ± 1	4 ± 1	24 ± 1	8 ± 1	13 ± 1	9 ± 1		
	plant outer							
Composition (avg.)	52	26	8				9	5
uniform D_{cut}	57 ± 1	22 ± 1	7 ± 1				9 ± 1	5 ± 1
customized D_{cut}	53 ± 2	24 ± 1	9 ± 1				9 ± 1	5 ± 1

Clustering analysis can also be utilized to study the affinity between lipids. Table 46 presents the propensity of two lipids to coexist in a cluster, with a focus on sitosterol and GluCer/GIPC. These numbers measure the probability to find the second lipid (the accompanying lipid) in the cluster where the first lipid (the kernel lipid) is present. Percentage of the accompanying lipid in the composition is also provided for comparison. When the kernel lipid is sitosterol, the percentages of other lipids are not significantly modified compared to the composition, though a slight decrease for

CluCer/GIPC is observed. When the kernel lipid is GluCer/GIPC, sitosterol is more favorable than the other lipids and GluCer/GIPC is less favorable, which is consistent with the repulsion between the glycolipids discovered by the 2D-RDF.

Table 46. Percentages of accompanying lipids in cluster where a kernel lipid is present. The minimum number of lipids to form a cluster is set to 4.

model/leaflet	kernel lipid	accompanying lipid(s)	composition (%)	same cluster (%)
inner	SITO	SITO	42	43 ± 1
outer	SITO	SITO	49	51 ± 1
	SITO	GluCer/GIPC	15	12 ± 1
	GluCer/GIPC	SITO	49	55 ± 1
	GluCer/GIPC	GluCer/GIPC	15	8 ± 1
plant inner	SITO	SITO	42	42 ± 1
plant outer	SITO	SITO	52	54 ± 1
	SITO	GluCer/GIPC	14	11 ± 1
	GluCer/GIPC	SITO	52	57 ± 3
	GluCer/GIPC	GluCer/GIPC	14	9 ± 2

HMM was used to assign states to individual lipids based on the local composition. The two putative states are the L_o and L_d phases, considering experimental observations and simulation results of similar systems.[269, 318] The representative atoms to locate a lipid in the local composition calculation are same as those used in the 2D-RDF and clustering analyses. In this analysis, we try to answer two questions: (1) Does certain type of lipid prefer to stay in one state (phase)? (2) Does HMM provide additional or different information regarding lipid organization compared to the clustering analysis? Since the focus here is potential phase separation, only the outer model and the outer leaflet from the asymmetric membrane were investigated because they contain both saturated and unsaturated tails.

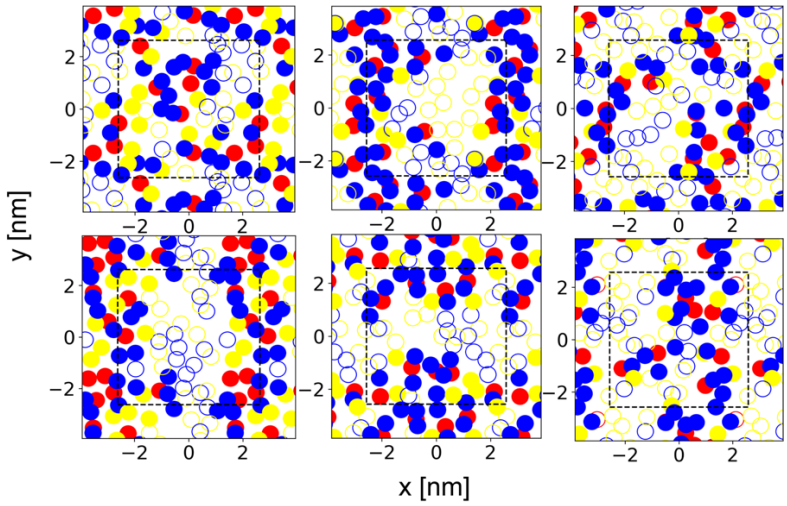

Figure 46. End-of-simulation states for lipids in the upper leaflet (upper row) and the lower leaflet (lower row) for the outer model. Each column represents one replica. Glycolipids in red, glycerophospholipids in yellow, and sitosterol in blue. Lipids in the glycolipid-enriched state are filled. The center boxes indicate the boundaries of the simulation cells.

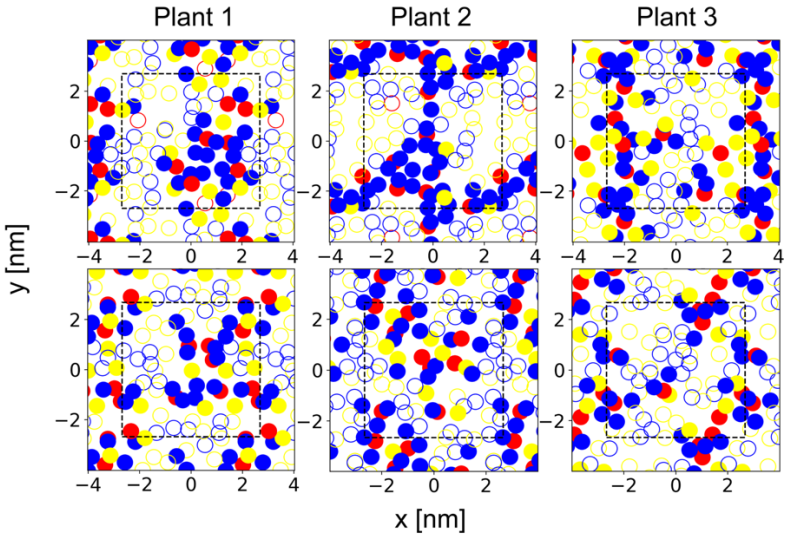

Figure 47. End-of-simulation states for lipids in the upper leaflet for the asymmetric model. Each row represents one replica. Glycolipids in red, glycerophospholipids in yellow, and sitosterol in blue. Lipids in the glycolipid-enriched state are filled. The center boxes indicate the boundaries of the simulation cells.

End-of-simulation states for all replicas and leaflets are plotted in Figures 46 and 47. The glycolipid-enriched state is filled with color in these subplots. It is interesting that a stripe-like pattern is found for almost all the replicas/leaflets of the outer model except the upper leaflet from replica 1. In the asymmetric membrane, this pattern is maintained. Besides, the glycolipid-enriched state also contains much more sitosterols than glycerophospholipids. To quantify this differentiation, the percentage of each lipid type that appeared in the glycolipid-enriched state was calculated (Table 47). On average, for the outer model, 98.0 % of the glycolipids and 58.4 % of the sitosterol appeared in the glycolipid-enriched state, and it is only 41.2 % for glycerophospholipids. These averages are 94.9 %, 54.7 % and 37.0 % for the asymmetric models, respectively. It should be noted that the stripe-like pattern was not observed when visualizing lipid clustering (not shown). This is not surprising because different metrics were used. In the clustering analysis, the cutoff distances determined the assignment of state; in the phase separation analysis, however, the local composition played an important role. With glycolipids being less populated in the model, using local composition as the metric could potentially introduce bias into HMM so that the lipids around glycolipids are more likely to be assigned to one state.

Table 47. Percentage of lipid in the glycolipid-abundant phase in the outer model (outer) and the outer leaflets of the asymmetric models (asym.)

lipid	replica 1 (outer)	replica 2 (outer)	replica 3 (outer)	model 1 (asym.)	model 2 (asym.)	model 3 (asym.)
glycerophospholipids	46.8 %	41.9 %	35.1 %	37.0 %	34.2 %	53.8 %
sitosterol	55.1 %	62.6 %	57.4 %	52.2 %	58.1 %	39.8 %
GIPC/GluCer	100 %	98.3 %	95.6 %	95.0 %	89.8 %	100 %

The tangential (to the bilayer normal) component of the pressure tensor is shown in Figure 48, which exhibits huge oscillations between positive and negative values. Compared to the pressure profiles obtained for pure phosphoglycerolipids,[108] there are additional peaks at $z = \sim \pm 10$ Å, which might be caused by sitosterols. The outer membrane has larger overall oscillations compared to the inner membrane, which is consistent with the larger compressibility modulus it exhibits. In the asymmetric membrane, pressure profile for each leaflet follows their corresponding symmetric models, though slightly modulated by the opposite leaflet.

For the asymmetric membrane, the surface tension of each leaflet was computed based on Eq. 28.The average surface tension for the inner leaflet is 6.0 ± 2.0 dyn/cm and the average for the outer leaflet is -6.0 ± 2 dyn/cm, which are small compared to the compressibility modulus, indicating a good match between leaflet areas. $K_c^m c_0$ is -0.15 ± 0.03 kcal/mol/Å for each leaflet in the inner model and -0.30 ± 0.03 kcal/mol/Å in the outer model, both indicating a positive spontaneous curvature of the monolayer. These numbers did not change drastically in the asymmetric model, as $K_c^m c_0$ values for the inner and outer leaflet are -0.21 ± 0.03 kcal/mol/Å and -0.29 ± 0.04 kcal/mol/Å, respectively.

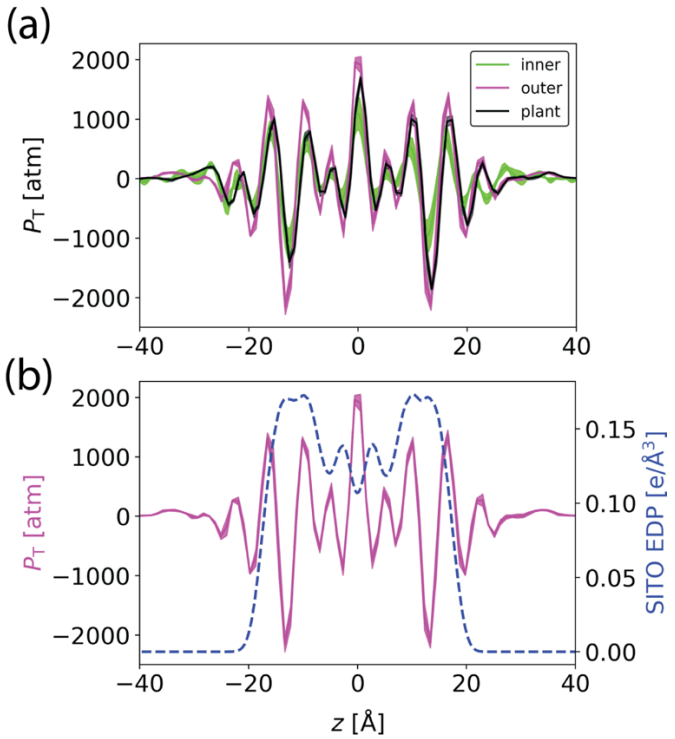

Figure 48. (a) the tangential pressure profile for the three model membranes. (b) The tangential pressure profile for the outer model aligned with the sitosterol EDP.

5.2.4. Discussion

In this study, the *Arabidopsis thaliana* plasma membrane was modeled using the CHARMM36 all-atom lipid force field. Two symmetric bilayer models representing the inner and outer leaflets of the membrane were simulated before merged to form the asymmetric membrane model. The major difference between the two leaflets in our model is that the outer leaflet contains glycolipids while the inner leaflet contains more glycerophospholipids but no glycolipid. The lipid composition was primarily determined based on Uemura et al.,[323] where the concentration of sphingolipids was determined to be 7 mol%. It should be noted that the proportion of sphingolipids varies among different plants. For example, Cacas et al.[317] measured GIPCs representing up to 40 mol % of total tobacco PM lipids, a number much higher

than *Arabidopsis thaliana* reported by Uemura et al. Apart from the species gap which must have contributed to this difference, the lipid extraction method used by Uemura et al. might have caused incomplete solubilization of GIPCs as pointed out later by Markham et al.[326] Despite these uncertainties from experiments, MD simulation can provide valuable insights about how different lipid types interact with each other and how the lipid composition influences membrane properties.

Due to the existence of sitosterol, pressure profiles for the model membranes exhibit more oscillations at z = ~ ± 10 Å compared to bilayers composed of pure glycerophospholipids or sphingolipids.[108] As discussed by Ollila et al.,[339] these oscillations are likely utilized by the membrane to modulate the structure of membrane proteins since the associated energetics is at the order of a few k_BT. The calculated spontaneous curvature is negative for both the inner and outer leaflets, though the exact numbers were not determined. While a direct comparison with experiment or other simulation study is not available, this result is qualitatively in agreement with the result for DOPC (80%)/CHOL (20%) bilayer simulated by Ollila et al.[339] using a united-atom force field. Using X-ray diffraction, Chen et al.[340] observed that 5 mol% cholesterol in DOPC induced the inverted hexagonal (H_{II}) phases of very large dimension, and that increasing the level of cholesterol resulted in reduced lattice dimensions hence more negative spontaneous curvatures. In addition, using small-angle X-ray scattering (SAXS), Kollmitzer et al.[341] measured the (partial) spontaneous curvature of cholesterol in the H_{II} phase to be -4.94 ± 0.13 Å, which was significantly more negative compared to PC, PE and sphingolipid. These results indicate that sterols may also introduce negative spontaneous curvature in a lamellar phase. Since the asymmetric

membrane studied here is relatively rigid so that a high K_c^m is expected, which means a small $|c_0|$. Even though, one might expect the net spontaneous curvature to be positive in order to form a cell. One reasonable argument is that membrane proteins can modulate the curvature of the bilayer; another possible reason is that the percentage of GIPC in the outer leaflet might be higher than what Uemura et al.[323] measured, which likely has a positive (partial) spontaneous curvature because $K_c^m c_0$ for palmitoylsphingomyelin (PSM) is positive in the lamellar phase.[108]

The DBSCAN algorithm with pairwise cutoffs was used to study lipid clustering, which showed stability against different settings. While there was no lipid type showing significantly higher probability to form clusters, additional analysis has shown that sitosterol favors glycolipid-containing clusters. This is consistent with the 2D-RDFs, where sitosterol showed high affinity with glycolipids. HMM was used to detect potential compositional phase separation by grouping all the lipids into three major types (i.e., glycerophospholipids, glycolipids and sitosterol). Through this analysis, sitosterols were found to prefer the glycolipid-enriched state, which agrees with the 2D-RDFs. However, the high probability (> 90 %) for glycolipids to be in the glycolipid-enriched state is concerning. Because of the small number of glycolipids in the model, one can imagine that lipids close to any glycolipid would have a much higher chance to be assigned to the glycolipid-enriched state. This could have biased the HMM process such that the glycolipid positions become definitive and other information is overlooked. Nevertheless, the affinity between sitosterol and glycolipids support a potential phase separation which could eventually lead to lipid rafts. Interestingly,

glycolipids have unfavorable interactions with each other. In order to form lipid rafts, sitosterols must act as bridges connecting the glycolipids.

Chapter 6: Conclusion and Outlook

Three lipid force fields within the CHARMM community were parameterized in this dissertation. The C36/LJ-PME additive and the Drude/LJ-PME polarizable force field use LJ-PME to account for the long-range dispersions which are important for membrane simulations. With these parameterizations, both force fields achieve consistency between bilayers and monolayers. However, the TIP3P water model used in the C36/LJ-PME force field has a too low surface tension, which means the correct isotherm for monolayer involve a cancellation of error. When computational power allows, the Drude lipid force field will likely be a better choice for many membrane simulations, thought the greater efficiency of the additive force fields will make them attractive for a number of years. Nevertheless, additional dynamic properties should be carefully checked for the Drude/LJ-PME force field. In addition, a concurrent parametrization of the Drude double bond and neighboring methylene groups is taking place in the MacKerell group. When the parameterization is completed, the current Drude lipid force field should be tested with the new double bond parameters on unsaturated lipids. Dynamic properties including NMR T_1, lipid diffusion and membrane shear viscosity will also be added to the validation set.

A semi-automated approach (FFLiP) for optimizing the CHARMM and Drude force fields was developed. Unlike the early developments of these force fields which started from parametrizations of small model compounds, FFLiP takes advantage of thermodynamic reweighting to estimate the first derivatives of membrane-related properties with respect to parameters of interest, and iteratively changes the parameters to generate better predictions for these properties. Compared to the conventional ("by-

hand") method which may take months or longer to parametrize a single lipid, the new semi-automated method only takes a few weeks to generate high-quality parameters for several lipids at the same time, and it can be modified easily to include more lipid types and to parameterize protein and nucleic acid FFs as long as experimental references that can be calculated as ensemble averages in simulations are available. Despite the efficiency of FFLiP, it should be used with care if there is a potential phase change (a concern in lipid bilayer simulations), or any other large structural changes when parametrizing a macromolecular force field. In such cases, the trajectory sampled by the simulation may not contain sufficient useful information of the target state, and therefore a meaningful evaluation of the parameter sensitivity is not guaranteed. It should also be noted that most gradient-based optimization algorithms are only capable of finding the local minimum rather than the global minimum. In other words, this method cannot be used to build a new FF from scratch. However, FFLiP can be applied to improve existing force fields in agreement with newly published experimental data or state-of-the-art simulation protocols such as LJ-PME.

FFLiP is by nature based on inference from existing data, which is quite similar to the idea of machine leaning (ML). In recent years, there is a growing trend to use ML in computational chemistry. ML-based FFs have been designed to study electronic effects such as those brought by the lone pairs,[342] thermodynamic properties of bulk systems such as water[343] and bulk materials,[344] chemical reactions which involve bond breaking and formation,[345, 346] and even excited states.[347] These ML-based FFs are mostly trained on *ab initio* data and are intended to close the gap between the accuracy of *ab initio* methods and the efficiency of classical FFs. Training such a force field

requires sampling of many different configurations of the targeted molecule and large-scale *ab initio* calculations that can be computationally expensive. When it comes to large molecules/systems like lipids and biological membranes, a cutoff of interactions is usually needed, and the many-body problem is still approximated in a "mean-field approximation" way.[348] Transferability is another issue with ML-based FFs. Even in the simple example of the LJ potential, Unke et al.[348] have shown that three independent models are needed for the different regions (repulsive, around minimum, and attractive) of the interaction when using the atomic distance r as a descriptor. One possible solution is to introduce physics into the ML model. One example is from Morawietz et al.[349] where ML is used to predict partial charges for each atom based on their local environment, which can then be used in the calculation of the long-range electrostatic energy.

Another force field parameterized in this dissertation is the CHARMM united atom chain model (C36UAr). As a united atom force field, C36UAr is very competitive among existing united atom lipid force fields. As mentioned in the introduction, GROMOS is currently the most popular and versatile UA lipid force field. Unlike the CHARMM UA lipid force fields, only polar hydrogens in the lipid heads are modeled explicitly in the GROMOS family. Although not tested in this dissertation, the headgroup parameters taken directly from C36 would make the S_{CD} values for the headgroup atoms and carbon-2 (C22, C32) more accurate than commonly used lipid force fields. In terms of surface area, C36UAr is better than GROMOS. After the original publication of GROMOS 43A1-S3 by Chiu et al. for DMPC, DPPC and DOPC,[13] several other studies tried to include PG and PE lipids into the FF.[109,110] While

surface areas for PC and PG lipids were generally within 3% deviations from experimental values, Pluhackova et al.[111] observed a too low A_l (50.2 ± 0.3 Å^2) at 308.15 K for POPE (GROMOS 54a7) compared to the experimental value (58.0 ± 0.3 Å^2). For C36UAr, the reported surface areas deviate from experimental values by 1.5% on average, although the average deviation for the two PE lipids is 3.5%. C36UAr is also very accurate for bilayer thicknesses. Given a fixed volume per lipid, the thickness of a lipid bilayer is supposed to be anticorrelated with the A_l. With this in mind, it is not surprising to find out that the thicknesses for the GROMOS family are generally good with the outlier being POPE.[3] The K_A is more of a concern for UA lipid models. While C36UAr underestimated K_A for all bilayers tested in this work by ~25%, the variations of the GROMOS FFs noted above typically overestimate this property by more than 50%.[3] Despite the accuracy of C36UAr for glycerophospholipids, sphingomyelin and sterols specific parameters are missing in the current set. Future parameterization of the force field should be focused on these lipids.

Two sets of membrane models were developed in this dissertation. The first is for the *P. aeruginosa* inner plasma membrane in planktonic and biofilm modes. Compositions were taken from an electrospray ionization mass spectroscopy; eight major lipid types were chosen for the biofilm model and five for the planktonic model. S_{CD} profiles for the cyclic-containing lipids were compared to the preceding experiment and showed a good agreement. The hydrophobic thicknesses agree with the *P. aeruginosa* transmembrane proteins in the OPM database. Compared to *E. coli*, the inner plasma membrane for *P. aeruginosa* is thicker and less rigid. Overall, these detailed models provide a very good representation of the *P. aeruginosa* inner plasma

membrane and should be considered when studying the behavior of membrane-associated proteins. The second model is for the plasma membrane of *Arabidopsis thaliana*. In the asymmetric model, the major difference between the two leaflets is that the outer leaflet contains glycolipids while the inner leaflet only has glycerophospholipids and sitosterol. The saturated tails from the glycolipids and the increased amount of sitosterol lead to a tighter packing of the outer leaflet. Clustering and compositional phase separation analyses have confirmed that sitosterol has a high affinity with glycolipids and potential phase separations were detected using HMM. These models will aid in future simulation studies related to plant membranes, such as membrane partitioning and trans-membrane transport of small molecules.

In conclusion, this dissertation reaches a milestone in the history of the CHARMM lipid force field. The inclusion of the long-range dispersions unifies the treatment of the Lennard Jones potential between different parts of the force field, and the semi-automated method developed in this dissertation allows fast and accurate future parameterization of other lipids. As computing power grows, an era where the CHARMM polarizable (Drude) force field can be more frequently used with proteins, nucleic acids, carbohydrates, and lipids is in the sight.

Funding and Computational Resources

This research was supported by the Intramural Research Program of the NIH, National Heart, Lung and Blood Institute, NSF Grant Nos. MCB-1149187 and CBET-1604576, and the use of the high-performance computational capabilities at the National Institutes of Health, Bethesda, MD (NHLBI LoBoS and Biowulf clusters), the Maryland Advanced Research Computing Center (MARCC), and the University of Maryland (Deepthought2).

Bibliography

(1) van Meer, G.; Voelker, D. R.; Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nature Reviews Molecular Cell Biology* **2008**, *9* (2), 112-124. DOI: 10.1038/nrm2330.
(2) Furt, F.; Simon-Plas, F.; Mongrand, S. Lipids of the Plant Plasma Membrane. In *The Plant Plasma Membrane*, Murphy, A. S., Schulz, B., Peer, W. Eds.; Springer Berlin Heidelberg, 2011; pp 3-30.
(3) van Meer, G.; de Kroon, A. I. P. M. Lipid map of the mammalian cell. *Journal of Cell Science* **2011**, *124* (1), 5-8. DOI: 10.1242/jcs.071233 (acccessed 6/16/2022).
(4) Fahy, E.; Subramaniam, S.; Brown, H. A.; Glass, C. K.; Merrill, A. H., Jr.; Murphy, R. C.; Raetz, C. R. H.; Russell, D. W.; Seyama, Y.; Shaw, W.; et al. A comprehensive classification system for lipids. *Journal of Lipid Research* **2005**, *46* (5), 839-861. DOI: 10.1194/jlr.E400004-JLR200 (acccessed 2022/06/15).
(5) Chaurio, R. A.; Janko, C.; Muñoz, L. E.; Frey, B.; Herrmann, M.; Gaipl, U. S. Phospholipids: key players in apoptosis and immune regulation. *Molecules* **2009**, *14* (12), 4892-4914. DOI: 10.3390/molecules14124892 PubMed.
(6) Cevc, G. How membrane chain-melting phase-transition temperature is affected by the lipid chain asymmetry and degree of unsaturation: an effective chain-length model. *Biochemistry* **1991**, *30* (29), 7186-7193. DOI: 10.1021/bi00243a021.
(7) Andersen, O. S.; Koeppe, R. E. Bilayer Thickness and Membrane Protein Function: An Energetic Perspective. *Annual Review of Biophysics and Biomolecular Structure* **2007**, *36* (1), 107-130. DOI: 10.1146/annurev.biophys.36.040306.132643 (acccessed 2022/06/15).
(8) Cho, W.; Stahelin, R. V. Membrane-Protein Interactions in Cell Signaling and Membrane Trafficking. *Annual Review of Biophysics and Biomolecular Structure* **2005**, *34* (1), 119-151. DOI: 10.1146/annurev.biophys.33.110502.133337 (acccessed 2022/06/15).
(9) Pandit, K. R.; Klauda, J. B. Membrane models of *E. coli* containing cyclic moieties in the aliphatic lipid chain. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818* (5), 1205-1210.
(10) Khakbaz, P.; Klauda, J. B. Probing the importance of lipid diversity in cell membranes via molecular simulation. *Chemistry and Physics of Lipids* **2015**, *192*, 12-22. DOI: https://doi.org/10.1016/j.chemphyslip.2015.08.003.
(11) Kaneda, T. Iso- and anteiso-fatty acids in bacteria: biosynthesis, function, and taxonomic significance. *Microbiol Rev* **1991**, *55* (2), 288-302. DOI: 10.1128/mr.55.2.288-302.1991 PubMed.
(12) Lim, J. B.; Klauda, J. B. Branching at the Iso- and Anteiso- Positions in Complex Chlamydia Membranes: A Molecular Dynamics Study. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808* (1), 323-331
(13) Gronnier, J.; Germain, V.; Gouguet, P.; Cacas, J.-L.; Mongrand, S. GIPC: Glycosyl Inositol Phospho Ceramides, the major sphingolipids on earth. *Plant Signal Behav* **2016**, *11* (4), e1152438-e1152438. DOI: 10.1080/15592324.2016.1152438 PubMed.

(14) Futerman, A. H.; Hannun, Y. A. The complex life of simple sphingolipids. *EMBO Rep* **2004**, *5* (8), 777-782. DOI: 10.1038/sj.embor.7400208 PubMed.
(15) Hofsäss, C.; Lindahl, E.; Edholm, O. Molecular dynamics simulations of phospholipid bilayers with cholesterol. *Biophysical journal* **2003**, *84* (4), 2192-2206. DOI: 10.1016/S0006-3495(03)75025-5 PubMed.
(16) Manno, S.; Takakuwa, Y.; Mohandas, N. Identification of a functional role for lipid asymmetry in biological membranes: Phosphatidylserine-skeletal protein interactions modulate membrane stability. *Proceedings of the National Academy of Sciences* **2002**, *99* (4), 1943-1948. DOI: 10.1073/pnas.042688399 (acccessed 2022/06/15).
(17) Jämbeck, J. P. M.; Lyubartsev, A. P. Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *The Journal of Physical Chemistry B* **2012**, *116* (10), 3164-3179. DOI: 10.1021/jp212503e.
(18) Dickson, C. J.; Madej, B. D.; Skjevik, Å. A.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *Journal of Chemical Theory and Computation* **2014**, *10* (2), 865-879. DOI: 10.1021/ct4010307.
(19) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* **2010**, *114* (23), 7830-7843. DOI: 10.1021/jp101759q.
(20) Chiu, S.-W.; A. Pandit, S.; L. Scott, H.; Jakobsson, E. An Improved United Atom Force Field for Simulation of Mixed Lipid Bilayers. *The Journal of Physical Chemistry B* **2009**, *113* (9), 2748-2763. DOI: 10.1021/jp807056c.
(21) Javanainen, M.; Lamberg, A.; Cwiklik, L.; Vattulainen, I.; Ollila, O. H. S. Atomistic Model for Nearly Quantitative Simulations of Langmuir Monolayers. *Langmuir* **2018**, *34* (7), 2565-2572. DOI: 10.1021/acs.langmuir.7b02855.
(22) Jämbeck, J. P. M.; Lyubartsev, A. P. An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *Journal of Chemical Theory and Computation* **2012**, *8* (8), 2938-2948. DOI: 10.1021/ct300342n.
(23) Jämbeck, J. P. M.; Lyubartsev, A. P. Another Piece of the Membrane Puzzle: Extending Slipids Further. *Journal of Chemical Theory and Computation* **2013**, *9* (1), 774-784. DOI: 10.1021/ct300777p.
(24) Dickson, C. J.; Rosso, L.; Betz, R. M.; Walker, R. C.; Gould, I. R. GAFFlipid: a General Amber Force Field for the accurate molecular dynamics simulation of phospholipid. *Soft Matter* **2012**, *8* (37), 9617-9627, 10.1039/C2SM26007G. DOI: 10.1039/C2SM26007G.
(25) Maciejewski, A.; Pasenkiewicz-Gierula, M.; Cramariuc, O.; Vattulainen, I.; Rog, T. Refined OPLS All-Atom Force Field for Saturated Phosphatidylcholine Bilayers at Full Hydration. *The Journal of Physical Chemistry B* **2014**, *118* (17), 4571-4581. DOI: 10.1021/jp5016627.
(26) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; et al. Current Status of the AMOEBA Polarizable Force Field. *The Journal of Physical Chemistry B* **2010**, *114* (8), 2549-2564. DOI: 10.1021/jp910674d.

(27) Chowdhary, J.; Harder, E.; Lopes, P. E. M.; Huang, L.; MacKerell, A. D.; Roux, B. A Polarizable Force Field of Dipalmitoylphosphatidylcholine Based on the Classical Drude Model for Molecular Dynamics Simulations of Lipids. *The Journal of Physical Chemistry B* **2013**, *117* (31), 9142-9160. DOI: 10.1021/jp402860e.
(28) Li, H.; Chowdhary, J.; Huang, L.; He, X.; MacKerell, A. D.; Roux, B. Drude Polarizable Force Field for Molecular Dynamics Simulations of Saturated and Unsaturated Zwitterionic Lipids. *Journal of Chemical Theory and Computation* **2017**, *13* (9), 4535-4552. DOI: 10.1021/acs.jctc.7b00262.
(29) Lee, S.; Tran, A.; Allsopp, M.; Lim, J. B.; Hénin, J.; Klauda, J. B. CHARMM36 United Atom Chain Model for Lipids and Surfactants. *The Journal of Physical Chemistry B* **2014**, *118* (2), 547-556. DOI: 10.1021/jp410344g.
(30) Yu, Y.; Klauda, J. B. Update of the CHARMM36 United Atom Chain Model for Hydrocarbons and Phospholipids. *Journal of Physical Chemistry B* **2020**, *124* (31), 6797-6812. DOI: 10.1021/acs.jpcb.0c04795.
(31) Berger, O.; Edholm, O.; Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **1997**, *72* (5), 2002-2013. DOI: https://doi.org/10.1016/S0006-3495(97)78845-3.
(32) Chiu, S. W.; Clark, M. M.; Jakobsson, E.; Subramaniam, S.; Scott, H. L. Optimization of Hydrocarbon Chain Interaction Parameters: Application to the Simulation of Fluid Phase Lipid Bilayers. *The Journal of Physical Chemistry B* **1999**, *103* (30), 6323-6327. DOI: 10.1021/jp983219x.
(33) Kukol, A. Lipid Models for United-Atom Molecular Dynamics Simulations of Proteins. *Journal of Chemical Theory and Computation* **2009**, *5* (3), 615-626. DOI: 10.1021/ct8003468.
(34) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *The Journal of Physical Chemistry B* **2004**, *108* (2), 750-760. DOI: 10.1021/jp036508g.
(35) Arnarez, C.; Uusitalo, J. J.; Masman, M. F.; Ingólfsson, H. I.; de Jong, D. H.; Melo, M. N.; Periole, X.; de Vries, A. H.; Marrink, S. J. Dry Martini, a Coarse-Grained Force Field for Lipid Membrane Simulations with Implicit Solvent. *Journal of Chemical Theory and Computation* **2015**, *11* (1), 260-275. DOI: 10.1021/ct500477k.
(36) Madej, B. D.; Gould, I. R.; Walker, R. C. A Parameterization of Cholesterol for Mixed Lipid Bilayer Simulation within the Amber Lipid14 Force Field. *The Journal of Physical Chemistry B* **2015**, *119* (38), 12424-12435. DOI: 10.1021/acs.jpcb.5b04924.
(37) Dickson, C. J.; Walker, R. C.; Gould, I. R. Lipid21: Complex Lipid Membrane Simulations with AMBER. *Journal of Chemical Theory and Computation* **2022**, *18* (3), 1726-1736. DOI: 10.1021/acs.jctc.1c01217.
(38) Grote, F.; Lyubartsev, A. P. Optimization of Slipids Force Field Parameters Describing Headgroups of Phospholipids. *The Journal of Physical Chemistry B* **2020**, *124* (40), 8784-8793. DOI: 10.1021/acs.jpcb.0c06386.
(39) Poger, D.; Van Gunsteren, W. F.; Mark, A. E. A new force field for simulating phosphatidylcholine bilayers. *Journal of Computational Chemistry* **2010**, *31* (6), 1117-1125. DOI: 10.1002/jcc.21396 (acccessed 2020/11/05).

(40) Poger, D.; Mark, A. E. On the Validation of Molecular Dynamics Simulations of Saturated and cis-Monounsaturated Phosphatidylcholine Lipid Bilayers: A Comparison with Experiment. *Journal of Chemical Theory and Computation* **2010**, *6* (1), 325-336. DOI: 10.1021/ct900487a.
(41) Piggot, T. J.; Piñeiro, Á.; Khalid, S. Molecular Dynamics Simulations of Phosphatidylcholine Membranes: A Comparative Force Field Study. *Journal of Chemical Theory and Computation* **2012**, *8* (11), 4593-4609. DOI: 10.1021/ct3003157.
(42) Antila, H.; Buslaev, P.; Favela-Rosales, F.; Ferreira, T. M.; Gushchin, I.; Javanainen, M.; Kav, B.; Madsen, J. J.; Melcr, J.; Miettinen, M. S.; et al. Headgroup Structure and Cation Binding in Phosphatidylserine Lipid Bilayers. *The Journal of Physical Chemistry B* **2019**, *123* (43), 9066-9079. DOI: 10.1021/acs.jpcb.9b06091.
(43) Piggot, T. J.; Holdbrook, D. A.; Khalid, S. Electroporation of the E. coli and S. Aureus Membranes: Molecular Dynamics Simulations of Complex Bacterial Membranes. *The Journal of Physical Chemistry B* **2011**, *115* (45), 13381-13388. DOI: 10.1021/jp207013v.
(44) Chiu, S. W.; Vasudevan, S.; Jakobsson, E.; Mashl, R. J.; Scott, H. L. Structure of Sphingomyelin Bilayers: A Simulation Study. *Biophysical Journal* **2003**, *85*, 3624-3635. DOI: 10.1016/S0006-3495(03)74780-8.
(45) Davis, J. E.; Patel, S. Charge Equilibration Force Fields for Lipid Environments: Applications to Fully Hydrated DPPC Bilayers and DMPC-Embedded Gramicidin A. *The Journal of Physical Chemistry B* **2009**, *113* (27), 9183-9196. DOI: 10.1021/jp901088g.
(46) Venable, R. M.; Krämer, A.; Pastor, R. W. Molecular Dynamics Simulations of Membrane Permeability. *Chemical Reviews* **2019**, *119* (9), 5954-5997. DOI: 10.1021/acs.chemrev.8b00486.
(47) Krämer, A.; C. Pickard, F.; Huang, J.; M. Venable, R.; C. Simmonett, A.; Reith, D.; N. Kirschner, K.; W. Pastor, R.; R. Brooks, B. Interactions of Water and Alkanes: Modifying Additive Force Fields to Account for Polarization Effects. *Journal of Chemical Theory and Computation* **2019**, *15* (6), 3854-3867. DOI: 10.1021/acs.jctc.9b00016.
(48) Orsi, M.; Sanderson, W. E.; Essex, J. W. Permeability of Small Molecules through a Lipid Bilayer: A Multiscale Simulation Study. *The Journal of Physical Chemistry B* **2009**, *113* (35), 12019-12029. DOI: 10.1021/jp903248s.
(49) Sugii, T.; Takagi, S.; Matsumoto, Y. A molecular-dynamics study of lipid bilayers: Effects of the hydrocarbon chain length on permeability. *The Journal of Chemical Physics* **2005**, *123* (18), 184714. DOI: 10.1063/1.2102900 (acccessed 2020/11/05).
(50) Peng, X.; Zhang, Y.; Chu, H.; Li, Y.; Zhang, D.; Cao, L.; Li, G. Accurate Evaluation of Ion Conductivity of the Gramicidin A Channel Using a Polarizable Force Field without Any Corrections. *Journal of Chemical Theory and Computation* **2016**, *12* (6), 2973-2982. DOI: 10.1021/acs.jctc.6b00128.
(51) Chu, H.; Peng, X.; Li, Y.; Zhang, Y.; Min, H.; Li, G. Polarizable atomic multipole-based force field for DOPC and POPE membrane lipids. *Molecular Physics* **2018**, *116* (7-8), 1037-1050. DOI: 10.1080/00268976.2018.1436201.

(52) Davis, J. E.; Rahaman, O.; Patel, S. Molecular Dynamics Simulations of a DMPC Bilayer Using Nonadditive Interaction Models. *Biophysical Journal* **2009**, *96* (2), 385-402. DOI: https://doi.org/10.1016/j.bpj.2008.09.048.
(53) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chemical Reviews* **2016**, *116* (9), 4983-5013. DOI: 10.1021/acs.chemrev.5b00505.
(54) Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *Journal of Chemical Theory and Computation* **2013**, *9* (8), 3527-3537. DOI: 10.1021/ct400140n.
(55) Alvares, D. S.; dos Santos Cabrera, M. P.; Ruggiero Neto, J. Chapter Two - Strategies for Exploring Electrostatic and Nonelectrostatic Contributions to the Interaction of Helical Antimicrobial Peptides with Model Membranes. In *Advances in Biomembranes and Lipid Self-Assembly*, Iglič, A., Kulkarni, C. V., Rappolt, M. Eds.; Vol. 24; Academic Press, 2016; pp 43-73.
(56) Schlenkrich, M.; Brickmann, J.; MacKerell, A. D.; Karplus, M. An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications. In *Biological Membranes: A Molecular Perspective from Computation and Experiment*, Merz, K. M., Roux, B. Eds.; Birkhäuser Boston, 1996; pp 31-81.
(57) Feller, S. E.; Yin, D.; Pastor, R. W.; MacKerell, A. D., Jr. Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies. *Biophysical journal* **1997**, *73* (5), 2269-2279. DOI: 10.1016/S0006-3495(97)78259-6 PubMed.
(58) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **1983**, *79* (2), 926-935. DOI: 10.1063/1.445869.
(59) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-Induced Forces between Two Hydrophilic Groups. *Journal of Physical Chemistry* **1994**, *98* (8), 2198-2202. DOI: 10.1021/j100059a038.
(60) Venable, R. M.; Hatcher, E.; Guvench, O.; Mackerell, A. D., Jr.; Pastor, R. W. Comparing simulated and experimental translation and rotation constants: range of validity for viscosity scaling. *J Phys Chem B* **2010**, *114* (39), 12501-12507. DOI: 10.1021/jp105549s PubMed.
(61) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103* (19), 8577-8593. DOI: 10.1063/1.470117.
(62) Zhuang, X.; Dávila-Contreras, E. M.; Beaven, A. H.; Im, W.; Klauda, J. B. An extensive simulation study of lipid bilayer properties with different head groups, acyl chain lengths, and chain saturations. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858* (12), 3093-3104. DOI: https://doi.org/10.1016/j.bbamem.2016.09.016.
(63) Wu, E. L.; Qi, Y.; Song, K. C.; Klauda, J. B.; Im, W. Preferred Orientations of Phosphoinositides in Bilayers and Their Implications in Protein Recognition Mechanisms. *The Journal of Physical Chemistry B* **2014**, *118* (16), 4315-4325. DOI: 10.1021/jp500610t.

(64) Venable, Richard M.; Sodt, Alexander J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, Alexander D., Jr.; Pastor, Richard W.; Klauda, Jeffery B. CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature. *Biophysical Journal* **2014**, *107* (1), 134-145. DOI: 10.1016/j.bpj.2014.05.034 (acccessed 2020/07/29).
(65) Wang, E.; Klauda, J. B. Molecular Dynamics Simulations of Ceramide and Ceramide-Phosphatidylcholine Bilayers. *The Journal of Physical Chemistry B* **2017**, *121* (43), 10091-10104. DOI: 10.1021/acs.jpcb.7b08967.
(66) Wang, E.; Klauda, J. B. Examination of Mixtures Containing Sphingomyelin and Cholesterol by Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2017**, *121* (18), 4833-4844. DOI: 10.1021/acs.jpcb.7b01832.
(67) Leonard, A. N.; Pastor, R. W.; Klauda, J. B. Parameterization of the CHARMM All-Atom Force Field for Ether Lipids and Model Linear Ethers. *Journal of Physical Chemistry B* **2018**, *122* (26), 6744-6754. DOI: 10.1021/acs.jpcb.8b02743.
(68) West, A.; Zoni, V.; Teague, W. E.; Leonard, A. N.; Vanni, S.; Gawrisch, K.; Tristram-Nagle, S.; Sachs, J. N.; Klauda, J. B. How Do Ethanolamine Plasmalogens Contribute to Order and Structure of Neurological Membranes? *The Journal of Physical Chemistry B* **2020**, *124* (5), 828-839. DOI: 10.1021/acs.jpcb.9b08850.
(69) Wu, Emilia L.; Fleming, Patrick J.; Yeom, Min S.; Widmalm, G.; Klauda, Jeffery B.; Fleming, Karen G.; Im, W. E. coli Outer Membrane and Interactions with OmpLA. *Biophysical Journal* **2014**, *106* (11), 2493-2502. DOI: http://dx.doi.org/10.1016/j.bpj.2014.04.024.
(70) Jo, S.; Wu, E. L.; Stuhlsatz, D.; Klauda, J. B.; MacKerell, A. D.; Widmalm, G.; Im, W. Lipopolysaccharide Membrane Building and Simulation. In *Glycoinformatics*, Lütteke, T., Frank, M. Eds.; Springer New York, 2015; pp 391-406.
(71) Kim, S.; Patel, Dhilon S.; Park, S.; Slusky, J.; Klauda, Jeffery B.; Widmalm, G.; Im, W. Bilayer Properties of Lipid A from Various Gram-Negative Bacteria. *Biophysical Journal* **2016**, *111* (8), 1750-1760. DOI: 10.1016/j.bpj.2016.09.001 (acccessed 2016/10/19).
(72) Patel, Dhilon S.; Park, S.; Wu, Emilia L.; Yeom, Min S.; Widmalm, G.; Klauda, Jeffery B.; Im, W. Influence of Ganglioside GM1 Concentration on Lipid Clustering and Membrane Properties and Curvature. *Biophysical Journal* **2016**, *111* (9), 1987-1999. DOI: https://doi.org/10.1016/j.bpj.2016.09.021.
(73) Guvench, O.; S. Mallajosyula, S.; Prabhu Raman, E.; Hatcher, E.; Vanommeslaeghe, K.; J. Foster, T.; W. Jamison, F.; D. MacKerell, A. CHARMM Additive All-Atom Force Field for Carbohydrate Derivatives and Its Utility in Polysaccharide and Carbohydrate–Protein Modeling. *Journal of Chemical Theory and Computation* **2011**, *7* (10), 3162-3180. DOI: 10.1021/ct200328p.
(74) Mallajosyula, S. S.; Guvench, O.; Hatcher, E.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *Journal of Chemical Theory and Computation* **2012**, *8* (2), 759-776. DOI: 10.1021/ct200792v.
(75) Raman, E. P.; Guvench, O.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *The Journal of Physical Chemistry B* **2010**, *114* (40), 12981-12994. DOI: 10.1021/jp105758h.

(76) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *Journal of Chemical Theory and Computation* **2009**, *5* (9), 2353-2370. DOI: 10.1021/ct900242e.
(77) Klauda, J. B.; Monje, V.; Kim, T.; Im, W. Improving the CHARMM force field for polyunsaturated fatty acid chains. *Journal of Physical Chemistry B* **2012**, *116* (31), 9424-9431. DOI: 10.1021/jp304056p.
(78) Yu, Y.; Klauda, J. B. Modeling Pseudomonas aeruginosa inner plasma membrane in planktonic and biofilm modes. *The Journal of Chemical Physics* **2018**, *149* (21), 215102. DOI: 10.1063/1.5052629 (acccessed 2020/07/27).
(79) Lim, J. B.; Rogaski, B.; Klauda, J. B. Update of the Cholesterol Force Field Parameters in CHARMM. *The Journal of Physical Chemistry B* **2012**, *116* (1), 203-210. DOI: 10.1021/jp207925m.
(80) Monje-Galvan, V.; Klauda, J. B. Modelling Yeast Organelle Membranes and How Lipid Diversity influences Bilayer Properties. *Biochemistry* **2015**, *54* (45), 6852-6861. DOI: 10.1021/acs.biochem.5b00718.
(81) Zhuang, X.; Ou, A.; Klauda, J. B. Simulations of simple linoleic acid-containing lipid membranes and models for the soybean plasma membranes. *The Journal of Chemical Physics* **2017**, *146* (21), 215103. DOI: 10.1063/1.4983655 (acccessed 2021/03/16).
(82) Harder, E.; MacKerell, A. D.; Roux, B. Many-Body Polarization Effects and the Membrane Dipole Potential. *Journal of the American Chemical Society* **2009**, *131* (8), 2760-2761. DOI: 10.1021/ja806825g.
(83) Vorobyov, I. V.; Anisimov, V. M.; MacKerell, A. D. Polarizable Empirical Force Field for Alkanes Based on the Classical Drude Oscillator Model. *The Journal of Physical Chemistry B* **2005**, *109* (40), 18988-18999. DOI: 10.1021/jp053182y.
(84) Yu, H.; Whitfield, T. W.; Harder, E.; Lamoureux, G.; Vorobyov, I.; Anisimov, V. M.; MacKerell, A. D.; Roux, B. Simulating Monovalent and Divalent Ions in Aqueous Solution Using a Drude Polarizable Force Field. *Journal of Chemical Theory and Computation* **2010**, *6* (3), 774-786. DOI: 10.1021/ct900576a.
(85) Schneider, R.; Brugger, B.; Sandhoff, R.; Zellnig, G.; Leber, A.; Lampl, M.; Athenstaedt, K.; Hrastnik, C.; Eder, S.; Duam, G.; et al. Electrospray Ionization Tandem Mass Spectrometry (ESI-MS/MS) Analysis of the Lipid Molecular Species Composition of Yeast Subcellular Membranes Reveals Acyl Chain-based Sorting/Remodeling of Distinct Molecular Species En Route to the Plasma Membrane. *The Journal of Cell Biology* **1999**, *146* (4), 741-754.
(86) Tuller, G.; Nemec, T.; Hrastnik, C.; Daum, G. Lipid composition of subcellular membranes of an FY1679-derived haploid yeast wild-type strain grown on different carbon sources. *Yeast* **1999**, *15* (14), 1555-1564.
(87) Kamp, J. A. F. O. d. Lipid Asymmetry in Membranes. *Annual Review of Biochemistry* **1979**, *48* (1), 47-71. DOI: 10.1146/annurev.bi.48.070179.000403.
(88) Marrink, S. J.; Corradi, V.; Souza, P. C. T.; Ingólfsson, H. I.; Tieleman, D. P.; Sansom, M. S. P. Computational Modeling of Realistic Cell Membranes. *Chemical Reviews* **2019**, *119* (9), 6184-6226. DOI: 10.1021/acs.chemrev.8b00460.

(89) Venable, R. M.; Zhang, Y.; Hardy, B. J.; Pastor, R. W. Molecular dynamics simulations of a lipid bilayer and of hexadecane: An investigation of membrane fluidity. *Science* **1993**, *262*, 223-226. DOI: 10.1126/science.8211140.
(90) Róg, T.; Pasenkiewicz-Gierula, M. Cholesterol Effects on the Phosphatidylcholine Bilayer Nonpolar Region: A Molecular Simulation Study. *Biophysical Journal* **2001**, *81* (4), 2190-2202. DOI: 10.1016/S0006-3495(01)75867-5 (acccessed 2022/06/21).
(91) Murzyn, K.; Róg, T.; Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane. *Biophysical journal* **2005**, *88* (2), 1091-1103. DOI: 10.1529/biophysj.104.048835 PubMed.
(92) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29* (11), 1859-1865. DOI: 10.1002/jcc.20945 (acccessed 2020/07/29).
(93) Wassenaar, T. A.; Ingólfsson, H. I.; Böckmann, R. A.; Tieleman, D. P.; Marrink, S. J. Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *Journal of Chemical Theory and Computation* **2015**, *11* (5), 2144-2155. DOI: 10.1021/acs.jctc.5b00209.
(94) Venable, R. M.; Luo, Y.; Gawrisch, K.; Roux, B.; Pastor, R. W. Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *The Journal of Physical Chemistry B* **2013**, *117* (35), 10183-10192. DOI: 10.1021/jp401512z.
(95) Wu, Emilia L.; Engström, O.; Jo, S.; Stuhlsatz, D.; Yeom, Min S.; Klauda, Jeffery B.; Widmalm, G.; Im, W. Molecular Dynamics and NMR Spectroscopy Studies of E.coli Lipopolysaccharide Structure and Dynamics. *Biophysical Journal* **2013**, *105* (6), 1444-1455. DOI: 10.1016/j.bpj.2013.08.002 (acccessed 2020/07/29).
(96) Klauda, J. B.; Wu, X.; Pastor, R. W.; Brooks, B. R. Long-Range Lennard-Jones and Electrostatic Interactions in Interfaces: Application of the Isotropic Periodic Sum Method. *The Journal of Physical Chemistry B* **2007**, *111* (17), 4393-4400. DOI: 10.1021/jp068767m.
(97) Mecke, M.; Winkelmann, J.; Fischer, J. Molecular Dynamics Simulation of the Liquid-Vapor Interface: The Lennard-Jones Fluid. *Journal of Chemical Physics* **1997**, *107*, 9264-9270. DOI: 10.1063/1.475217.
(98) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry* **2009**, *30* (10), 1545-1614. DOI: 10.1002/jcc.21287.
(99) Wennberg, C. L.; Murtola, T.; Páll, S.; Abraham, M. J.; Hess, B.; Lindahl, E. Direct-Space Corrections Enable Fast and Accurate Lorentz-Berthelot Combination Rule Lennard-Jones Lattice Summation. *Journal of Chemical Theory and Computation* **2015**, *11* (12), 5737-5746. DOI: 10.1021/acs.jctc.5b00726.
(100) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13* (7), e1005659. DOI: 10.1371/journal.pcbi.1005659.

(101) Leonard, A. N.; Simmonett, A. C.; Pickard, F. C.; Huang, J.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Comparison of Additive and Polarizable Models with Explicit Treatment of Long-Range Lennard-Jones Interactions Using Alkane Simulations. *Journal of Chemical Theory and Computation* **2018**, *14* (2), 948-958. DOI: 10.1021/acs.jctc.7b00948.
(102) Pastor, R. W.; MacKerell, A. D. Development of the CHARMM Force Field for Lipids. *The Journal of Physical Chemistry Letters* **2011**, *2* (13), 1526-1532. DOI: 10.1021/jz200167q.
(103) Tu, K.; Tobias, D. J.; Blasie, J. K.; Klein, M. L. Molecular dynamics investigation of the structure of a fully hydrated gel-phase dipalmitoylphosphatidylcholine bilayer. *Biophysical Journal* **1996**, *70* (2), 595-608. DOI: https://doi.org/10.1016/S0006-3495(96)79623-6.
(104) Tieleman, D. P.; Marrink, S. J.; Berendsen, H. J. C. A computer perspective of membranes: molecular dynamics studies of lipid bilayer systems. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* **1997**, *1331* (3), 235-270. DOI: https://doi.org/10.1016/S0304-4157(97)00008-7.
(105) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19-25. DOI: https://doi.org/10.1016/j.softx.2015.06.001.
(106) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **2013**, *3* (2), 198-210, https://doi.org/10.1002/wcms.1121. DOI: https://doi.org/10.1002/wcms.1121 (acccessed 2021/01/29).
(107) Edholm, O. Chapter 3 Time and Length Scales in Lipid Bilayer Simulations. In *Current Topics in Membranes*, Feller, S. E. Ed.; Vol. 60; Academic Press, 2008; pp 91-110.
(108) Venable, R. M.; Brown, F. L. H.; Pastor, R. W. Mechanical Properties of Lipid Bilayers from Molecular Dynamics Simulation. *Chemistry and Physics of Lipids* **2015**, *192*, 60-74. DOI: https://doi.org/10.1016/j.chemphyslip.2015.07.014.
(109) Zhuang, X.; Makover, J. R.; Im, W.; Klauda, J. B. A Systematic Molecular Dynamics Simulation Study of Temperature Dependent Bilayer Structural Properties. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2014**, *1838* (10), 2520-2529. DOI: https://doi.org/10.1016/j.bbamem.2014.06.010.
(110) Krämer, A.; Hülsmann, M.; Köddermann, T.; Reith, D. Automated parameterization of intermolecular pair potentials using global optimization techniques. *Computer Physics Communications* **2014**, *185* (12), 3228-3239. DOI: https://doi.org/10.1016/j.cpc.2014.08.022.
(111) Hülsmann, M.; Kirschner, K. N.; Krämer, A.; Heinrich, D. D.; Krämer-Fuhrmann, O.; Reith, D. Optimizing Molecular Models Through Force-Field Parameterization via the Efficient Combination of Modular Program Packages. In *Foundations of Molecular Modeling and Simulation: Select Papers from FOMMS 2015*, Snurr, R. Q., Adjiman, C. S., Kofke, D. A. Eds.; Springer Singapore, 2016; pp 53-77.

(112) Wang, L.-P.; J. Martinez, T.; S. Pande, V. Building Force Fields: An Automatic, Systematic, and Reproducible Approach. *The Journal of Physical Chemistry Letters* **2014**, *5* (11), 1885-1891. DOI: 10.1021/jz500737m.
(113) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **2008**, *129* (12), 124105. DOI: 10.1063/1.2978177 (acccessed 2020/07/28).
(114) Messerly, R. A.; Razavi, S. M.; Shirts, M. R. Configuration-Sampling-Based Surrogate Models for Rapid Parameterization of Non-Bonded Interactions. *Journal of Chemical Theory and Computation* **2018**, *14* (6), 3144-3162. DOI: 10.1021/acs.jctc.8b00223.
(115) Oliveira, M. P.; Andrey, M.; Rieder, S. R.; Kern, L.; Hahn, D. F.; Riniker, S.; Horta, B. A. C.; Hünenberger, P. H. Systematic Optimization of a Fragment-Based Force Field against Experimental Pure-Liquid Properties Considering Large Compound Families: Application to Saturated Haloalkanes. *Journal of Chemical Theory and Computation* **2020**, *16* (12), 7525-7555. DOI: 10.1021/acs.jctc.0c00683.
(116) Grimme, S. A General Quantum Mechanically Derived Force Field (QMDFF) for Molecules and Condensed Phase Simulations. *Journal of Chemical Theory and Computation* **2014**, *10* (10), 4497-4514. DOI: 10.1021/ct500573f.
(117) Horton, J. T.; Allen, A. E. A.; Dodda, L. S.; Cole, D. J. QUBEKit: Automating the Derivation of Force Field Parameters from Quantum Mechanics. *Journal of Chemical Information and Modeling* **2019**, *59* (4), 1366-1381. DOI: 10.1021/acs.jcim.8b00767.
(118) Ghahremanpour, M. M.; van Maaren, P. J.; Caleman, C.; Hutchison, G. R.; van der Spoel, D. Polarizable Drude Model with s-Type Gaussian or Slater Charge Density for General Molecular Mechanics Force Fields. *Journal of Chemical Theory and Computation* **2018**, *14* (11), 5553-5566. DOI: 10.1021/acs.jctc.8b00430.
(119) Campetella, M.; De Mitri, N.; Prampolini, G. Automated parameterization of quantum-mechanically derived force-fields including explicit sigma holes: A pathway to energetic and structural features of halogen bonds in gas and condensed phase. *The Journal of Chemical Physics* **2020**, *153* (4), 044106. DOI: 10.1063/5.0014280 (acccessed 2021/01/26).
(120) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; et al. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *Journal of Chemical Theory and Computation* **2018**, *14* (11), 6076-6092. DOI: 10.1021/acs.jctc.8b00640.
(121) Qiu, Y. S., Daniel; Boothroyd, Simon; Jang, Hyesu; Wagner, Jeffrey; Bannan, Caitlin C.; et al.. Development and Benchmarking of Open Force Field v1.0.0, the Parsley Small Molecule Force Field. *ChemRxiv* **2020**. (acccessed Jan 10, 2004).
(122) McKiernan, K. A.; Wang, L.-P.; Pande, V. S. Training and Validation of a Liquid-Crystalline Phospholipid Bilayer Force Field. *Journal of Chemical Theory and Computation* **2016**, *12* (12), 5960-5967. DOI: 10.1021/acs.jctc.6b00801.
(123) Hülsmann, M.; Köddermann, T.; Vrabec, J.; Reith, D. GROW: A gradient-based optimization workflow for the automated development of molecular models. *Computer Physics Communications* **2010**, *181* (3), 499-513. DOI: https://doi.org/10.1016/j.cpc.2009.10.024.

(124) Bourasseau, E.; Haboudou, M.; Boutin, A.; Fuchs, A. H.; Ungerer, P. New optimization method for intermolecular potentials: Optimization of a new anisotropic united atoms potential for olefins: Prediction of equilibrium properties. *The Journal of Chemical Physics* **2003**, *118* (7), 3020-3034. DOI: 10.1063/1.1537245 (acccessed 2020/11/20).
(125) Reith, D.; Meyer, H.; Müller-Plathe, F. CG-OPT: A software package for automatic force field design. *Computer Physics Communications* **2002**, *148* (3), 299-313. DOI: https://doi.org/10.1016/S0010-4655(02)00562-3.
(126) Sundaralihgam, M. DISCUSSION PAPER: MOLECULAR STRUCTURES AND CONFORMATIONS OF THE PHOSPHOLIPIDS AND SPHINGOMYELINS. *Ann N Y Acad Sci* **1972**, *195* (1), 324-355. DOI: 10.1111/j.1749-6632.1972.tb54814.x From NLM.
(127) Nomenclature of Lipids. *J. Lipid Res.* **1967**, *8* (5), 523-528.
(128) The Nomenclature of Lipids (Recommendations 1976). IUPAC-IUB Commission on Biochemical Nomenclature. *Journal of lipid research* **1978**, *19* (1), 114-128.
(129) Foglia, F.; Lawrence, M. J.; Lorenz, C. D.; McLain, S. E. On the Hydration of the Phosphocholine Headgroup in Aqueous Solution. *The Journal of Chemical Physics* **2010**, *133* (14), 145103. DOI: 10.1063/1.3488998 (acccessed 2020/07/28).
(130) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105* (43), 9954-9960. DOI: 10.1021/jp003020w.
(131) Kučerka, N.; Nieh, M.-P.; Katsaras, J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808* (11), 2761-2771. DOI: https://doi.org/10.1016/j.bbamem.2011.07.022.
(132) Seelig, A.; Seelig, J. Dynamic structure of fatty acyl chains in a phospholipid bilayer measured by deuterium magnetic resonance. *Biochemistry* **1974**, *13* (23), 4839-4845. DOI: 10.1021/bi00720a024.
(133) Seelig, A.; Seelig, J. Bilayers of Dipalmitoyl-3-sn-Phosphatidylcholine: Conformational Differences between the Fatty Acyl Chains. *BBA - Biomembranes* **1975**, *406* (1), 1-5. DOI: 10.1016/0005-2736(75)90037-1.
(134) Klauda, J. B.; Eldho, N. V.; Gawrisch, K.; Brooks, B. R.; Pastor, R. W. Collective and Noncollective Models of NMR Relaxation in Lipid Vesicles and Multilayers. *Journal of Physical Chemistry B* **2008**, *112* (19), 5924-5929. DOI: 10.1021/jp075641w.
(135) Kučerka, N.; Liu, Y.; Chu, N.; Petrache, H. I.; Tristram-Nagle, S.; Nagle, J. F. Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using x-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophysical Journal* **2005**, *88* (4), 2626-2637. DOI: 10.1529/biophysj.104.056606.
(136) Pan, J.; Cheng, X.; Heberle, F. A.; Mostofian, B.; Kučerka, N.; Drazba, P.; Katsaras, J. Interactions between Ether Phospholipids and Cholesterol As Determined by Scattering and Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **2012**, *116* (51), 14829-14838. DOI: 10.1021/jp310345j.
(137) Guler, S. D.; Ghosh, D. D.; Pan, J.; Mathai, J. C.; Zeidel, M. L.; Nagle, J. F.; Tristram-Nagle, S. Effects of Ether vs. Ester Linkage on Lipid Bilayer Structure and

Water Permeability. *Chemistry and Physics of Lipids* **2009**, *160* (1), 33-44. DOI: https://doi.org/10.1016/j.chemphyslip.2009.04.003.
(138) Nagle, J. F.; Tristram-Nagle, S. Structure of Lipid Bilayers. *Biochimica et Biophysica Acta - Reviews on Biomembranes* **2000**, *1469* (3), 159-195. DOI: 10.1016/S0304-4157(00)00016-2.
(139) Zwanzig, R. W. High‐Temperature Equation of State by a Perturbation Method. I. Nonpolar Gases. *The Journal of Chemical Physics* **1954**, *22* (8), 1420-1426. DOI: 10.1063/1.1740409 (acccessed 2020/11/20).
(140) Oliphant, T. E. Python for Scientific Computing. *Computing in Science & Engineering* **2007**, *9* (3), 10-20. DOI: 10.1109/MCSE.2007.58.
(141) Botan, A.; Favela-Rosales, F.; Fuchs, P. F. J.; Javanainen, M.; Kanduč, M.; Kulig, W.; Lamberg, A.; Loison, C.; Lyubartsev, A.; Miettinen, M. S.; et al. Toward Atomistic Resolution Structure of Phosphatidylcholine Headgroup and Glycerol Backbone at Different Ambient Conditions. *The Journal of Physical Chemistry B* **2015**, *119* (49), 15075-15088. DOI: 10.1021/acs.jpcb.5b04878.
(142) Oliphant, T. E. Python for scientific computing. *Computing in Science and Engineering* **2007**. DOI: 10.1109/MCSE.2007.58.
(143) Guvench, O.; MacKerell, A. D. Automated conformational energy fitting for force-field development. *Journal of Molecular Modeling* **2008**, *14* (8), 667-679. DOI: 10.1007/s00894-008-0305-0.
(144) Jo, S.; Lim, J. B.; Klauda, J. B.; Im, W. CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes. *Biophysical Journal* **2009**, *97* (1), 50-58. DOI: 10.1016/j.bpj.2009.04.013 (acccessed 2020/07/29).
(145) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable, R. M.; et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014**, *35* (27), 1997-2004. DOI: 10.1002/jcc.23702 (acccessed 2020/07/29).
(146) Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; et al. 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* **2016**, *12* (1), 405-413. DOI: 10.1021/acs.jctc.5b00935.
(147) Wu, X.; Brooks, B. R. Self-guided Langevin dynamics simulation method. *Chemical Physics Letters* **2003**, *381* (3), 512-518. DOI: https://doi.org/10.1016/j.cplett.2003.10.013.
(148) Kučerka, N.; Katsaras, J.; Nagle, J. F. Comparing Membrane Simulations to Scattering Experiments: Introducing the SIMtoEXP Software. *Journal of Membrane Biology* **2010**, *235* (1), 43-50. DOI: 10.1007/s00232-010-9254-5.
(149) Lipari, G.; Szabo, A. Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules. 2. Analysis of Experimental Results. *Journal of the American Chemical Society* **1982**, *104* (17), 4559-4570. DOI: 10.1021/ja00381a010.
(150) Yeh, I. C.; Hummer, G. System-size dependence of diffusion coefficients and viscosities from molecular dynamics simulations with periodic boundary conditions.

Journal of Physical Chemistry B **2004**, *108* (40), 15873-15879. DOI: 10.1021/jp0477147.
(151) Camley, B. A.; Lerner, M. G.; Pastor, R. W.; Brown, F. L. H. Strong Influence of Periodic Boundary Conditions on Lateral Diffusion in Lipid Bilayer Membranes. *The Journal of Chemical Physics* **2015**, *143* (24), 243113. DOI: 10.1063/1.4932980 (acccessed 2020/08/19).
(152) Venable, R. M.; Ingólfsson, H. I.; Lerner, M. G.; Perrin, B. S.; Camley, B. A.; Marrink, S. J.; Brown, F. L. H.; Pastor, R. W. Lipid and Peptide Diffusion in Bilayers: The Saffman-Delbrück Model and Periodic Boundary Conditions. *Journal of Physical Chemistry B* **2017**, *121* (15), 3443-3457. DOI: 10.1021/acs.jpcb.6b09111.
(153) Zgorski, A.; Pastor, R. W.; Lyman, E. Surface Shear Viscosity and Interleaflet Friction from Nonequilibrium Simulations of Lipid Bilayers. *Journal of Chemical Theory and Computation* **2019**, *15* (11), 6471-6481. DOI: 10.1021/acs.jctc.9b00683.
(154) Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Dynamical motions of lipids and a finite size effect in simulations of bilayers. *Journal of Chemical Physics* **2006**, *125* (14), 144710-144710. DOI: 10.1063/1.2354486.
(155) Yu, Y.; Krämer, A.; Venable, R. M.; Simmonett, A. C.; MacKerell, A. D.; Klauda, J. B.; Pastor, R. W.; Brooks, B. R. Semi-automated Optimization of the CHARMM36 Lipid Force Field to Include Explicit Treatment of Long-Range Dispersion. *Journal of Chemical Theory and Computation* **2021**. DOI: 10.1021/acs.jctc.0c01326.
(156) Kučerka, N.; Nieh, M. P.; Katsaras, J. Fluid phase lipid areas and bilayer thicknesses of commonly used phosphatidylcholines as a function of temperature. *Biochimica et Biophysica Acta - Biomembranes* **2011**, *1808* (11), 2761-2771. DOI: 10.1016/j.bbamem.2011.07.022.
(157) Kučerka, N.; Nagle, J. F.; Sachs, J. N.; Feller, S. E.; Pencer, J.; Jackson, A.; Katsaras, J. Lipid Bilayer Structure Determined by the Simultaneous Analysis of Neutron and X-Ray Scattering Data. *Biophysical Journal* **2008**, *95* (5), 2356-2367. DOI: https://doi.org/10.1529/biophysj.108.132662.
(158) Pan, J.; Tristram-Nagle, S.; Kučerka, N.; Nagle, J. F. Temperature Dependence of Structure, Bending Rigidity, and Bilayer Interactions of Dioleoylphosphatidylcholine Bilayers. *Biophysical Journal* **2008**, *94* (1), 117-124. DOI: 10.1529/biophysj.107.115691 (acccessed 2020/08/04).
(159) Kučerka, N.; Van Oosten, B.; Pan, J.; Heberle, F. A.; Harroun, T. A.; Katsaras, J. Molecular Structures of Fluid Phosphatidylethanolamine Bilayers Obtained from Simulation-to-Experiment Comparisons and Experimental Scattering Density Profiles. *Journal of Physical Chemistry A* **2015**, *119* (5), 1947-1956. DOI: 10.1021/jp511159q.
(160) Pan, J.; Heberle, F. A.; Tristram-Nagle, S.; Szymanski, M.; Koepfinger, M.; Katsaras, J.; Kučerka, N. Molecular Structures of Fluid Phase Phosphatidylglycerol Bilayers as Determined by Small Angle Neutron and X-ray Scattering. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2012**, *1818* (9), 2135-2148. DOI: https://doi.org/10.1016/j.bbamem.2012.05.007.
(161) Pan, J.; Marquardt, D.; Heberle, F. A.; Kučerka, N.; Katsaras, J. Revisiting the Bilayer Structures of Fluid Phase Phosphatidylglycerol Lipids: Accounting for

Exchangeable Hydrogens. *Biochimica et Biophysica Acta - Biomembranes* **2014**, *1838* (11), 2966-2969. DOI: 10.1016/j.bbamem.2014.08.009.
(162) Somerharju, P. J.; Virtanen, J. A.; Eklund, K. K.; Vainio, P.; Kinnunen, P. K. J. 1-Palmitoyl-2-Pyrenedecanoyl Glycerophospholipids as Membrane Probes: Evidence for Regular Distribution in Liquid-Crystalline Phosphatidylcholine Bilayers. *Biochemistry* **1985**, *24* (11), 2773-2781. DOI: 10.1021/bi00332a027.
(163) Kučerka, N.; Holland, B. W.; Gray, C. G.; Tomberli, B.; Katsaras, J. Scattering Density Profile Model of POPG Bilayers As Determined by Molecular Dynamics Simulations and Small-Angle Neutron and X-ray Scattering Experiments. *The Journal of Physical Chemistry B* **2012**, *116* (1), 232-239. DOI: 10.1021/jp208920h.
(164) Pan, J.; Tristram-Nagle, S.; Nagle, J. F. Effect of cholesterol on structural and mechanical properties of membranes depends on lipid chain saturation. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **2009**, *80*, 021931-021931. DOI: 10.1103/PhysRevE.80.021931.
(165) Shaikh, S. R.; Brzustowicz, M. R.; Gustafson, N.; Stillwell, W.; Wassall, S. R. Monounsaturated PE Does Not Phase-Separate from the Lipid Raft Molecules Sphingomyelin and Cholesterol: Role for Polyunsaturation? *Biochemistry* **2002**, *41* (34), 10593-10602. DOI: 10.1021/bi025712b.
(166) Borle, F.; Seelig, J. Ca2+ Binding to Phosphatidylglycerol Bilayers as Studied by Differential Scanning Calorimetry and 2H- and 31P-nuclear Magnetic Resonance. *Chemistry and Physics of Lipids* **1985**, *36* (3), 263-283. DOI: https://doi.org/10.1016/0009-3084(85)90007-6.
(167) Seelig, A.; Seelig, J. Bilayers of dipalmitoyl-3-sn-phosphatidylcholine. Conformational differences between the fatty acyl chains. *BBA - Biomembranes* **1975**, *406* (1), 1-5. DOI: 10.1016/0005-2736(75)90037-1.
(168) Perly, B.; Smith, I. C. P.; Jarrell, H. C. Acyl chain dynamics of phosphatidylethanolamines containing oleic acid and dihydrosterculic acid: deuteron NMR relaxation studies. *Biochemistry* **1985**, *24* (17), 4659-4665. DOI: 10.1021/bi00338a027.
(169) Cao, Z. T.; Tester, J. W.; Trout, B. L. Sensitivity analysis of hydrate thermodynamic reference properties using experimental data and ab initio methods. *Journal of Physical Chemistry B* **2002**, *106* (31), 7681-7687.
(170) Rawicz, W.; Olbrich, K. C.; McIntosh, T.; Needham, D.; Evans, E. A. Effect of chain length and unsaturation on elasticity of lipid bilayers. *Biophysical Journal* **2000**, *79* (1), 328-339. DOI: 10.1016/S0006-3495(00)76295-3.
(171) Evans, E.; Rawicz, W.; Smith, B. A. Concluding remarks back to the future: Mechanics and thermodynamics of lipid biomembranes. *Faraday Discussions* **2012**, *161*, 591-611. DOI: 10.1039/c2fd20127e.
(172) Binder, H.; Gawrisch, K. Effect of Unsaturated Lipid Chains on Dimensions, Molecular Order and Hydration of Membranes. *The Journal of Physical Chemistry B* **2001**, *105* (49), 12378-12390. DOI: 10.1021/jp010118h.
(173) Rand, R. P.; Parsegian, V. A. Hydration forces between phospholipid bilayers. *BBA - Reviews on Biomembranes* **1989**, *988* (3), 351-376. DOI: 10.1016/0304-4157(89)90010-5.

(174) Leftin, A.; Brown, M. F. An NMR Database for Simulations of Membrane Dynamics. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808* (3), 818-839. DOI: https://doi.org/10.1016/j.bbamem.2010.11.027.
(175) Brown, M. F.; Ribeiro, A. A.; Williams, G. D. New View of Lipid Bilayer Dynamics from 2H and 13C NMR Relaxation Time Measurements. *Proceedings of the National Academy of Sciences of the United States of America* **1983**, *80* (14), 4325-4329. DOI: 10.1073/pnas.80.14.4325 PubMed.
(176) Szabo, A. Nuclear Magnetic Resonance Relaxation and the Dynamics of Proteins and Membranes: Theory and Experiment. *Annals of the New York Academy of Sciences* **1986**, *482* (1), 44-50. DOI: 10.1111/j.1749-6632.1986.tb20935.x (acccessed 2020/08/17).
(177) Szabo, A. Theory of Fluorescence Depolarization in Macromolecules and Membranes. *The Journal of Chemical Physics* **1984**, *81* (1), 150-167. DOI: 10.1063/1.447378 (acccessed 2020/08/17).
(178) Klauda, J. B.; Roberts, M. F.; Redfield, A. G.; Brooks, B. R.; Pastor, R. W. Rotation of Lipids in Membranes: Molecular Dynamics Simulation, 31P Spin-Lattice Relaxation, and Rigid-Body Dynamics. *Biophysical Journal* **2008**, *94* (8), 3074-3083. DOI: 10.1529/biophysj.107.121806.
(179) Camley, B. A.; Brown, F. L. H. Diffusion of Complex Objects Embedded in Free and Supported Lipid Bilayer Membranes: Role of Shape Anisotropy and Leaflet Structure. *Soft Matter* **2013**, *9* (19), 4767-4779, 10.1039/C3SM00073G. DOI: 10.1039/C3SM00073G.
(180) Scheidt, H. A.; Huster, D.; Gawrisch, K. Diffusion of Cholesterol and Its Precursors in Lipid Membranes Studied by 1H Pulsed Field Gradient Magic Angle Spinning NMR. *Biophysical journal* **2005**, *89* (4), 2504-2512. DOI: 10.1529/biophysj.105.062018.
(181) Lindblom, G.; Orädd, G. Lipid Lateral Diffusion and Membrane Heterogeneity. *Biochimica et Biophysica Acta - Biomembranes* **2009**, *1788* (1), 234-244. DOI: 10.1016/j.bbamem.2008.08.016.
(182) You, X.; Jing, H. *On the Transferability of CHARMM36m Protein Force Field with LJ-PME: Hydrogen Bonding Dynamics under Elevated Pressures*; 2021. DOI: 10.26434/chemrxiv.13653203.v1.
(183) Hilgemann, D. W. Local PIP2 signals: when, where, and how? *Pflügers Archiv - European Journal of Physiology* **2007**, *455* (1), 55-67. DOI: 10.1007/s00424-007-0280-9.
(184) van den Bogaart, G.; Meyenberg, K.; Risselada, H. J.; Amin, H.; Willig, K. I.; Hubrich, B. E.; Dier, M.; Hell, S. W.; Grubmüller, H.; Diederichsen, U.; et al. Membrane protein sequestering by ionic protein–lipid interactions. *Nature* **2011**, *479* (7374), 552-555. DOI: 10.1038/nature10545.
(185) Toner, M.; Vaio, G.; McLaughlin, A.; McLaughlin, S. Adsorption of cations to phosphatidylinositol 4,5-bisphosphate. *Biochemistry* **1988**, *27* (19), 7435-7443. DOI: 10.1021/bi00419a039.
(186) Jiang, Z.; Redfern, R. E.; Isler, Y.; Ross, A. H.; Gericke, A. Cholesterol stabilizes fluid phosphoinositide domains. *Chemistry and Physics of Lipids* **2014**, *182*, 52-61. DOI: https://doi.org/10.1016/j.chemphyslip.2014.02.003.

(187) Wang, Y.-H.; Collins, A.; Guo, L.; Smith-Dupont, K. B.; Gai, F.; Svitkina, T.; Janmey, P. A. Divalent Cation-Induced Cluster Formation by Polyphosphoinositides in Model Membranes. *Journal of the American Chemical Society* **2012**, *134* (7), 3387-3395. DOI: https://doi.org/10.1021/ja208640t.
(188) Levental, I.; Janmey, P. A.; Cēbers, A. Electrostatic Contribution to the Surface Pressure of Charged Monolayers Containing Polyphosphoinositides. *Biophysical Journal* **2008**, *95* (3), 1199-1205. DOI: https://doi.org/10.1529/biophysj.107.126615.
(189) Petelska, A. D.; Naumowicz, M.; Figaszewski, Z. A. The Influence of pH on Phosphatidylethanolamine Monolayer at the Air/Aqueous Solution Interface. *Cell Biochemistry and Biophysics* **2013**, *65* (2), 229-235. DOI: 10.1007/s12013-012-9424-4.
(190) Cseh, R.; Benz, R. The Adsorption of Phloretin to Lipid Monolayers and Bilayers Cannot Be Explained by Langmuir Adsorption Isotherms Alone. *Biophysical Journal* **1998**, *74* (3), 1399-1408. DOI: https://doi.org/10.1016/S0006-3495(98)77852-X.
(191) Gawrisch, K.; Parsegian, V. A.; Hajduk, D. A.; Tate, M. W.; Gruner, S. M.; Fuller, N. L.; Rand, R. P. Energetics of a hexagonal-lamellar-hexagonal-phase transition sequence in dioleoylphosphatidylethanolamine membranes. *Biochemistry* **1992**, *31* (11), 2856-2864. DOI: 10.1021/bi00126a003.
(192) Loew, C.; Riske, K. A.; Lamy, M. T.; Seelig, J. Thermal Phase Behavior of DMPG Bilayers in Aqueous Dispersions as Revealed by 2H- and 31P-NMR. *Langmuir* **2011**, *27* (16), 10041-10049. DOI: 10.1021/la201027p.
(193) Antila, H. S.; Wurl, A.; Ollila, O. H. S.; Miettinen, M. S.; Ferreira, T. M. Quasi-uncoupled rotational diffusion of phospholipid headgroups from the main molecular frame. *arXiv:2009.06774* **2020**.
(194) Antila, H. S.; M. Ferreira, T.; Ollila, O. H. S.; Miettinen, M. S. Using Open Data to Rapidly Benchmark Biomolecular Simulations: Phospholipid Conformational Dynamics. *Journal of Chemical Information and Modeling* **2021**. DOI: 10.1021/acs.jcim.0c01299.
(195) Mabrey, S.; Sturtevant, J. M. Investigation of phase transitions of lipids and lipid mixtures by high sensitivity differential scanning calorimetry. *Proceedings of the National Academy of Sciences of the United States of America* **1976**, *73* (11), 3862-3866. DOI: 10.1073/pnas.73.11.3862.
(196) Clarke, R. J. The dipole potential of phospholipid membranes and methods for its detection. *Advances in Colloid and Interface Science* **2001**, *89-90*, 263-281. DOI: https://doi.org/10.1016/S0001-8686(00)00061-0.
(197) Lamoureux, G.; Roux, B. t. Modeling induced polarization with classical Drude oscillators: Theory and molecular dynamics simulation algorithm. *The Journal of Chemical Physics* **2003**, *119* (6), 3025-3039. DOI: 10.1063/1.1589749 (acccessed 2020/07/29).
(198) Jiang, W.; Hardy, D. J.; Phillips, J. C.; MacKerell, A. D.; Schulten, K.; Roux, B. High-Performance Scalable Molecular Dynamics Simulations of a Polarizable Force Field Based on Classical Drude Oscillators in NAMD. *The Journal of Physical Chemistry Letters* **2011**, *2* (2), 87-92. DOI: 10.1021/jz101461d.
(199) Lemkul, J. A.; Roux, B.; van der Spoel, D.; MacKerell Jr, A. D. Implementation of extended Lagrangian dynamics in GROMACS for polarizable

simulations using the classical Drude oscillator model. *Journal of Computational Chemistry* **2015**, *36* (19), 1473-1479, https://doi.org/10.1002/jcc.23937. DOI: https://doi.org/10.1002/jcc.23937 (acccessed 2022/06/16).
(200) Huang, J.; Lemkul, J. A.; Eastman, P. K.; MacKerell Jr, A. D. Molecular dynamics simulations using the drude polarizable force field on GPUs with OpenMM: Implementation, validation, and benchmarks. *Journal of Computational Chemistry* **2018**, *39* (21), 1682-1689, https://doi.org/10.1002/jcc.25339. DOI: https://doi.org/10.1002/jcc.25339 (acccessed 2022/06/16).
(201) Huang, J.; Lopes, P. E. M.; Roux, B.; MacKerell, A. D. Recent Advances in Polarizable Force Fields for Macromolecules: Microsecond Simulations of Proteins Using the Classical Drude Oscillator Model. *The Journal of Physical Chemistry Letters* **2014**, *5* (18), 3144-3150. DOI: 10.1021/jz501315h.
(202) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B.; MacKerell, A. D. A polarizable model of water for molecular dynamics simulations of biomolecules. *Chemical Physics Letters* **2006**, *418* (1-3), 245-249. DOI: 10.1016/j.cplett.2005.10.135.
(203) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802, https://doi.org/10.1002/jcc.20289. DOI: https://doi.org/10.1002/jcc.20289 (acccessed 2020/12/09).
(204) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press, 1989.
(205) Kognole, A. A.; Aytenfisu, A. H.; MacKerell, A. D. Balanced polarizable Drude force field parameters for molecular anions: phosphates, sulfates, sulfamates, and oxides. *Journal of Molecular Modeling* **2020**, *26* (6), 152. DOI: 10.1007/s00894-020-04399-0.
(206) Gross, J. D.; Warschawski, D. E.; Griffin, R. G. Dipolar Recoupling in MAS NMR: A Probe for Segmental Order in Lipid Bilayers. *Journal of the American Chemical Society* **1997**, *119* (4), 796-802. DOI: 10.1021/ja962951b.
(207) Hong, M.; Schmidt-Rohr, K.; Pines, A. NMR Measurement of Signs and Magnitudes of C-H Dipolar Couplings in Lecithin. *Journal of the American Chemical Society* **1995**, *117* (11), 3310-3311. DOI: 10.1021/ja00116a051.
(208) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; et al. PSI4 1.4: Open-source software for high-throughput quantum chemistry. *The Journal of Chemical Physics* **2020**, *152* (18), 184108. DOI: 10.1063/5.0006002 (acccessed 2022/06/18).
(209) Wang, L.-P.; Song, C. Geometry optimization made simple with translation and rotation coordinates. *The Journal of Chemical Physics* **2016**, *144* (21), 214108. DOI: 10.1063/1.4952956 (acccessed 2022/06/18).
(210) Yu, Y.; Krämer, A.; Venable, R. M.; Brooks, B. R.; Klauda, J. B.; Pastor, R. W. CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *Journal of Chemical Theory and Computation* **2021**. DOI: 10.1021/acs.jctc.0c01327.

(211) Kognole, A. A.; Lee, J.; Park, S.-J.; Jo, S.; Chatterjee, P.; Lemkul, J. A.; Huang, J.; MacKerell Jr, A. D.; Im, W. CHARMM-GUI Drude prepper for molecular dynamics simulation using the classical Drude polarizable force field. *Journal of Computational Chemistry* **2022**, *43* (5), 359-375, https://doi.org/10.1002/jcc.26795. DOI: https://doi.org/10.1002/jcc.26795 (acccessed 2022/06/19).
(212) Zhang, Y.; Feller, S. E.; Brooks, B. R.; Pastor, R. W. Computer simulation of liquid/liquid interfaces. I. Theory and application to octane/water. *The Journal of Chemical Physics* **1995**, *103* (23), 10252-10266. DOI: 10.1063/1.469927.
(213) Krämer, A.; Ghysels, A.; Wang, E.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Membrane permeability of small molecules from unbiased molecular dynamics simulations. *The Journal of Chemical Physics* **2020**, *153* (12), 124107. DOI: 10.1063/5.0013429 (acccessed 2022/06/19).
(214) Smaby, J. M.; Brockman, H. L. Surface dipole moments of lipids at the argon-water interface. Similarities among glycerol-ester-based lipids. *Biophysical Journal* **1990**, *58* (1), 195-204. DOI: https://doi.org/10.1016/S0006-3495(90)82365-1.
(215) Mathai, J. C.; Tristram-Nagle, S.; Nagle, J. F.; Zeidel, M. L. Structural Determinants of Water Permeability through the Lipid Membrane. *Journal of General Physiology* **2007**, *131* (1), 69-76. DOI: 10.1085/jgp.200709848 (acccessed 6/21/2022).
(216) Crane, J. M.; Putz, G.; Hall, S. B. Persistence of Phase Coexistence in Disaturated Phosphatidylcholine Monolayers at High Surface Pressures. *Biophysical Journal* **1999**, *77* (6), 3134-3143. DOI: https://doi.org/10.1016/S0006-3495(99)77143-2.
(217) Leonard, A. N.; Wang, E.; Monje-Galvan, V.; Klauda, J. B. Developing and Testing of Lipid Force Fields with Applications to Modeling Cellular Membranes. *Chemical Reviews* **2019**, *119* (9), 6227-6269. DOI: 10.1021/acs.chemrev.8b00384.
(218) Chiu, S. W.; Clark, M.; Balaji, V.; Subramaniam, S.; Scott, H. L.; Jakobsson, E. Incorporation of surface tension into molecular dynamics simulation of an interface: a fluid phase lipid bilayer membrane. *Biophysical Journal* **1995**, *69*, 1230-1245. DOI: 10.1016/S0006-3495(95)80005-6.
(219) Hénin, J.; Shinoda, W.; Klein, M. L. United-Atom Acyl Chains for CHARMM Phospholipids. *The Journal of Physical Chemistry B* **2008**, *112* (23), 7008-7015. DOI: 10.1021/jp800687p.
(220) Klauda, J. B.; Brooks, B. R.; MacKerell, A. D.; Venable, R. M.; Pastor, R. W. An ab Initio Study on the Torsional Surface of Alkanes and Its Effect on Molecular Simulations of Alkanes and a DPPC Bilayer. *Journal of Physical Chemistry B* **2005**, *109* (11), 5300-5311. DOI: 10.1021/jp0468096.
(221) Jorgensen, W. L.; Madura, J. D.; Swenson, C. J. Optimized Intermolecular Potential Functions for Liquid Hydrocarbons. *Journal of the American Chemical Society* **1984**, *106* (22), 6638-6646. DOI: 10.1021/ja00334a030.
(222) Seelig, J.; Waespe-Šarčević, N. Molecular Order in Cis and Trans Unsaturated Phospholipid Bilayers†. *Biochemistry* **1978**, *17* (16), 3310-3315. DOI: 10.1021/bi00609a021.
(223) Rowan, T. H. Functional stability analysis of numerical algorithms. Ph.D. Dissertation. 1990.
(224) Steven, G. J. The NLopt nonlinear-optimization package.

(225) Martinez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30* (13), 2157-2164. DOI: 10.1002/jcc.21224.
(226) Martyna, G. J.; Tuckerman, M. E.; Tobias, D. J.; Klein, M. L. Explicit reversible integrators for extended systems dynamics. *Molecular Physics* **1996**, *87* (5), 1117-1157. DOI: 10.1080/00268979600100761.
(227) Phillips, J. C.; Gengbin, Z.; Kumar, S.; Kale, L. V. NAMD: Biomolecular Simulation on Thousands of Processors. 2002, IEEE Computer Society Press: pp 1-18. DOI: 10.1109/sc.2002.10019.
(228) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995**, *103* (11), 4613-4621. DOI: 10.1063/1.470648.
(229) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics* **1994**, *101* (5), 4177-4189. DOI: 10.1063/1.467468.
(230) H. de Vries, A.; Chandrasekhar, I.; F. van Gunsteren, W.; H. Hünenberger, P. Molecular Dynamics Simulations of Phospholipid Bilayers: Influence of Artificial Periodicity, System Size, and Simulation Time. *The Journal of Physical Chemistry B* **2005**, *109* (23), 11643-11652. DOI: 10.1021/jp0507952.
(231) Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11* (4), 1864-1874. DOI: 10.1021/ct5010406.
(232) Balusek, C.; Hwang, H.; Hon Lau, C.; Lundquist, K.; Hazel, A.; Pavlova, A.; L. Lynch, D.; H. Reggio, P.; Wang, Y.; C. Gumbart, J. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2019**, *15* (8), 4673-4686. DOI: 10.1021/acs.jctc.9b00160.
(233) Lemmon, E. W.; McLinden, M. O.; Friend, D. G. Thermophysical Properties of Fluid Systems. Linstrom, P. J., Mallard, W. G. Eds.
(234) Lide, D. R. *CRC Handbook of Chemistry and Physics 86TH Edition*; CRC Press, Inc., 2005.
(235) Small, D. M. *Handbook of lipid research 4: The physical chemistry of lipids*; Plenum Press, 1986.
(236) Camin, D. L.; Rossini, F. D. Physical properties of the 17 isomeric hexenes of the API research series. *Journal of Physical Chemistry* **1956**, *60* (10), 1446-1451. DOI: 10.1021/j150544a029.
(237) Yaws, C. L. Chapter 1 - Physical Properties – Organic Compounds. Yaws, C. L. B. T. T. Y. H. o. P. P. f. H. a. C. Ed.; Gulf Professional Publishing, 2015; pp 1-683.
(238) Majer, V. r.; Svoboda, V. c.; Kehiaian, H. V.; Chemistry, I. U. o. P. a. A. *Enthalpies of vaporization of organic compounds : a critical review and data compilation*; Blackwell Scientific, 1985.
(239) Chickos, J. S.; Acree, W. E. Enthalpies of vaporization of organic and organometallic compounds, 1880-2002. *Journal of Physical and Chemical Reference Data* **2003**, *32* (2), 519-878. DOI: 10.1063/1.1529214.

(240) Stephenson, R. M.; Malanowski, S. *Handbook of the Thermodynamics of Organic Compounds*; 1987. DOI: 10.1007/978-94-009-3173-2.
(241) Aralaguppi, M. I.; Aminabhavi, T. M.; Balundgi, R. H.; Joshi, S. S. Thermodynamic interactions in mixtures of bromoform with hydrocarbons. *Journal of Physical Chemistry* **1991**, *95* (13), 5299-5308. DOI: 10.1021/j100166a070.
(242) Michailidou, E. K.; Assael, M. J.; Huber, M. L.; Abdulagatov, I. M.; Perkins, R. A. Reference correlation of the viscosity of n-Heptane from the Triple Point to 600 K and up to 248 MPa. *Journal of Physical and Chemical Reference Data* **2014**, *43* (2). DOI: 10.1063/1.4875930.
(243) Fermeglia, M.; Torriano, G. Density, Viscosity, and Refractive Index for Binary Systems of n-C16 and Four Nonlinear Alkanes at 298.15 K. *Journal of Chemical & Engineering Data* **1999**, *44* (5), 965-969. DOI: 10.1021/je9900171.
(244) Klein, T.; Yan, S.; Cui, J.; Magee, J. W.; Kroenlein, K.; Rausch, M. H.; Koller, T. M.; Fröba, A. P. Liquid Viscosity and Surface Tension of n-Hexane, n-Octane, n-Decane, and n-Hexadecane up to 573 K by Surface Light Scattering. *Journal of Chemical & Engineering Data* **2019**, *64* (9), 4116-4131. DOI: 10.1021/acs.jced.9b00525.
(245) Ertl, H.; Dullien, F. A. L. Self‐diffusion and viscosity of some liquids as a function of temperature. *AIChE Journal* **1973**, *19* (6), 1215-1223. DOI: 10.1002/aic.690190619.
(246) Fishman, E. Self-diffusion in liquid n-pentane and n-heptane. *Journal of Physical Chemistry* **1955**, *59* (5), 469-472. DOI: 10.1021/j150527a022.
(247) Dymond, J. H.; Harris, K. R. The temperature and density dependence of the self-diffusion coefficient of n-hexadecane. *Molecular Physics* **1992**, *75* (2), 461-466. DOI: 10.1080/00268979200100351.
(248) Eldho, N. V.; Feller, S. E.; Tristram-Nagle, S.; Polozov, I. V.; Gawrisch, K. Polyunsaturated docosahexaenoic vs docosapentaenoic acid - Differences in lipid matrix properties from the loss of one double bond. *Journal of the American Chemical Society* **2003**. DOI: 10.1021/ja029029o.
(249) Nagle, J. F. Area/lipid of bilayers from NMR. *Biophysical Journal* **1993**, *64* (5), 1476-1481. DOI: https://doi.org/10.1016/S0006-3495(93)81514-5.
(250) Khakbaz, P.; Klauda, J. B. Investigation of phase transitions of saturated phosphocholine lipid bilayers via molecular dynamics simulations. *Biochimica et Biophysica Acta - Biomembranes* **2018**, *1860* (8), 1489-1501. DOI: 10.1016/j.bbamem.2018.04.014.
(251) Tristram-Nagle, S.; Zhang, R.; Suter, R. M.; Worthington, C. R.; Sun, W. J.; Nagle, J. F. Measurement of chain tilt angle in fully hydrated bilayers of gel phase lecithins. *Biophysical journal* **1993**, *64* (4), 1097-1109. DOI: 10.1016/S0006-3495(93)81475-9 PubMed.
(252) Das, C.; Noro, M. G.; Olmsted, P. D. Simulation studies of stratum corneum lipid mixtures. *Biophysical Journal* **2009**, *97* (7), 1941-1951. DOI: 10.1016/j.bpj.2009.06.054.
(253) Krämer, A.; Ghysels, A.; Wang, E.; Venable, R. M.; Klauda, J. B.; Brooks, B. R.; Pastor, R. W. Membrane Permeability of Small Molecules from Unbiased Molecular Dynamics Simulations. *J. Chem. Phys.*

(254) Lee, C. T.; Comer, J.; Herndon, C.; Leung, N.; Pavlova, A.; Swift, R. V.; Tung, C.; Rowley, C. N.; Amaro, R. E.; Chipot, C.; et al. Simulation-Based Approaches for Determining Membrane Permeability of Small Compounds. *Journal of Chemical Information and Modeling* **2016**, *56* (4), 721-733. DOI: 10.1021/acs.jcim.6b00022.
(255) Awoonor-Williams, E.; Rowley, C. N. Molecular simulation of nonfacilitated membrane permeation. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2016**, *1858* (7, Part B), 1672-1687. DOI: https://doi.org/10.1016/j.bbamem.2015.12.014.
(256) Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving efficiency of large time-scale molecular dynamics simulations of hydrogen-rich systems. *Journal of Computational Chemistry* **1999**, *20* (8), 786-798. DOI: 10.1002/(SICI)1096-987X(199906)20:8<786::AID-JCC5>3.0.CO;2-B.
(257) Lindahl, E.; Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B. Implementation of the charmm force field in GROMACS: Analysis of protein stability effects from correction maps, virtual interaction sites, and water models. *Journal of Chemical Theory and Computation* **2010**, *6* (2), 459-466. DOI: 10.1021/ct900549r.
(258) Loubet, B.; Kopec, W.; Khandelia, H. Accelerating All-Atom MD Simulations of Lipids Using a Modified Virtual-Sites Technique. *Journal of Chemical Theory and Computation* **2014**, *10* (12), 5690-5695. DOI: 10.1021/ct500100f.
(259) Saliba, A.-E.; Vonkova, I.; Gavin, A.-C. The systematic analysis of protein–lipid interactions comes of age. *Nature Reviews Molecular Cell Biology* **2015**, *16* (12), 753-761. DOI: 10.1038/nrm4080.
(260) Saliba, A.-E.; Vonkova, I.; Ceschia, S.; Findlay, G. M.; Maeda, K.; Tischer, C.; Deghou, S.; van Noort, V.; Bork, P.; Pawson, T.; et al. A quantitative liposome microarray to systematically characterize protein-lipid interactions. *Nature Methods* **2014**, *11* (1), 47-50. DOI: 10.1038/nmeth.2734.
(261) Li, X.; Gianoulis, T. A.; Yip, K. Y.; Gerstein, M.; Snyder, M. Extensive In Vivo Metabolite-Protein Interactions Revealed by Large-Scale Systematic Analyses. *Cell* **2010**, *143* (4), 639-650. DOI: 10.1016/j.cell.2010.09.048 (acccessed 2020/10/13).
(262) Maeda, K.; Anand, K.; Chiapparino, A.; Kumar, A.; Poletto, M.; Kaksonen, M.; Gavin, A.-C. Interactome map uncovers phosphatidylserine transport by oxysterol-binding proteins. *Nature* **2013**, *501* (7466), 257-261. DOI: 10.1038/nature12430.
(263) Laufer, C.; Fischer, B.; Huber, W.; Boutros, M. Measuring genetic interactions in human cells by RNAi and imaging. *Nature Protocols* **2014**, *9* (10), 2341-2353. DOI: 10.1038/nprot.2014.160.
(264) Isakoff, S. J.; Cardozo, T.; Andreev, J.; Li, Z.; Ferguson, K. M.; Abagyan, R.; Lemmon, M. A.; Aronheim, A.; Skolnik, E. Y. Identification and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast. *The EMBO Journal* **1998**, *17* (18), 5374-5387. DOI: 10.1093/emboj/17.18.5374 (acccessed 2020/10/13).
(265) Contreras, F. X.; Ernst, A. M.; Haberkant, P.; Björkholm, P.; Lindahl, E.; Gönen, B.; Tischer, C.; Elofsson, A.; von Heijne, G.; Thiele, C.; et al. Molecular recognition of a single sphingolipid species by a protein's transmembrane domain. *Nature* **2012**, *481* (7382), 525-529. DOI: 10.1038/nature10742.

(266) Hulce, J. J.; Cognetta, A. B.; Niphakis, M. J.; Tully, S. E.; Cravatt, B. F. Proteome-wide mapping of cholesterol-interacting proteins in mammalian cells. *Nature Methods* **2013**, *10* (3), 259-264. DOI: 10.1038/nmeth.2368.
(267) Haberkant, P.; Holthuis, J. C. M. Fat & fabulous: Bifunctional lipids in the spotlight. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **2014**, *1841* (8), 1022-1030. DOI: https://doi.org/10.1016/j.bbalip.2014.01.003.
(268) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* **2012**, *41* (1), 429-452. DOI: 10.1146/annurev-biophys-042910-155245.
(269) Sodt, A. J.; Sandar, M. L.; Gawrisch, K.; Pastor, R. W.; Lyman, E. The Molecular Structure of the Liquid-Ordered Phase of Lipid Bilayers. *Journal of the American Chemical Society* **2014**, *136* (2), 725-732. DOI: 10.1021/ja4105667.
(270) Dror, R. O.; Green, H. F.; Valant, C.; Borhani, D. W.; Valcourt, J. R.; Pan, A. C.; Arlow, D. H.; Canals, M.; Lane, J. R.; Rahmani, R.; et al. Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. *Nature* **2013**, *503* (7475), 295-299. DOI: 10.1038/nature12595.
(271) Park, S.; Beaven, A. H.; Klauda, J. B.; Im, W. How Tolerant are Membrane Simulations with Mismatch in Area per Lipid between Leaflets? *Journal of Chemical Theory and Computation* **2015**, *11* (7), 3466-3477. DOI: 10.1021/acs.jctc.5b00232.
(272) Gerald P. Bodey, R. B., Victor Fainstein, and Leena Jadeja. Infections Caused by Pseudomonas aeruginosa. *Reviews of Infectious Disease* **1983**, *5* (2), 279 - 313.
(273) Lederberg, J. e. a. Pseudomona. In *Encyclopedia of Microbiology*, second ed.; San Diego 2000; pp 876-891.
(274) de Bentzmann, S.; Plésiat, P. The Pseudomonas aeruginosa opportunistic pathogen and human infections. *Environmental Microbiology* **2011**, *13* (7), 1655-1665. DOI: 10.1111/j.1462-2920.2011.02469.x (acccessed 2018/07/26).
(275) Mah, T. F.; O'Toole, G. A. Mechanisms of biofilm resistance to antimicrobial agents. *Trends in microbiology* **2001**, *9* (1), 34-39.
(276) Schierholz, J. M.; Beuth, J.; Pulverer, G. Adherent bacteria and activity of antibiotics. *The Journal of antimicrobial chemotherapy* **1999**, *43* (1), 158-160.
(277) Stewart, P. S. Mechanisms of antibiotic resistance in bacterial biofilms. *International journal of medical microbiology : IJMM* **2002**, *292* (2), 107-113. DOI: 10.1078/1438-4221-00196.
(278) Drenkard, E. Antimicrobial resistance of Pseudomonas aeruginosa biofilms. *Microbes Infect* **2003**, *5* (13), 1213-1219. DOI: 10.1016/j.micinf.2003.08.009.
(279) Folsom, J. P.; Richards, L.; Pitts, B.; Roe, F.; Ehrlich, G. D.; Parker, A.; Mazurie, A.; Stewart, P. S. Physiology of Pseudomonas aeruginosa in biofilms as revealed by transcriptome analysis. *BMC Microbiology* **2010**, *10*, 294-294. DOI: 10.1186/1471-2180-10-294 PMC.
(280) Zhang, B.; Powers, R. Analysis of bacterial biofilms using NMR-based metabolomics. *Future medicinal chemistry* **2012**, *4* (10), 1273-1306. DOI: 10.4155/fmc.12.59 PMC.
(281) Junter, G.-A.; Jouenne, T. Immobilized viable microbial cells: from the process to the proteome… or the cart before the horse. *Biotechnology Advances* **2004**, *22* (8), 633-658. DOI: https://doi.org/10.1016/j.biotechadv.2004.06.003.

(282) Denich, T. J.; Beaudette, L. A.; Lee, H.; Trevors, J. T. Effect of selected environmental and physico-chemical factors on bacterial cytoplasmic membranes. *Journal of Microbiological Methods* **2003**, *52* (2), 149-182. DOI: https://doi.org/10.1016/S0167-7012(02)00155-0.
(283) Fuqua, W. C.; Winans, S. C.; Greenberg, E. P. Quorum sensing in bacteria: the LuxR-LuxI family of cell density-responsive transcriptional regulators. *Journal of bacteriology* **1994**, *176* (2), 269-275.
(284) Parkin, J.; Chavent, M.; Khalid, S. Molecular Simulations of Gram-Negative Bacterial Membranes: A Vignette of Some Recent Successes. *Biophysical Journal* **2015**, *109* (3), 461-468. DOI: 10.1016/j.bpj.2015.06.050 PMC.
(285) Benamara, H.; Rihouey, C.; Jouenne, T.; Alexandre, S. Impact of the biofilm mode of growth on the inner membrane phospholipid composition and lipid domains in Pseudomonas aeruginosa. *Biochimica et biophysica acta* **2011**, *1808* (1), 98-105. DOI: 10.1016/j.bbamem.2010.09.004.
(286) Benamara, H.; Rihouey, C.; Abbes, I.; Ben Mlouka, M. A.; Hardouin, J.; Jouenne, T.; Alexandre, S. Characterization of membrane lipidome changes in Pseudomonas aeruginosa during biofilm growth on glass wool. *PloS one* **2014**, *9* (9), e108478. DOI: 10.1371/journal.pone.0108478.
(287) Mizuno, T.; Kageyama, M. Separation and Characterization of the Outer Membrane of Pseudomonas aeruginosa. *The Journal of Biochemistry* **1978**, *84* (1), 179-191.
(288) Jo, S.; Kim, T.; Iyer Vidyashankara, G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29* (11), 1859-1865. DOI: 10.1002/jcc.20945 (acccessed 2018/07/25).
(289) Jo, S.; Kim, T.; Im, W. Automated builder and database of Protein/Membrane complexes for molecular dynamics simulations. *PLoS One* **2007**, *2* (9), e880.
(290) Wu Emilia, L.; Cheng, X.; Jo, S.; Rui, H.; Song Kevin, C.; Dávila-Contreras Eder, M.; Qi, Y.; Lee, J.; Monje-Galvan, V.; Venable Richard, M.; et al. CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *Journal of Computational Chemistry* **2014**, *35* (27), 1997-2004. DOI: 10.1002/jcc.23702 (acccessed 2018/07/25).
(291) Phillips James, C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel Robert, D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26* (16), 1781-1802. DOI: 10.1002/jcc.20289 (acccessed 2018/07/25).
(292) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *The journal of physical chemistry. B* **2010**, *114* (23), 7830-7843. DOI: 10.1021/jp101759q PMC.
(293) Durell, S. R.; Brooks, B. R.; Ben-Naim, A. Solvent-Induced Forces between Two Hydrophilic Groups. *The Journal of Physical Chemistry* **1994**, *98* (8), 2198-2202. DOI: 10.1021/j100059a038.
(294) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-

alkanes. *Journal of Computational Physics* **1977**, *23* (3), 327-341. DOI: https://doi.org/10.1016/0021-9991(77)90098-5.
(295) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397 (acccessed 2021/01/13).
(296) Steinbach Peter, J.; Brooks Bernard, R. New spherical-cutoff methods for long-range forces in macromolecular simulation. *Journal of Computational Chemistry* **1994**, *15* (7), 667-683. DOI: 10.1002/jcc.540150702 (acccessed 2018/07/25).
(297) Shinoda, W.; Okazaki, S. A Voronoi analysis of lipid area fluctuation in a bilayer. *The Journal of Chemical Physics* **1998**, *109* (4), 1517-1521. DOI: 10.1063/1.476702 (acccessed 2018/07/26).
(298) Perly, B.; Smith, I. C. P.; Jarrell, H. C. Effects of the replacement of a double bond by a cyclopropane ring in phosphatidylethanolamines: a deuterium NMR study of phase transitions and molecular organization. *Biochemistry* **1985**, *24* (4), 1055-1063. DOI: 10.1021/bi00325a038.
(299) Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes database. *Bioinformatics* **2006**, *22* (5), 623-625. DOI: 10.1093/bioinformatics/btk023.
(300) Róg, T.; Pasenkiewicz-Gierula, M.; Vattulainen, I.; Karttunen, M. Ordering effects of cholesterol and its analogues. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2009**, *1788* (1), 97-121. DOI: http://dx.doi.org/10.1016/j.bbamem.2008.08.022.
(301) Lim, J. B.; Klauda, J. B. Lipid chain branching at the iso- and anteiso-positions in complex chlamydia membranes: A molecular dynamics study. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **2011**, *1808* (1), 323-331. DOI: https://doi.org/10.1016/j.bbamem.2010.07.036.
(302) Koornneef, M.; Meinke, D. The Development of Arabidopsis as A Model Plant. *The Plant Journal* **2010**, *61* (6), 909-921. DOI: 10.1111/j.1365-313X.2009.04086.x (acccessed 2020/10/10).
(303) Lloyd, A. M.; Barnason, A. R.; Rogers, S. G.; Byrne, M. C.; Fraley, R. T.; Horsch, R. B. Transformation of Arabidopsis thaliana with Agrobacterium tumefaciens. *Science* **1986**, *234* (4775), 464. DOI: 10.1126/science.234.4775.464.
(304) Valvekens, D.; Van Montagu, M.; Van Lijsebettens, M. Agrobacterium tumefaciens-Mediated Transformation of Arabidopsis thaliana Root Explants by Using Kanamycin Selection. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85* (15), 5536-5540. DOI: 10.1073/pnas.85.15.5536 PubMed.
(305) Koncz, C.; Martini, N.; Mayerhofer, R.; Koncz-Kalman, Z.; Körber, H.; Redei, G. P.; Schell, J. High-Frequency T-DNA-Mediated Gene Tagging in Plants. *Proceedings of the National Academy of Sciences* **1989**, *86* (21), 8467. DOI: 10.1073/pnas.86.21.8467.
(306) Feldmann, K. A.; David Marks, M. Agrobacterium-Mediated Transformation of Germinating Seeds of Arabidopsis thaliana: A Non-Tissue Culture Approach. *Molecular and General Genetics MGG* **1987**, *208* (1), 1-9. DOI: 10.1007/BF00330414.

(307) Leutwiler, L. S.; Hough-Evans, B. R.; Meyerowitz, E. M. The DNA of Arabidopsis thaliana. *Molecular and General Genetics MGG* **1984**, *194* (1), 15-23. DOI: 10.1007/BF00383491.
(308) Cheng, C.-Y.; Krishnakumar, V.; Chan, A. P.; Thibaud-Nissen, F.; Schobel, S.; Town, C. D. Araport11: A Complete Reannotation of the Arabidopsis thaliana Reference Genome. *The Plant Journal* **2017**, *89* (4), 789-804. DOI: 10.1111/tpj.13415 (acccessed 2020/10/10).
(309) **Meyerowitz, E. M. a. S. C. R. e.** *Arabidopsis*; Cold Spring Harbor Press, 1994.
(310) *The Arabidopsis Information Resource*. https://www.arabidopsis.org (accessed 2020 Oct 10).
(311) Jones, A. M.; Chory, J.; Dangl, J. L.; Estelle, M.; Jacobsen, S. E.; Meyerowitz, E. M.; Nordborg, M.; Weigel, D. The Impact of Arabidopsis on Human Health: Diversifying Our Portfolio. *Cell* **2008**, *133* (6), 939-943. DOI: 10.1016/j.cell.2008.05.040 PubMed.
(312) Jones, J. D. G.; Dangl, J. L. The Plant Immune System. *Nature* **2006**, *444* (7117), 323-329. DOI: 10.1038/nature05286.
(313) Parry, G.; Estelle, M. Auxin Receptors: A New Role for F-box Proteins. *Current Opinion in Cell Biology* **2006**, *18* (2), 152-156. DOI: https://doi.org/10.1016/j.ceb.2006.02.001.
(314) Deng, X. W.; Caspar, T.; Quail, P. H. cop1: a regulatory locus involved in light-controlled development and gene expression in Arabidopsis. *Genes & Development* **1991**, *5* (7), 1172-1182. DOI: 10.1101/gad.5.7.1172.
(315) Yadeta, K. A.; Elmore, J. M.; Coaker, G. Advancements in the Analysis of the Arabidopsis Plasma Membrane Proteome. *Front Plant Sci* **2013**, *4*, 86-86. DOI: 10.3389/fpls.2013.00086 PubMed.
(316) Clouse, S. D. Arabidopsis Mutants Reveal Multiple Roles for Sterols in Plant Development. *The Plant Cell* **2002**, *14* (9), 1995. DOI: 10.1105/tpc.140930.
(317) Cacas, J.-L.; Buré, C.; Grosjean, K.; Gerbeau-Pissot, P.; Lherminier, J.; Rombouts, Y.; Maes, E.; Bossard, C.; Gronnier, J.; Furt, F.; et al. Revisiting Plant Plasma Membrane Lipids in Tobacco: A Focus on Sphingolipids. *Plant Physiology* **2016**, *170* (1), 367. DOI: 10.1104/pp.15.00564.
(318) Mongrand, S.; Morel, J.; Laroche, J.; Claverol, S.; Carde, J.-P.; Hartmann, M.-A.; Bonneu, M.; Simon-Plas, F.; Lessire, R.; Bessoule, J.-J. Lipid Rafts in Higher Plant Cells: PURIFICATION AND CHARACTERIZATION OF TRITON X-100-INSOLUBLE MICRODOMAINS FROM TOBACCO PLASMA MEMBRANE. *Journal of Biological Chemistry* **2004**, *279* (35), 36277-36286. DOI: 10.1074/jbc.M403440200.
(319) Stanislas, T.; Bouyssie, D.; Rossignol, M.; Vesa, S.; Fromentin, J.; Morel, J.; Pichereaux, C.; Monsarrat, B.; Simon-Plas, F. Quantitative Proteomics Reveals a Dynamic Association of Proteins to Detergent-resistant Membranes upon Elicitor Signaling in Tobacco. *Molecular & Cellular Proteomics* **2009**, *8* (9), 2186. DOI: 10.1074/mcp.M900090-MCP200.
(320) Titapiwatanakun, B.; Blakeslee, J. J.; Bandyopadhyay, A.; Yang, H.; Mravec, J.; Sauer, M.; Cheng, Y.; Adamec, J.; Nagashima, A.; Geisler, M.; et al. ABCB19/PGP19 stabilises PIN1 in membrane microdomains in Arabidopsis. *The*

Plant Journal **2009**, *57* (1), 27-44. DOI: 10.1111/j.1365-313X.2008.03668.x (acccessed 2020/10/11).
(321) Tjellström, H.; Hellgren, L. I.; Wieslander, Å.; Sandelius, A. S. Lipid asymmetry in plant plasma membranes: phosphate deficiency-induced phospholipid replacement is restricted to the cytosolic leaflet. *The FASEB Journal* **2010**, *24* (4), 1128-1138. DOI: 10.1096/fj.09-139410 (acccessed 2020/10/13).
(322) Gilmour, S. J.; Hajela, R. K.; Thomashow, M. F. Cold Acclimation in Arabidopsis thaliana. *Plant Physiology* **1988**, *87* (3), 745. DOI: 10.1104/pp.87.3.745.
(323) Uemura, M.; Joseph, R. A.; Steponkus, P. L. Cold Acclimation of Arabidopsis thaliana (Effect on Plasma Membrane Lipid Composition and Freeze-Induced Lesions). *Plant Physiology* **1995**, *109* (1), 15. DOI: 10.1104/pp.109.1.15.
(324) Guo, D.-a.; Venkatramesh, M.; Nes, W. D. Developmental regulation of sterol biosynthesis in Zea mays. *Lipids* **1995**, *30* (3), 203, https://doi.org/10.1007/BF02537823. DOI: https://doi.org/10.1007/BF02537823 (acccessed 2021/01/03).
(325) Grosjean, K.; Mongrand, S.; Beney, L.; Simon-Plas, F.; Gerbeau-Pissot, P. Differential Effect of Plant Lipids on Membrane Organization: SPECIFICITIES OF PHYTOSPHINGOLIPIDS AND PHYTOSTEROLS*. *Journal of Biological Chemistry* **2015**, *290* (9), 5810-5825. DOI: https://doi.org/10.1074/jbc.M114.598805.
(326) Markham, J. E.; Li, J.; Cahoon, E. B.; Jaworski, J. G. Separation and Identification of Major Plant Sphingolipid Classes from Leaves. *Journal of Biological Chemistry* **2006**, *281* (32), 22684-22694. DOI: 10.1074/jbc.M604050200.
(327) Hsieh, T. C. Y.; Kaul, K.; Laine, R. A.; Lester, R. L. Structure of a major glycophosphoceramide from tobacco leaves, PSL-I: 2-deoxy-2-acetamido-D-glucopyranosyl(α 1 → 4)-D-glucuronopyranosyl(α 1 → 2)myoinositol-1-O-phosphoceramide. *Biochemistry* **1978**, *17* (17), 3575-3581. DOI: 10.1021/bi00610a024.
(328) Hsieh, T. C.; Lester, R. L.; Laine, R. A. Glycophosphoceramides from plants. Purification and characterization of a novel tetrasaccharide derived from tobacco leaf glycolipids. *Journal of Biological Chemistry* **1981**, *256* (15), 7747-7755.
(329) Cacas, J.-L.; Furt, F.; Le Guédard, M.; Schmitter, J.-M.; Buré, C.; Gerbeau-Pissot, P.; Moreau, P.; Bessoule, J.-J.; Simon-Plas, F.; Mongrand, S. Lipids of plant membrane rafts. *Progress in Lipid Research* **2012**, *51* (3), 272-299. DOI: https://doi.org/10.1016/j.plipres.2012.04.001.
(330) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics* **2020**, *153* (4), 044130. DOI: 10.1063/5.0014475 (acccessed 2021/03/16).
(331) Yu, Y.; Klauda, J. B. Symmetric and Asymmetric Models for the Arabidopsis thaliana Plasma Membrane: A Simulation Study. *The Journal of Physical Chemistry B* **2021**, *125* (41), 11418-11431. DOI: 10.1021/acs.jpcb.1c04704.
(332) Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon; 1996.

(333) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (null), 2825–2830.
(334) Leonard, E. B.; Ted, P.; George, S.; Norman, W. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *The Annals of Mathematical Statistics* **1970**, *41* (1), 164-171. DOI: 10.1214/aoms/1177697196.
(335) Welch, L. R. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter* **2003**, *53* (4), 10-13.
(336) Viterbi, A. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* **1967**, *13* (2), 260-269. DOI: 10.1109/TIT.1967.1054010.
(337) Gracià, R. S.; Bezlyepkina, N.; Knorr, R. L.; Lipowsky, R.; Dimova, R. Effect of cholesterol on the rigidity of saturated and unsaturated membranes: fluctuation and electrodeformation analysis of giant vesicles. *Soft Matter* **2010**, *6* (7), 1472-1482, 10.1039/B920629A. DOI: 10.1039/B920629A.
(338) Mamode Cassim, A.; Navon, Y.; Gao, Y.; Decossas, M.; Fouillen, L.; Grélard, A.; Nagano, M.; Lambert, O.; Bahammou, D.; Van Delft, P.; et al. Biophysical analysis of the plant-specific GIPC sphingolipids reveals multiple modes of membrane regulation. *Journal of Biological Chemistry* **2021**, *296*. DOI: 10.1016/j.jbc.2021.100602 (acccessed 2021/09/21).
(339) Samuli Ollila, O. H.; Róg, T.; Karttunen, M.; Vattulainen, I. Role of sterol type on lateral pressure profiles of lipid membranes affecting membrane protein functionality: Comparison between cholesterol, desmosterol, 7-dehydrocholesterol and ketosterol. *Journal of Structural Biology* **2007**, *159* (2), 311-323. DOI: https://doi.org/10.1016/j.jsb.2007.01.012.
(340) Chen, Z.; Rand, R. P. The influence of cholesterol on phospholipid membrane curvature and bending elasticity. *Biophysical Journal* **1997**, *73* (1), 267-276. DOI: https://doi.org/10.1016/S0006-3495(97)78067-6.
(341) Kollmitzer, B.; Heftberger, P.; Rappolt, M.; Pabst, G. Monolayer spontaneous curvature of raft-forming membrane lipids. *Soft Matter* **2013**, *9* (45), 10877-10884, 10.1039/C3SM51829A. DOI: 10.1039/C3SM51829A.
(342) Sauceda, H. E.; Chmiela, S.; Poltavsky, I.; Müller, K.-R.; Tkatchenko, A. Molecular force fields with gradient-domain machine learning: Construction and application to dynamics of small molecules with coupled cluster forces. *The Journal of Chemical Physics* **2019**, *150* (11), 114102. DOI: 10.1063/1.5078687 (acccessed 2022/07/19).
(343) Hermann, J.; DiStasio, R. A.; Tkatchenko, A. First-Principles Models for van der Waals Interactions in Molecules and Materials: Concepts, Theory, and Applications. *Chemical Reviews* **2017**, *117* (6), 4714-4758. DOI: 10.1021/acs.chemrev.6b00446.
(344) Deringer, V. L.; Csányi, G. Machine learning based interatomic potential for amorphous carbon. *Physical Review B* **2017**, *95* (9), 094203. DOI: 10.1103/PhysRevB.95.094203.
(345) Brickel, S.; Das, A. K.; Unke, O. T.; Turan, H. T.; Meuwly, M. Reactive molecular dynamics for the [Cl–CH_3–Br]− reaction in the gas phase and in solution: a

comparative study using empirical and neural network force fields. *Electronic Structure* **2019**, *1* (2), 024002. DOI: 10.1088/2516-1075/ab1edb.
(346) Sweeny, B. C.; Pan, H.; Kassem, A.; Sawyer, J. C.; Ard, S. G.; Shuman, N. S.; Viggiano, A. A.; Brickel, S.; Unke, O. T.; Upadhyay, M.; et al. Thermal activation of methane by MgO+: temperature dependent kinetics, reactive molecular dynamics simulations and statistical modeling. *Physical Chemistry Chemical Physics* **2020**, *22* (16), 8913-8923, 10.1039/D0CP00668H. DOI: 10.1039/D0CP00668H.
(347) Chen, W.-K.; Liu, X.-Y.; Fang, W.-H.; Dral, P. O.; Cui, G. Deep Learning for Nonadiabatic Excited-State Dynamics. *The Journal of Physical Chemistry Letters* **2018**, *9* (23), 6702-6708. DOI: 10.1021/acs.jpclett.8b03026.
(348) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chemical Reviews* **2021**, *121* (16), 10142-10186. DOI: 10.1021/acs.chemrev.0c01111.
(349) Morawietz, T.; Sharma, V.; Behler, J. A neural network potential-energy surface for the water dimer based on environment-dependent atomic energies and charges. *The Journal of Chemical Physics* **2012**, *136* (6), 064103. DOI: 10.1063/1.3682557 (acccessed 2022/07/19).