

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



## Understanding Lipid Membrane Biophysics Through Molecular Simulation

Smith, Paul

*Awarding institution:*  
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

### END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

### Take down policy

If you believe that this document breaches copyright please contact [librarypure@kcl.ac.uk](mailto:librarypure@kcl.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Understanding Lipid Membrane Biophysics Through Molecular Simulation

*Paul Smith*

A dissertation submitted in partial fulfillment  
of the requirements for the degree of  
**Doctor of Philosophy**  
of  
**King's College London.**

Department of Physics  
King's College London

December 19, 2021



and being there for me when I was depressed.

To my family — Mum, Dad, Sarah, Matt, Bec, Rafal, and all my wonderful nieces and nephews — I love you all. Thank you for your unconditional support and encouragement, even when I said I wanted to start a PhD at 30 years old. I promise I will get a job now.

To J-Bao, thank you for sticking with me during these past few months of thesis writing and toe-healing. With you, I feel seen, heard, understood. Long may our adventures continue.

And finally, to Jan Rogan, my sixth form philosophy teacher, thank you for starting me on this journey. When I had no plans for what to do after A-levels, you suggested I ‘might quite enjoy’ studying at university. In doing so, you have had a profound impact on my life and I will be eternally grateful.

*'To understand and be understood is to be free.'*

— Daniel Johnston, 'Go'

# Abstract

The plasma membrane is a dynamic interface between a cell and its external environment. It is a complex structure composed of lipids, proteins, carbohydrates, and RNAs. Interactions between these constituent molecules give rise to biochemical processes such as cell-signalling, antigen presentation, and vesiculation. Part of this thesis seeks to contribute to our current understanding of the molecular origin of both physiological and pathophysiological phenomena in cellular membranes. This is done through use of coarse-grained and all-atom molecular dynamics (MD) simulations. The other part of this thesis aims to improve and add to the set of software tools currently used for analysing MD simulations of lipid membranes.

In the plasma membrane, specific lipid species are thought to aggregate into functional platforms known as ‘lipid-rafts’. The lipid-raft hypothesis posits that these small regions, nanometers in size, are comprised of highly ordered lipids in which cell-signalling proteins are embedded. If the aggregation of cell-signalling proteins within lipid-rafts is required for signal transmission, the breakdown of these structures would disrupt signalling pathways and in turn bring about cell death. In Chapter 3, I describe how the oxidation of cholesterol leads to the disruption of these highly ordered nanodomains in model membranes. I reveal three potential mechanisms by which nanodomain formation is disrupted, and in doing so I provide a molecular level description of the means by which cholesterol oxidation may cause apoptosis in biological membranes.

Cholesterol and sphingomyelin are the two lipid species widely thought to be important in the formation of lipid-rafts in mammalian plasma membranes. These two

lipid species are known to have a high affinity for one another, and in complex lipid mixtures this leads to their co-localisation in highly ordered nanodomains. However, the precise origin of the preferential mixing of these two lipids is currently unknown. In Chapter 4, through an unsupervised clustering of cholesterol-sphingomyelin conformations I find there are four distinct modes of interaction between these lipid species. One of these modes is possible only with sphingomyelins — not other lipid species. This particular mode desolvates the hydrophobic core of cholesterol, thus reducing the free energy cost of exposing this core to the surrounding solvent. Therefore, I suggest this mode of interaction between cholesterol and sphingomyelin is the reason why cholesterol preferentially mixes with sphingomyelins over other lipid species.

In Chapter 5, I describe LiPyphilic - an open-source Python package I have created for analysing MD simulations of lipid membranes. LiPyphilic offers analyses that provide important structural and dynamical information about lipid membranes, but are not available in any other software. LiPyphilic is fast, fully-tested and easy to install. It is designed to be interoperable with the wider scientific Python stack, and was built following best practices in modern software development. The challenge now is to ensure the long-term sustainability of LiPyphilic by building a community of users and contributors to the project.

# Publications

During the course of my PhD, I have contributed to the publication of the following articles.

11. M. A. al-Baadri, **P. Smith**, K. T. al-Jamal, C. D. Lorenz, 'Nanomaterial functionalisation modulates hard protein corona formation: Atomistic simulations applied to graphitic materials', *Adv. Mater. Interfaces* **2021**, (In press).
10. **P. Smith**, C. D. Lorenz, 'LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations' *J. Chem. Theory Comput.* **2021**, DOI: 10.1021/acs.jctc.1c00447.
9. **P. Smith**, P. G. Petrov, C. D. Lorenz, 'Cholesterol Oxidation Modulates the Formation of Liquid-Ordered Domains in Model Membranes' *bioRxiv* **2021**, 2021.05.24.445501. [Retracted by all authors, 2021]
8. R. M. Ziolek, **P. Smith**, D. L. Pink, C. A. Dreiss, C. D. Lorenz, 'Unsupervised Learning Unravels the Structure of Four-Arm and Linear Block Copolymer Micelles' *Macromolecules* **2021**, 54, 8, 3755–3768.
7. M. A. al-Badri, **P. Smith**, R. C. Sinclair, K. T. al-Jamal, C. D. Lorenz, 'Accurate large scale modelling of Graphene Oxide: ion trapping and chaotropic potential at the interface' *Carbon* **2021**, 174, 266-275.
6. **P. Smith**, D. M. Owen, C. D. Lorenz, M. Makarova, 'Asymmetric glycerophospholipids impart distinctive biophysical properties to lipid bilayers' *Biophys. J.* **2021**, 120, 9, 1746-1754.



5. D. Ahmadi, R. Ledder, N. Mahmoudi, P. Li, J. Tellam, D. Robinson, R. Heenan, **P. Smith**, C. D. Lorenz, D. Barlow, J. M. Lawrence, 'Supramolecular Architecture of a Multicomponent Biomimetic Lipid Barrier Formulation' *J. Colloid Interface Sci.* **2021**, 587, 597-612.
4. F. Cipcigan, **P. Smith**, J. Crain, A. Hogner, L. De Maria, A. Llinas, E. Ratkova, 'Membrane permeability in cyclic peptides is modulated by core conformations' *J. Chem. Inf. Model.* **2020**, 61, 1, 263-269.
3. **P. Smith**, P. J. Quinn, C. D. Lorenz, 'Two Coexisting Membrane Structures are Defined by Lateral and Transbilayer Interactions Between Sphingomyelin and Cholesterol' *Langmuir* **2020**, 36, 33, 9786-9799.
2. I. E. Steinmark, P-H. Chung, R. M. Ziolek, B. Cornell, **P. Smith**, J. A. Levitt, C. Tregidgo, C. Molteni, G. Yahioğlu, C. D. Lorenz, K. Suhling, 'Time-Resolved Fluorescence Anisotropy of a Molecular Rotor Resolves Microscopic Viscosity Parameters in Complex Environments' *Small* **2020**, 1907139.
1. **P. Smith**, R. M. Ziolek, E. Gazzarrini, D. M. Owen, C. D. Lorenz, 'On the interaction of hyaluronic acid with synovial fluid lipid membranes' *Phys. Chem. Chem. Phys.* **2019**, 21, 9845-9857.

# Software

During the course of my PhD, I have contributed to development of the following open-source software packages:

- LiPyphilic (**creator**) <https://github.com/p-j-smith/lipyphilic>
- MDAnalysis (**contributor**) <https://github.com/MDAnalysis/mdanalysis>
- PyMOL (**contributor**) <https://github.com/schrodinger/pymol-open-source>

# Contents

<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Motivation . . . . .	13
1.2	The cellular membrane . . . . .	13
1.3	Lipids of the cellular membrane . . . . .	14
1.3.1	Glycerophospholipids . . . . .	16
1.3.2	Sphingolipids . . . . .	18
1.3.3	Acyl tails . . . . .	19
1.3.4	Sterols . . . . .	20
1.4	Supramolecular lipid structures . . . . .	21
1.5	Lipid bilayer phases . . . . .	21
1.6	Lateral heterogeneities . . . . .	25
1.7	Lipid-associated pathologies . . . . .	27
1.8	Molecular dynamics simulations . . . . .	29
1.8.1	The early history of computer simulation . . . . .	30
1.8.2	A brief overview of lipid membrane simulations . . . . .	32
1.9	Overview of this thesis . . . . .	34
<b>2</b>	<b>Methods</b>	<b>35</b>
2.1	Molecular dynamics . . . . .	35
2.1.1	Force field . . . . .	35
2.1.2	System creation . . . . .	40
2.1.3	Gradient descent . . . . .	40
2.1.4	Initialise velocities . . . . .	41

2.1.5	Numerical integration of the equations of motion . . . . .	42
2.1.6	Thermostats and barostats . . . . .	43
2.1.7	Improving computational efficiency . . . . .	45
2.1.8	Simulating bulk behaviour . . . . .	47
2.1.9	Coarse-grained molecular dynamics . . . . .	49
2.1.10	Summary . . . . .	51
2.2	Machine learning . . . . .	51
2.2.1	Dimensionality reduction and clustering . . . . .	52
2.2.2	Hidden Markov Models . . . . .	56
<b>3</b>	<b>Cholesterol Oxidation Modulates the Formation of Liquid-Ordered Domains in Model Membranes</b>	<b>58</b>
3.1	Introduction . . . . .	59
3.2	Methods . . . . .	60
3.2.1	Simulation protocol . . . . .	60
3.2.2	MARTINI parameters for 7-ketocholesterol . . . . .	61
3.2.3	Analysis methods . . . . .	62
3.3	Results . . . . .	65
3.4	Conclusion . . . . .	73
3.5	Addendum . . . . .	74
<b>4</b>	<b>Two Coexisting Membrane Structures Are Defined By Lateral and Transbilayer Interactions Between Sphingomyelin and Cholesterol</b>	<b>77</b>
<b>5</b>	<b>LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations</b>	<b>93</b>
<b>6</b>	<b>Conclusions</b>	<b>108</b>
<b>7</b>	<b>Further Work</b>	<b>III</b>
	<b>Appendices</b>	<b>II4</b>

<b>A</b>	<b>Supporting Information for ‘Two Coexisting Membrane Structures Are Defined By Lateral and Transbilayer Interactions Between Sphingomyelin and Cholesterol’</b>	<b>114</b>
<b>B</b>	<b>Supporting Information for ‘LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations’</b>	<b>136</b>
	<b>Bibliography</b>	<b>143</b>

## Chapter 1

# Introduction

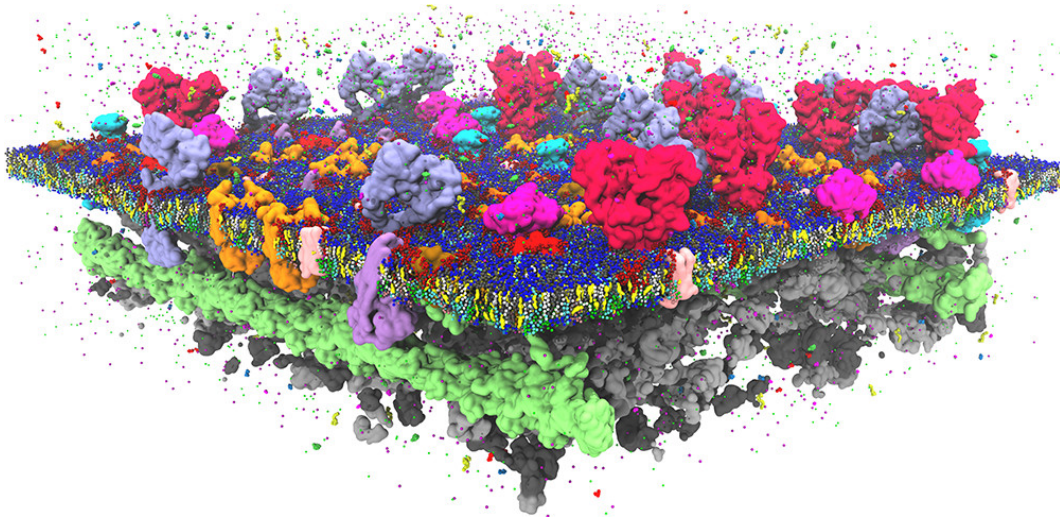
### 1.1 Motivation

The plasma membrane is a dynamic interface that is responsible for directing many cellular processes. It is comprised of many different lipids, proteins, and sugars that continually diffuse, aggregate, and exchange between leaflets. From the motion of these molecules, complex behaviour — such as cell-signalling, membrane trafficking, vesiculation, and apoptosis — arises. Part of this thesis aims to contribute to the current understanding of the molecular origin of both physiological and pathological phenomena in cellular membranes. The other part seeks to improve and add to the software currently available for analysing molecular dynamics simulations of lipid membranes.

### 1.2 The cellular membrane

The cellular membrane, also known as the plasma membrane, provides a physical barrier between a cell and its external environment. It is a quasi-two-dimensional, selectively permeable structure that is primarily comprised of a lipid matrix and embedded proteins (Figure 1.1). This physical barrier, however, is not inert — it is the site of many important cellular processes such as cell-signalling and membrane trafficking.

The mechanical properties of the membrane, such as surface tension and bending modulus, are closely related to cell function. These properties are modulated by the cytoskeleton, a network of protein filaments that attach via integral proteins to the intracellular leaflet of the plasma membrane. Different cell types, and cells at different stages of their life cycle, tune their bending modulus through increasing or decreasing the strength



**Figure 1.1:** The cellular membrane is composed of hundreds of different lipids, crowded with a large variety of embedded as well as peripherally bound proteins. The supporting actin cytoskeleton and an electrochemical gradient across the membrane are important for proper cell functioning. Reprinted with permission from Marrink et al.<sup>1</sup> (<https://pubs.acs.org/doi/10.1021/acs.chemrev.8b00460>). Further permission requests related to this figure should be directed to the American Chemical Society.

of interaction between the cytoskeleton and the plasma membrane.<sup>2</sup>

The cytoskeleton also plays a role in dissipating tensile forces, which leads to the local activation of mechanosensitive ion channels.<sup>3</sup> These channel proteins are responsible for maintaining an electrochemical ion gradient across the plasma membrane, which is necessary for mitochondrial energy production.

### 1.3 Lipids of the cellular membrane

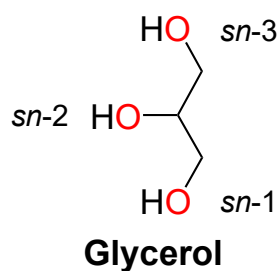
Biological membranes contain hundreds of different lipid species, and there are countless possible lipid compositions. Differences in lipid composition are seen at all levels of biological organisation — from organisms, organs, and tissues, to cell types, organelles, and membrane leaflets. Further, the composition of a leaflet can change in response to external stimuli. It is not known why so many different types of lipid exist in nature, nor is it fully understood how the lipid composition of a membrane determines its function. However, given the metabolic cost of precisely maintaining the lipid composition of biological membranes, there must be some evolutionary advantage to having such complex mixtures. The physiological importance of lipid homeostasis is further evidenced

**Table 1.1:** Levels of classification of lipid structures. Lower levels provide more structural detail about a lipid. The *category* defines the backbone of the lipid, the *class* the headgroup, the *species* the total number of C atoms and double bonds in the acyl tails, and the *subspecies* the number of C atoms and double bonds in each tail.

Level	Example	Added information
<b>Category</b>	Glycerophospholipid	Backbone
<b>Class</b>	Glycerophosphocholine	Headgroup
<b>Species</b>	Phosphatidylcholine (34:1)	Total number of C atoms and double bonds in the acyl tails
<b>Subspecies</b>	Phosphatidylcholine (16:0/18:1)	Number of C atoms and double bonds in each tail

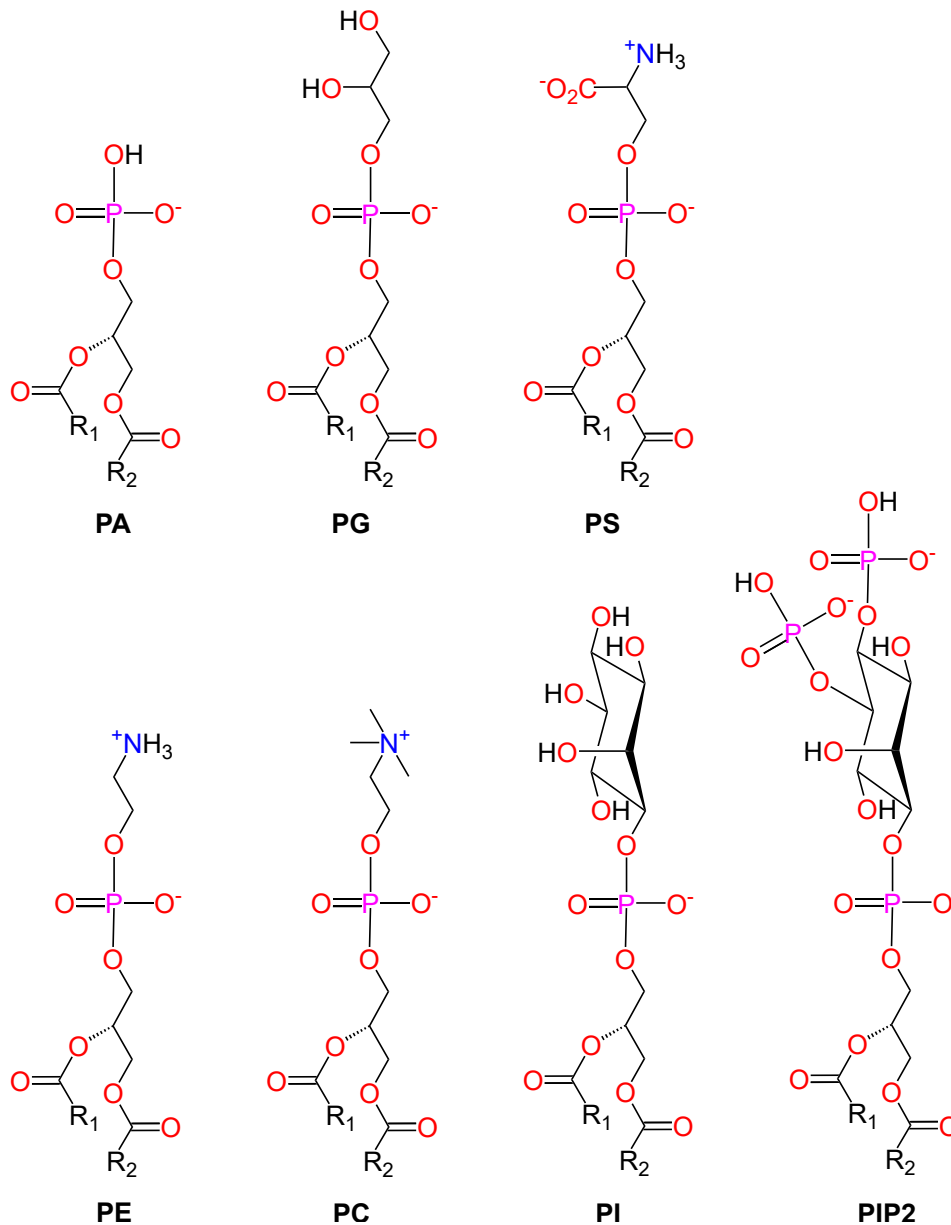
by the large number of diseases associated with changes in membrane lipid composition. With recent advances in lipidomics, we are now beginning to unravel the composition of different cell types at the leaflet level,<sup>4</sup> as well as identify lipid biomarkers for certain pathologies.<sup>5</sup>

Lipids are *amphiphilic* molecules, comprised of hydrophobic and hydrophilic moieties. Many lipids have two fatty acids (acyl tails) joined to a polar headgroup. Lipids can be classified based on the particular combination of their headgroup and acyl tails (Table 1.1).<sup>6</sup> The chemical structure of a lipid is very important; small changes to a lipid's chemistry, such as modifying the acyl tails, can have a drastic impact on the biophysics of the resultant membrane structure.<sup>7</sup> Below I will discuss some of the headgroups and acyl tails commonly found in biological lipids.



**Figure 1.2:** Chemical structure glycerol — the backbone of all glycerophospholipids. Fatty acids replace the *sn-1* and *sn-2* hydroxyl groups, whilst the *sn-3* hydroxyl group is replaced by a phosphate-containing headgroup.





**Figure 1.3:** Chemical structure of some of the glycerophospholipids found in biological membranes. The lipids shown are: phosphatidic acid (PA), phosphatidylglycerol (PG), phosphatidylserine (PS), phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylinositol (PI), and phosphatidylinositol 2,4-bisphosphate (PIP<sub>2</sub>).

### 1.3.1 Glycerophospholipids

Glycerophospholipids are a major category of lipids in the plasma membrane. They are all based on a glycerol backbone (Figure 1.2) and have a phosphosphate-containing head-group at the *sn*-3 position. Fatty acids are connected via ester linkages at the *sn*-1 and *sn*-2 positions. Phosphatidic acid (PA) is the simplest glycerophospholipid — it has a

phosphate headgroup that is singly deprotonated at physiological pH (Figure 1.3). PA is present in biological membranes in only a few mol %, but it plays an important role in cell-signalling and is required for the synthesis of other glycerophospholipid.<sup>8</sup>

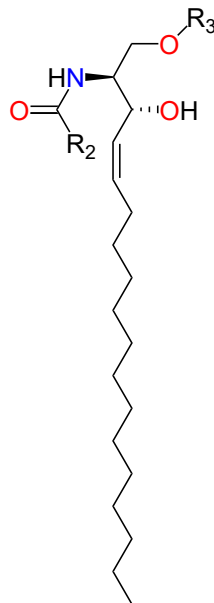
Phosphatidylglycerol (PG) is synthesised from PA by replacing the hydroxyl of the phosphate headgroup with glycerol. Like PA, PG is found only in small quantities in most eukaryotes, although it is present in much larger quantities in some prokaryotes, especially gram-positive bacteria.<sup>9</sup> PG is also present in larger quantities (around 10 mol %) in mammalian lung surfactant,<sup>10</sup> where its net negative charge serves to stabilise the absorption of positively charged surfactant proteins.<sup>11</sup>

Phosphatidylserine (PS) is also synthesised from PA. In this case, the hydroxyl group is replaced by the amino acid serine. PS is the most abundant anionic phospholipid in most eukaryotic cellular membranes, accounting for more than 10 mol % of all lipids. It is located almost entirely in the cytosolic leaflet of the plasma membrane, where it plays an important role in intracellular signalling.<sup>12</sup>

Phosphatidylcholine (PC) and phosphatidylethanolamine (PE) are the two most common phospholipids of the mammalian plasma membrane, together accounting for around half of all phospholipids. They are predominantly distributed across apposing leaflets — PC and PE are most abundant in the extracellular and intracellular leaflets, respectively. Both PC and PE are *zwitterionic*, meaning they have no net charge but contain an equal number of positively- and negatively-charged functional groups. PE is primarily synthesised from PS, and is in turn used to synthesise PC. Both classes of lipid serve important physiological functions, whilst the *ratio* of PC to PE within the plasma membrane is important for proper liver functioning in mammals.<sup>13</sup>

Phosphatidylinositol (PI) is produced by replacing the hydroxyl group of PA by inositol. It is found in the plasma membrane in only very small quantities (less than 1 mol %), and — under healthy conditions — only in the intracellular leaflet. PI is, however, more common in the intracellular leaflets of membrane-bound organelles.<sup>14</sup> In these organelles, the inositol group of PI can undergo enzymatic phosphorylation, producing *phosphoinositides* such as PI 2,4-bisphosphate (PIP<sub>2</sub>). The inositol group of phosphoinositides can be mono-, bi-, or tri-phosphorylated. Phosphoinositides are more common in

the plasma membrane than PI, although they still account for less than 10 mol % of phospholipids. They are, however, a hugely important class of lipid involved in many cellular processes including membrane trafficking,<sup>15</sup> cell proliferation,<sup>16</sup> and regulating the activity of integral membrane proteins.<sup>17</sup>



**Sphingolipid**

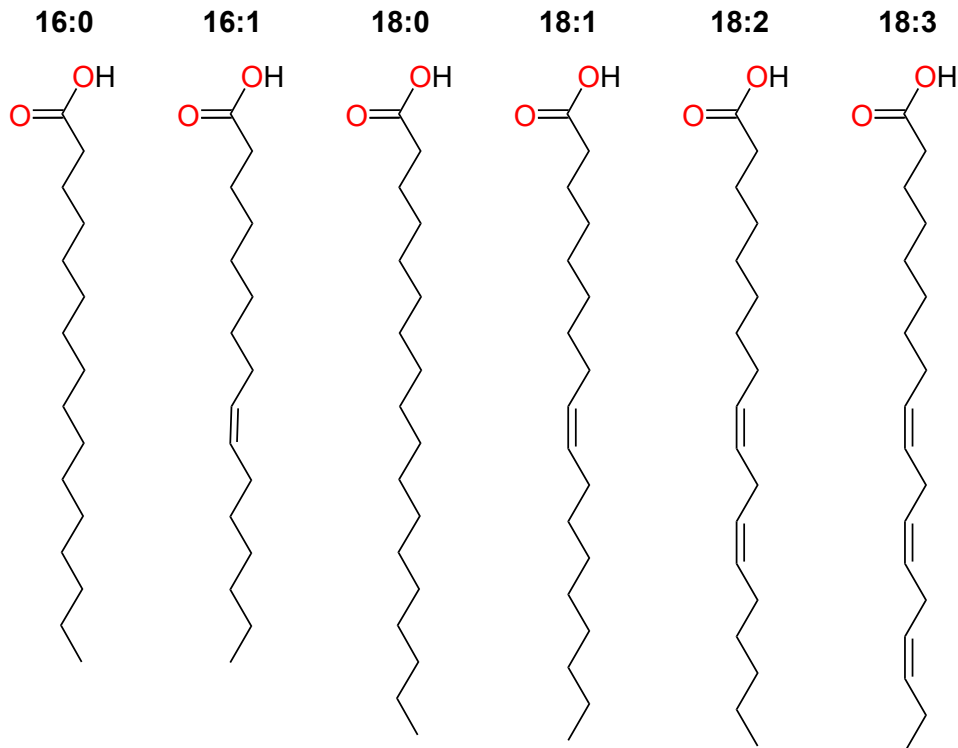
**Figure 1.4:** Chemical structure of a sphingolipid. Sphingolipids have a sphingoid base, a fatty acid for the R<sub>2</sub> group, and hydrophilic headgroup for the R<sub>3</sub> group.

### 1.3.2 Sphingolipids

Sphingolipids are another important category of lipids found in the plasma membrane. They are based on a sphingoid backbone rather than glycerol (Figure 1.4). As sphingolipids are not based on glycerol, they technically do not have *sn-1* and *sn-2* tails. However, the N-linked fatty acid and the acyl tail of the sphingoid base are analogous to the *sn-1* and *sn-2* tails of glycerophospholipids, respectively, and are often referred to as such.

In mammalian cell membranes, most sphingolipids are *sphingomyelins* — sphingolipids with a phosphate-containing headgroup, typically phosphocholine or phosphoethanolamine. Sphingomyelins are therefore also known as sphingophospholipids.

The sphingoid backbone is capable of both accepting and donating hydrogen bonds, whereas the glycerol backbone of glycerophospholipids can act only as a hydro-



**Figure 1.5:** Chemical structure of some of the fatty acids found in biological lipids. Each acyl tail can be described by two numbers XX:Y, where XX is the number of C atoms and Y is the number of C-C double bonds.

gen bond acceptor. This has profound implications on the interactions of sphingolipids with other lipids and proteins. Through such interactions, sphingolipids are especially important in driving the lateral organisation of the plasma membrane; they encourage the aggregation of signalling proteins, which is necessary for the transmission of signals across the membrane.

### 1.3.3 Acyl tails

Many biological lipids are diacyl-phospholipids, meaning they are comprised of two acyl tails and a phosphate-containing headgroup. Each acyl tail may be either *unsaturated* or *saturated* — either with or without C-C double bonds. Figure 1.5 shows some of the fatty acids of lipids found in biological membranes.

The *sn-1* tail of glycerophospholipids is typically saturated or monounsaturated and 16 to 18 C atoms in length. The *sn-1* tail of sphingolipids is typically 18 C atoms in length with one double bond located between the second and third C atoms. The *sn-2* tail in both glycerophospholipids and sphingolipids tends to be longer, although those of sph-

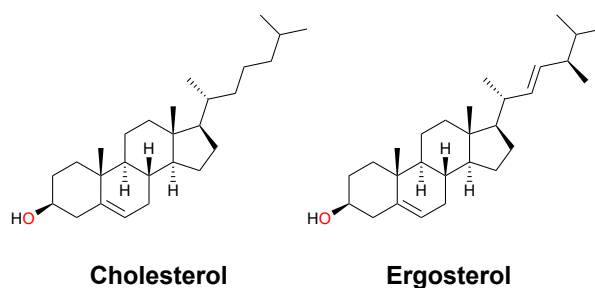
ingolipids can be significantly longer at more than 28 C atoms.<sup>18</sup> The *sn*-2 tail of sphingolipids is usually monounsaturated, whilst that of glycerophospholipids is polyunsaturated. All acyl tails, both *sn*-1 and *sn*-2, of biological phospholipids are comprised of an even number of carbon atoms.

The evolutionary purpose of having multiple double bonds, or of having acyl tails of different lengths, is not well understood. Recent studies, however, have shown the length of the acyl tails is important for interleaflet communication,<sup>7</sup> and that asymmetric lipids — with one short and one long acyl tail — may perform a similar role in certain yeast species to the more metabolically-expensive unsaturated lipids in mammalian plasma membranes.<sup>19</sup>

### 1.3.4 Sterols

Sterols are the most abundant lipid in most eukaryotic cellular membranes, comprising up to 50 mol % of the total lipid content.<sup>4</sup> Cholesterol and ergosterol are the major sterols of mammalian and yeast cellular membranes, respectively. Cholesterol has a flexible hydrocarbon tail, a rigid body comprised of a five-membered ring and three six-membered rings, as well as a hydroxyl headgroup (Figure 1.6). The rigid body of cholesterol is almost planar, and it has two methyl groups orthogonal to the plane. Ergosterol has a similar structure to cholesterol, although its tail has a C-C double bond and an extra methyl group.






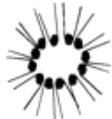
In the cellular membrane, sterols are important both as signalling molecules<sup>20</sup> and for regulating membrane fluidity.<sup>21</sup> In addition, sterols are metabolic precursors to steroid hormones<sup>22</sup> and bile salts.<sup>23</sup>



**Figure 1.6:** Chemical structure of cholesterol and ergosterol, the major sterol of mammalian and yeast plasma membranes, respectively.

## 1.4 Supramolecular lipid structures

In a polar solvent, lipids will self-assemble into supramolecular structures. In these structures, the polar headgroups form an interface with the solvent whilst the acyl tails aggregate in a hydrophobic core. Lipids may form monolayers such as micelles, or bilayers such as lamellar sheets (Figure 1.7).

Example	Molecular shape	Structure
PC		 Lamellar
lyso-PC		 Micelle
PE		 Inverted micelle

**Figure 1.7:** In a polar solvent, lipids may self-assemble into supramolecular structures such as bilayers, micelles, and inverted micelles. The structure formed depends on the relative cross-sectional areas of the hydrophobic and hydrophilic parts of the lipids in the mixture. Adapted from Escribá et al.<sup>24</sup> Copyright (1997) National Academy of Sciences.

The type of lipid structure formed depends on the lipid mixture. Lipids with equal hydrophilic and hydrophobic cross-sectional areas will aggregate into a lamellar bilayer sheet. This includes lipids such as PC, PS, and PG. Lipids with a larger hydrophobic lateral area, such as lyso-lipids that have no *sn*-2 acyl tail, will form a micelle. Other lipids such as PA and PE, which have comparatively smaller headgroups, form an inverted micelle structure. The cellular membrane itself is a large, round bilayer structure. In both experimental and simulation studies of lipid bilayers, however, a lamellar sheet is often used as an approximation of a small membrane patch.

## 1.5 Lipid bilayer phases

Lipid bilayers may exist in a number of phases, each distinguished by the degree of ordering of acyl tails and the rate of diffusion of lipids within a leaflet. The phase of a bilayer depends primarily on the strength of van der Waals (vdW) forces between the acyl tails,

**Table 1.2:** Nomenclature for some of the fatty acids present in biological membranes.  $N_C$  and  $N_{DB}$  refer to the number of carbon atoms and the number of double bonds, respectively.  $\Delta^d$  means there is a double bond at carbon  $d$  along the fatty acid. Melting point temperatures are for single-component bilayers of the corresponding diacylphosphatidylcholine. Biological membranes contain fatty acids from less than 10 carbon atoms to more than 28 carbon atoms in length.

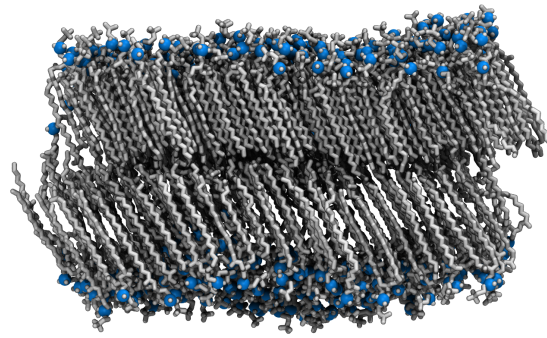
$N_C:N_{DB}$	Name		Melting point (K)
	Common	IUPAC	
16:0	palmitate	<i>n</i> -Hexadecanoate	315
16:1	palmitoleate	<i>n</i> - $\Delta^9$ -Hexadecanoate	-36
18:0	stearate	<i>n</i> -Octadecanoate	55
18:1	oleate	<i>n</i> - $\Delta^9$ -Octadecanoate	-22
18:2	linoleate	<i>n</i> - $\Delta^{9,12}$ -Octadecadienoate	-53
18:3	linolenate	<i>n</i> - $\Delta^{9,12,15}$ -Octadecatrienoate	-60

with stronger interactions leading to a more condensed and less fluid phase. In turn, the strength of vdW interactions between acyl tails depends on the length and degree of unsaturation of the tails.

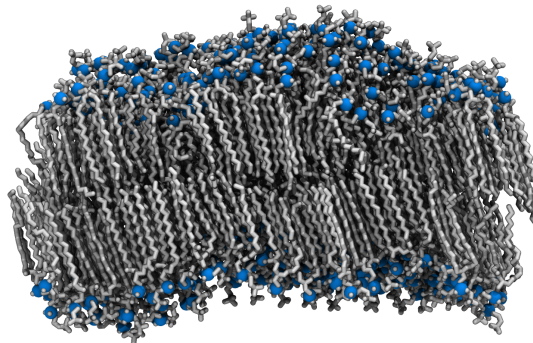
While vdW forces do not depend on temperature *per se*, as the temperature increases the vdW interactions can more easily be disrupted through thermal motions. Therefore, as the temperature is increased, a highly ordered *gel* phase bilayer will transition into a less ordered *liquid-disordered* phase bilayer. The specific temperature,  $T_m$ , at which this transition occurs depends on the length and saturation of the acyl tails. Generally, the longer and more saturated the tail, the higher the transitions temperature (Table 1.2).

**Gel phase** Saturated fatty acids have a high degree of conformational flexibility due to free rotation around all C-C bonds. This means that they can pack very tightly, fully extended, into a condensed and rigid bilayer. For example, at 303 K, distearoylphosphatidylcholine (DSPC) is in the gel-phase (Figure 1.8). This is characterised by the tight packing of elongated acyl tails, with little overlap — or *interdigitation* — between tails in apposing leaflets. Lipids in this phase diffuse slowly, on the order of  $1 \times 10^{-11} \text{ cm}^2 \text{ s}^{-1}$ . To ensure parity between the hydrophobic and hydrophilic cross-sectional areas, the acyl tails are tilted. This phase is therefore often referred to as the tilted gel-phase.

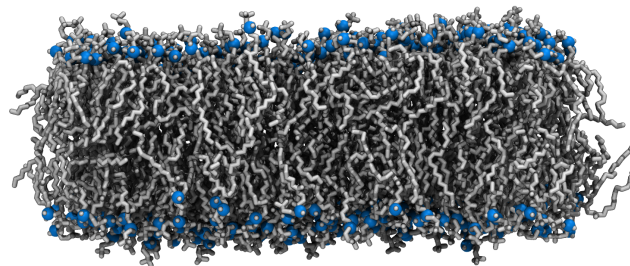
**Liquid-disordered phase** Fatty acids with C-C double bonds are unable to pack together tightly due to the kink at the location of the double bond. In biological mem-



DSPC (18:0/18:0): **gel**



DPPC (16:0/16:0): **ripple**



DOPC (18:1/18:1): **liquid-disordered**

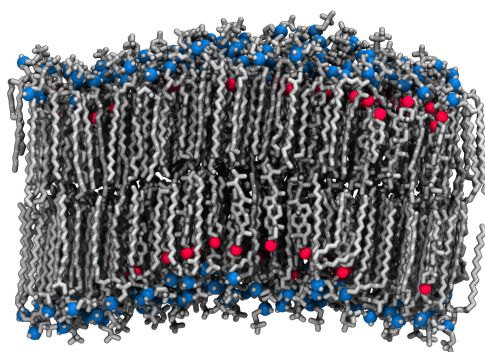
**Figure 1.8:** Phosphatidylcholine bilayers. At 303 K, distearoylphosphatidylcholine (DSPC), dipalmitoylphosphatidylcholine (DPPC), and dioleoylphosphatidylcholine (DOPC) are in the gel, ripple, and liquid-disordered phase, respectively. Blue spheres represent phosphorous atoms; all other atoms are coloured grey.

branes, almost all C-C double bonds are in the cis configuration, which results in a kink in the acyl tail of around  $30^\circ$ . The result is that there is a decrease in strength of the vdW interactions compared to lipids with fully saturated tails. Dioleoylphosphatidylcholine (DOPC), which has the same length acyl tails as DSPC but with a double bond in each tail, is in the liquid-disordered phase at 303 K. Membranes in this phase have lipids with more disordered and more interdigitated acyl tails as well as a faster rate of diffusion, on



the order of  $1 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$ .

**Ripple phase** Shortening the length of acyl tails also reduces the strength of vdW interactions between the acyl tails of neighbouring lipids. Dipalmitoylphosphatidylcholine (DPPC), which has two fully saturated tails 16 C atoms in length, is in the ripple phase at 303 K. In the ripple phase, most regions of the membrane are highly ordered with slow diffusion of lipids, on the order of  $1 \times 10^{-10} \text{ cm}^2 \text{ s}^{-1}$ . Some of these ordered regions are similar to the gel-phase, with a very tight packing of lipids and little interdigitation of acyl tails. Unique to the ripple phase, other highly ordered regions have *fully-interdigitated* acyl tails — the acyl tails of apposing leaflets overlap almost entirely. The fully-interdigitated regions of the membrane are much thinner than the gel-like regions. Thus, the presence of both interdigitated and non-interdigitated regions creates a ripple-like effect in the thickness of the membrane. Lipids at the interface between these two regions are significantly more disordered and have a larger diffusion coefficient than those in the gel-like and the fully-interdigitated regions.<sup>25</sup>



DPPC (16:0/16:0) + ergosterol:  
**liquid-ordered**

**Figure 1.9:** At 303 K, dipalmitoylphosphatidylcholine (DPPC) with 30 mol % ergosterol is in the liquid-ordered phase. Blue and red spheres represent the phosphorous and oxygen atoms of DPPC and ergosterol, respectively; all other atoms are coloured grey.

**Liquid-ordered phase** First described by Ipsen,<sup>26</sup> the liquid-ordered phase has fast diffusing lipids, yet these lipids have ordered acyl tails (Figure 1.9). Pure phospholipid bilayers cannot exist in this phase. Rather, the liquid-ordered phase is induced by the addition of a sterol, such as cholesterol or ergosterol, to membranes with highly saturated lipids. Sterols can intercalate between the tails of neighbouring phospholipids, disrupting the

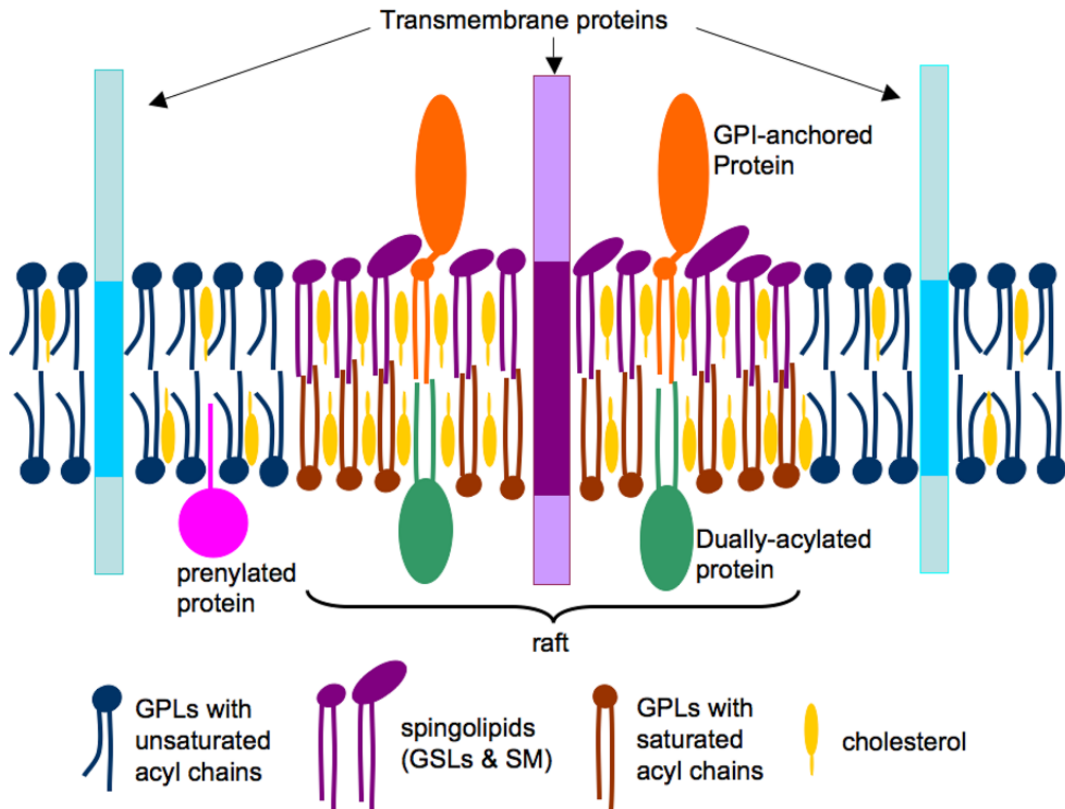
tight packing of their acyl tails. This increases the rate of lateral diffusion of the phospholipids. For example, the addition of 30 mol% ergosterol to a DPPC bilayer at 303 K leads to a diffusion coefficient  $1 \times 10^{-8} \text{ cm}^2 \text{ s}^{-1}$  — two orders of magnitude faster than in the pure DPPC bilayer. At the same time, the rigid body of the sterol ensures the acyl tails of the phospholipids remain highly ordered. As a result, the degree of order and rate of diffusion of lipids in the liquid-ordered phase are both intermediate between those of the gel and liquid-disordered phases. Sterols will also induce order in the acyl tails of unsaturated lipids, but they show a very strong preference for saturated lipids, *especially* saturated sphingolipids.

## 1.6 Lateral heterogeneities

The plasma membrane does not consist of a single homogeneous phase. Instead, there are dynamic, transient lateral heterogeneities that are thought to be important for the proper functioning of the cell. There are several theories on the origin, characteristics, and biological roles of these lateral heterogeneities.<sup>27–42</sup> In the lipid-raft hypothesis,<sup>43</sup> these heterogeneities — known as *lipid-rafts* — are functional platforms for the aggregation of cell-signalling proteins (Figure 1.10). The bulk of the mammalian plasma membrane is in the liquid-disordered phase, with raft regions in the liquid-ordered phase. Lipid-rafts are enriched in saturated glycerophospholipids, sphingolipids, and cholesterol, which preferentially mix with one another through both specific and non-specific interactions. The non-raft regions, on the other hand, are highly enriched in unsaturated glycerophospholipids.

Different proteins sort preferentially into either raft or non-raft regions of the plasma membrane. This sorting occurs by matching the hydrophobic thicknesses of integral proteins and the acyl tails of lipids.<sup>45</sup> As liquid-ordered raft-regions have lipids with more extended acyl tails, proteins with longer hydrophobic alpha-helices will preferentially move into these regions. Likewise, proteins with smaller hydrophobic regions will move in to the liquid-disordered non-raft regions.

Since Simons and Ikonen first described lipid rafts, there has been much debate about their existence and biological importance. This is because lipid rafts are difficult



**Figure 1.10:** Schematic of the plasma membrane with raft and non-raft regions. Unsaturated glycerophospholipids (blue) are enriched in the liquid-disordered non-raft regions. Saturated glycerophospholipids, spingolipids, and cholesterol are enriched in the liquid-ordered raft region. Different proteins preferentially sort into raft or non-raft regions based on the thickness of their hydrophobic regions. Image reprinted from Waheed and Freed,<sup>44</sup> Copyright (2009), with permission from Elsevier.

to detect — they are tens of nm in size and exist only for ms to s,<sup>46</sup> which makes them extremely difficult to study experimentally. As such, membranes that macroscopically phase separate into liquid-ordered and liquid-disordered regions — or *domains* — have been widely used as models of lipid rafts. A minimal model of domain formation is comprised of a high melting temperature lipid, a low melting temperature lipid, and cholesterol.

The most widely studied model of domain formation is a ternary mixture of DOPC, DPPC, and cholesterol. Given this mixture displays macroscopic phase separation, confocal fluorescence microscopy can be used to study domain formation in giant unilamellar vesicles of DPPC, DOPC, and cholesterol.<sup>47</sup> Through this approach, the effect of temperature, pressure, or other molecules on phase separation can be studied. This ap-

proach, however, neglects much of the complexity of the plasma membrane, and leads to macroscopic phase separation rather than nm sized liquid-ordered domains. However, as nanodomains behave surprisingly like genuine phases, model membranes that display macroscopic phase separation may still provide insight into the behaviour of their biological counterparts.<sup>48</sup>

## 1.7 Lipid-associated pathologies

Many pathologies are associated either with specific lipid species or with changes in the lipid composition of the cellular membrane. This includes neurodegenerative disorders such as Alzheimer's disease<sup>49</sup> and Parkinson's disease,<sup>50</sup> osteoporosis,<sup>51</sup> atherosclerosis,<sup>52</sup> cancer,<sup>53</sup> as well as bacterial,<sup>54</sup> viral,<sup>55</sup> and fungal<sup>56</sup> infections. In Chapter 3 of this thesis I will provide a molecular level description of how the oxidation of cholesterol disrupts the formation of liquid-ordered domains, which is associated with numerous pathologies. Below I will briefly discuss a few of the pathologies with which lipids are associated. While the following examples merely scratch the surface of lipid related pathologies, they do provide some indication of the range of diseases and disorders for which lipids play a role in the pathogenesis.

**Lipids in host-pathogen interactions** The high-density of receptor proteins in lipid rafts means these structures can act as binding sites for some pathogens. HIV, for example, binds to the CD<sub>4</sub> receptors in lipid rafts of activated T cells.<sup>57</sup> This binding has two effects: i) the immune response is stimulated, providing more T cells for HIV to bind to and ii) the infected cell releases molecules that encourage the formation of lipid rafts in non-infected cells, thus providing more platforms for other HIV particles to bind to.

In other cases, pathogens exploit the role of lipids in cell-signalling pathways. The *Mycobacterium tuberculosis* bacterium, for instance, is quickly engulfed by host macrophages. Usually, once a foreign particle has been engulfed, a macrophage would incorporate bile salts into its interior, with these bile salts then breaking down the pathogen. *M. tuberculosis*, however, binds to intracellular PI 3-phosphate lipids, which blocks the pathway responsible for recruiting bile salts to the cell's interior.<sup>58</sup> In this way, *M. tuberculosis* can lie dormant in macrophages for many decades, clinically undetectable yet still

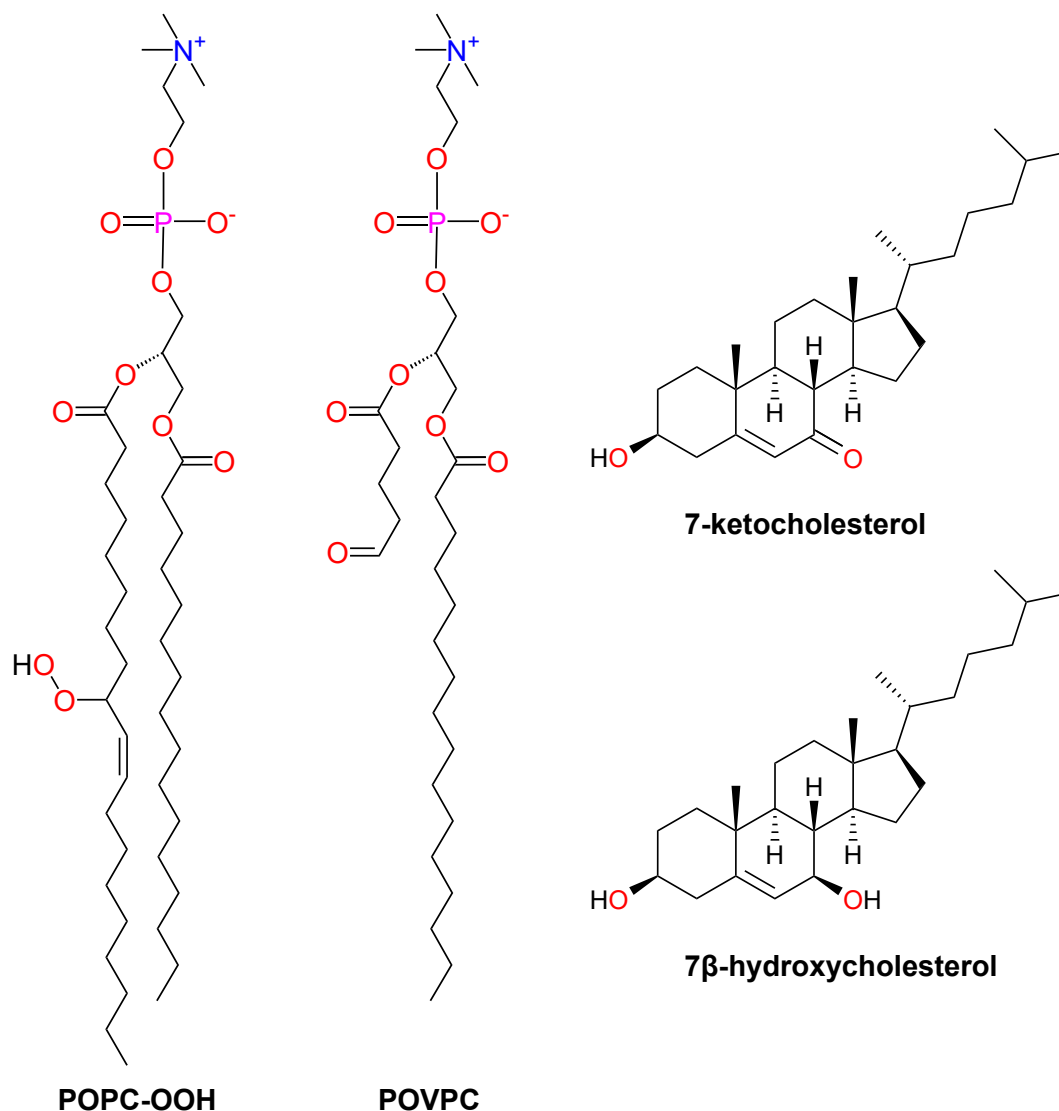
replicating. Like many pathogens, *M. tuberculosis* also metabolises host lipids as a source of nutrients.

**Pathologies related to phospholipid oxidation** Oxidative stress can lead to the non-enzymatic oxidation of phospholipids and sterols (Figure 1.11). Increased levels of oxidised phospholipids in the cellular membrane are associated with numerous pathologies. Hydroperoxides, for example, are thought to play a causal role in myocardial ischemia, whilst the saturated aldehyde 1-palmitoyl-2-(5-oxovaleroyl)-sn-glycero-3-phosphocholine (POVPC) is thought to activate apoptotic cell-signalling pathways.<sup>59</sup>

**Pathologies related to sterol oxidation** When subjected to oxidative stress, cholesterol may react with free radicals, such as reactive oxygen species, to form ring-oxidised sterols, most commonly 7-ketocholesterol and 7 $\beta$ -hydroxycholesterol.<sup>60</sup> Under oxidative stress, the unsaturated hydrocarbon lipid tails of phospholipids are more-readily oxidised than cholesterol. However, because the clearance of oxysterols is less efficient than that of oxidised phospholipids, oxysterols are more prevalent than their phospholipid counterparts in the plasma membrane.<sup>61</sup>

Oxysterols are present in the plasma membrane under physiological conditions; their physiological roles include the removal of excess cholesterol from the body, as well as roles in pro-inflammatory signaling and the metabolism of lipids.<sup>60</sup> However, elevated levels of oxysterols are associated with a number of diseases, such as Type 2 diabetes, Huntingtons disease and PKAN.<sup>60-63</sup>

**Lipid composition in cancerous tumours** Cancerous tumours exist in a hypoxic environment with reduced access to nutrients. This leads to changes in lipid metabolism, which in turn leads to changes in the lipid composition of the cellular membranes.<sup>64</sup> Some of these changes, such as elevated levels of PS lipids, are so significant that they can be used as biomarkers for cancer. PS, which is usually confined to the intracellular leaflet, becomes enriched in the extracellular leaflet. This surface exposure of PS lipids protects the cancerous environment against the response of the immune system.<sup>65</sup> Chemotherapy and radiotherapy increase the levels of exposed PS in the cancer microenvironment, further increasing the immunosuppressive capability of the tumour.<sup>66</sup> However, radiotherapy in combination with an antibody that targets PS lipids may improve patients



**Figure 1.11:** Oxidised lipids are associated with a number of pathologies. POPC-OOH and 1-palmitoyl-2-(5-oxovaleroyl)-sn-glycero-3-phosphorylcholine (POVPC) are hydroperoxide and aldehyde derivatives, respectively, of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC). 7-ketocholesterol and 7β-hydroxycholesterol are oxysterols produced by the non-enzymatic oxidation of cholesterol.

outcomes for those with late-stage cancer.<sup>67</sup>

## 1.8 Molecular dynamics simulations

Molecular dynamics simulations have been used since the second half of the 20<sup>th</sup> century to investigate how macroscopic properties of materials emerge from the interactions of their constituent particles. The atomic scale detail of the structural and dynamical properties of materials afforded by MD simulations cannot be resolved using current ex-

perimental techniques. As such, MD simulations have become widely used alongside experiments to provide complementary insight into the structure and dynamics of condensed matter systems.

### **1.8.1 The early history of computer simulation**

During World War II, electronic computers in the National Laboratories of the United States were used solely for the study of classified issues of national security.<sup>68</sup> At this time, good analytical descriptions of the behaviour of dilute gases and crystalline solids had been developed. The study of dense liquids, however, relied upon the use of macroscopic mechanical models, such as a set of ball bearings. Thus, when the use of electronic computers was later permitted for non-classified research, one of the very first applications was the study of condensed liquids. One particular question that lacked scientific consensus was whether there is a liquid-to-solid phase transition in a system of hard spheres as the density of the system increases. By 1957, this question had been answered by means of computer simulations.

In 1953, Metropolis et al. published a computational Monte Carlo method for simulating a system of hard spheres.<sup>69</sup> This involved considering each hard sphere in turn and applying a random translation to the particle's position. If the translation results in a decrease in the potential energy of the system, the translation is accepted as the new position of the particle. Otherwise, the translation is accepted with a probability of the change in potential energy weighted by the Boltzmann factor. Whilst this method was first applied to hard spheres, it is a general approach and one of the most important algorithms developed in the 20<sup>th</sup> Century.

The Monte Carlo method is often referred to as Metropolis Monte Carlo, named after the first author of the seminal paper. However, it was Edward Teller who had the idea to weight the acceptance/rejection criterion by the Boltzmann factor, his student Marshall Rosenbluth who further developed the theory, and Arianna Rosenbluth who implemented the algorithm on MANIAC I.<sup>70-72</sup> Nicholas Metropolis was an author on the paper as he was director of the Institute for Computing Research at Los Alamos National Laboratories; he was first author because his name is first alphabetically.<sup>73</sup>

The following year, in 1954, Arianna and Marshall Rosenbluth performed further

Monte Carlo simulations of hard spheres to look for a liquid-to-solid phase transition.<sup>74</sup> They studied systems of 256 hard spheres initially positioned on a face-centered-cubic lattice, at 20 different volumes. For each system, 100 trials moves were performed for each of the 256 particles. The authors found no evidence of a phase transition, although they concluded that the Monte Carlo method might not be powerful enough to study such phenomena.

Berni Alder played an important role both in the early development of computer simulations and in resolving the question of whether there is a phase transition in a system of hard spheres. Prior to the publication of the Monte Carlo method by Metropolis et al., Berni Alder, Stan Frankel, John Kirkwood, and V. A. Lewinson had developed a less-general approach to Monte Carlo simulations of hard spheres.<sup>69,75</sup> However, Alder's PhD supervisor, Kirkwood, was uncomfortable with the theoretical foundations of the method and so the group did not publish their findings.<sup>76,77</sup>

In 1955, Berni Alder and Thomas Wainwright began to think about implementing Newton's equations of motion on electronic computers to study the approach to equilibrium of a system of hard spheres. This is more computationally expensive than the Monte Carlo method, but can provide information on both equilibrium and time-dependent properties. The first simulations were performed on UNIVAC, but this was not powerful enough to provide sufficient statistics — in a system of 100 hard spheres, only around 100 collisions were observed per hour on UNIVAC.<sup>77</sup>

It was not until 1956 that Alder and Wainwright had access to the IBM-704, which was around 200 times faster than the UNIVAC. By running the simulations over an entire summer, they had enough simulation time to determine the equilibrium properties of the simulated systems. In 1957, Alder and Wainwright published their landmark paper on molecular dynamics simulations.<sup>78</sup> This paper provided both the first description of molecular dynamics and a demonstration of a first-order liquid-to-solid phase transition in a system of hard spheres. In the same issue of *J. Chem. Phys.*, Bill Wood and Jack Jacobson presented their findings after reproducing the Monte Carlo calculations of hard spheres earlier performed by Rosenbluth and Rosenbluth.<sup>79</sup> Wood and Jacobson showed that the simulations performed by Rosenbluth and Rosenbluth, with only 100



trial moves per hard sphere, had poor exploration of phase space. Wood and Jacobson's further analysis suggested that a first-order transition is indeed likely. This agreement between Monte Carlo and molecular dynamics eventually convinced the community of the existence of the phase transition, and in doing so demonstrated the power of computer simulation as a scientific research method.

Later, in 1959, Alder and Wainwright published a more detailed account of their method, and over the next 20 years Alder continued to develop the theoretical foundations of molecular dynamics.<sup>80</sup> However, it must be remembered that these researchers could not have performed the simulations without the — predominantly female — programmers who implemented the algorithms on early computers such as MACIAC. In their landmark 1957 paper, Alder and Wainwright acknowledge Shirley Campbell and Mary Shephard for their work on implementing the molecular dynamics algorithm.<sup>78</sup> Mary Ann Mansigh, meanwhile, worked with Alder for over two decades, creating a highly optimised molecular dynamics engine called STEP.<sup>81</sup>

Since the 1950s, the development and optimisation of more sophisticated Monte Carlo and molecular dynamics methods, along with the vast increase in computing power, has led to an explosion in the use of computer simulations in scientific research. For biological problems alone, Monte Carlo and molecular dynamics simulations are now routinely used for studying proteins,<sup>82</sup> lipids,<sup>39</sup> carbohydrates,<sup>83</sup> RNA,<sup>84</sup> and DNA,<sup>85</sup> as well as the interactions between these biomolecules.<sup>1,86–88</sup> Below I will provide a brief overview of computer simulations of lipid membranes from the past few decades.

### **1.8.2 A brief overview of lipid membrane simulations**

Alder and Wainwright's molecular dynamics simulations involved perfectly elastic collision between hard spheres. By the 1960s, however, more realistic potentials involving both repulsive and attractive terms were employed to simulate physical matter more accurately. In 1960, Gibson et al. simulated the crystal structures of copper,<sup>89</sup> and four years later Rahman used the Lennard-Jones potential for his simulations of liquid argon.<sup>90</sup> Then, in the 1980s, the first simulations of lipid membranes appeared.

The first simulation of an approximate lipid membrane was performed by Kox et

al in 1980.<sup>91</sup> The authors modelled a monolayer of 90 lipids, each with a single seven C chain and a headgroup that was constrained to move in the  $xy$ -plane. While simple, this model managed to reproduce the expected first-order phase transition from the gaseous to the liquid-expanded monolayer phase.

In 1982, van der Ploeg and Berendsen reported on the first molecular dynamics simulations of a lipid bilayer.<sup>92</sup> The bilayer was comprised of 32 lipids in total, with the headgroups restrained along the  $z$ -axis using a harmonic potential to mimic the lipid-water interface. From the 80 ps trajectory, the authors calculated the deuterium order parameter, which is a measure of how far the vectors made by the C-H bonds in the acyl tails deviate from the membrane normal. The results were in excellent agreement with <sup>2</sup>H NMR spectroscopy measurements.

It was not until the end of the 1980s that simulations of fully-hydrated membranes were possible.<sup>93,94</sup> These simulations, however, were still limited to ps timescale. Nonetheless, measurements from these simulations led to improvements in the theoretical models of hydration forces at the membrane-water interface.

By the mid 1990s, there was interest in using a particle-mesh Ewald summation to account for long-range electrostatics, but introducing these interactions into the calculations would require the reparameterisation of a force field.<sup>95</sup> Since this time, a number of atomistic and coarse-grained force fields have been developed for the simulation of lipids and other components of the plasma membrane.<sup>96-105</sup>

Today, molecular dynamics simulations provide a powerful means of studying the cellular membrane, especially since the time and length scales accessible to simulation and experiment are beginning to converge. In recent years, we have seen simulations of bilayers with realistic lipid compositions,<sup>4,106-108</sup> including those that incorporate integral proteins,<sup>109</sup> as well as simulations of an entire virion.<sup>110</sup> One of the next challenges for the field will be to include models of the extracellular matrix, the cytoskeleton, and the cytosol. In this direction, Harker-Kirschneck et al.<sup>111</sup> recently created a coarse-grained model of ECSRT-III — a complex of cytosol proteins involved in membrane remodelling — and used this model to study cell division in archaea.<sup>112</sup>

## 1.9 Overview of this thesis

In the following chapter I will discuss the theoretical foundations of MD simulations, along with some of the practical aspects of performing simulations. I will also briefly discuss the use of machine learning methods for analysing MD simulations. In Chapter 4 I will present my investigation into the effect of cholesterol oxidation on domain formation in model membranes. In Chapter 4 I will look at the interactions between cholesterol and sphingomyelin, identifying specific modes of interaction that may drive their preferential mixing biological membranes. In Chapter 5 I will present LiPyphilic, a Python package I have created for the analysis of MD simulations of lipid membranes. In the final chapter this thesis I will summarise its key points and briefly discuss potential avenues for future research.

## Chapter 2

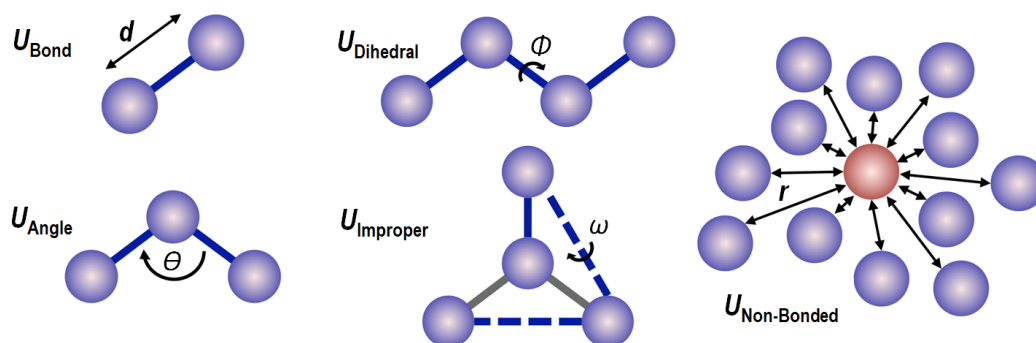
# Methods

### 2.1 Molecular dynamics

Molecular dynamics simulations involve the iterative, numerical integration of Newton's equations of motion to propagate the coordinates and velocities of a system of interacting particles through time. From the resultant trajectory, statistical mechanics can be used to calculate any thermodynamic observable. The calculated properties of the system can be used to assess the accuracy of the simulated model, provide new insight into experimental findings, or test and refine theoretical models. MD simulations thus provide a third way of undertaking scientific research, alongside the two original pillars of experiment and theory. In this chapter, I will discuss some of the important theoretical and practical aspects of MD simulations.

#### 2.1.1 Force field

For classical MD simulations, the ultrafast electronic motions of electrons are assumed to average out over the timescale of nuclear motions. Each atom is therefore treated as a point particle, and the microscopic state of the system can be described solely as a function of the positions and momenta of these particles. The equilibrium distribution of atomic positions and momenta is determined entirely by a force field — a set of interatomic potentials and their parameters that define the interactions between atoms. This set of potentials is comprised of non-bonded and bonded contributions to the total po-



**Figure 2.1:** Schematic of bonded and non-bonded terms in a classical molecular dynamics forcefield for biomolecular simulations. Adapted with permission from Riniker.<sup>114</sup> Copyright 2018 American Chemical Society.

tential energy,  $U(\mathbf{r})$ :

$$U(\mathbf{r}) = U_{\text{Non-Bonded}}(\mathbf{r}) + U_{\text{Bonded}}(\mathbf{r}) \quad (2.1)$$

**Non-Bonded potentials** In biomolecular simulations, non-bonded interactions are typically modelled as a combination of Lennard-Jones,  $U_{\text{LJ}}(\mathbf{r})$ , and Coulomb potentials between pairs of atoms:

$$U_{\text{Non-Bonded}}(\mathbf{r}) = U_{\text{LJ}}(\mathbf{r}) + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.2)$$

In the Coulomb potential,  $r_{ij}$  is the distance between atoms  $i$  and  $j$ , and  $\epsilon_0$  is the permittivity of free space.  $q_i$  and  $q_j$  are the fixed-point *partial* charges of atoms  $i$  and  $j$ , which are non-integer values of the positive elementary charge,  $q_e$ . These partial charges reflect the asymmetric distribution of electron density across covalent bonds; thus, the partial charge of an atom depends on its local chemical environment. For example, the oxygen and hydrogen atoms of the TIP3P water model have partial charges of  $-0.834 q_e$  and  $0.417 q_e$ , respectively, to account for the greater electronegativity of oxygen.<sup>113</sup>

The Lennard-Jones potential is a phenomenological model that captures both the attractive, long-range, van der Waals forces and the repulsive, short-range forces due to overlapping electron orbitals. Most force fields employ the Lennard-Jones 12-6 potential,  $U_{\text{LJ}}^{12,6}(\mathbf{r})$ , which provides a good compromise between accuracy and computational efficiency. For two atoms,  $i$  and  $j$ , separated by a distance of  $r_{ij}$ , the Lennard Jones 12-6

potential is given by:

$$U_{\text{LJ}}^{12,6}(\mathbf{r}) = \sum_{i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.3)$$

$\sigma_{ij}$  defines the distance at which the potential is equal to zero; for  $r_{ij} < \sigma_{ij}$  the repulsive  $r_{ij}^{-12}$  term dominates, whereas for  $r_{ij} > \sigma_{ij}$  the attractive  $-r_{ij}^{-6}$  term dominates. There is a single minimum in the Lennard-Jones 12-6 potential at  $r_{ij} = 2\sigma_{ij}^{1/6}$ , which determines the equilibrium distance between two atoms. The value of  $\epsilon_{ij}$  determines the depth of the potential well at its minimum, and thus defines the strength of the interaction between the two atoms.

Values of  $\sigma$  and  $\epsilon$  are defined for each *atom type* based on the element and its local chemical environment. For interactions between different atoms types, in which  $\sigma_i \neq \sigma_j$  and  $\epsilon_i \neq \epsilon_j$ , the equilibrium separation distance and depth of the potential well are usually calculated using the Lorentz-Berthelot mixing rules. This means using the arithmetic mean of the van der Waals radii and the geometric mean of the interaction strengths:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \quad (2.4)$$

$$\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j} \quad (2.5)$$

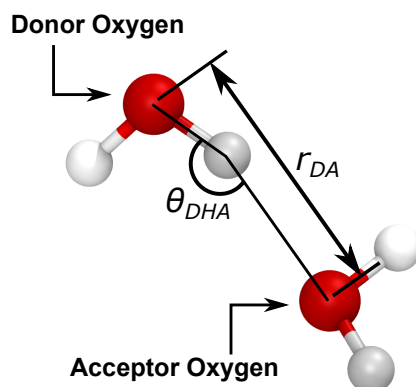
**Bonded potentials**  $U_{\text{Non-Bonded}}(\mathbf{r})$  describes the interactions between pairs of atoms that are not covalently bonded to one another. For atoms that are within the same molecule, and connected via three or fewer bonds, these non-bonded interactions are typically not accounted for or are scaled down. Instead, atoms that are near one another on a molecular graph have their interactions modelled by the bonded potentials,  $U_{\text{Bonded}}(\mathbf{r})$ . The bonded interactions are typically comprised of both harmonic and cosine potentials:

$$\begin{aligned}
U_{\text{Bonded}}(r) = & \underbrace{\frac{1}{2} \sum_{ij} k_r (r_{ij} - r_0)^2}_{\text{covalent bonds}} + \underbrace{\frac{1}{2} \sum_{ijk} k_\theta (\theta_{ijk} - \theta_0)^2}_{\text{bond angles}} \\
& + \underbrace{\frac{1}{2} \sum_{ik} k_u (u_{ik} - u_0)^2}_{\text{Urey-Bradley}} \\
& + \underbrace{\sum_{ijkl} \sum_{n=1}^N k_{\phi,n} [1 + \cos(n\phi_{ijkl} - \delta_n)]}_{\text{bond dihedrals}} + \underbrace{\sum_{ijkl} k_\omega (\omega_{ijkl} - \omega_0)^2}_{\text{improper dihedrals}}
\end{aligned} \tag{2.6}$$

A harmonic potential with a force constant,  $k_r$ , is used to represent the stretching of covalent bonds. The force constant is usually large, on the order of  $100 \text{ kcal mol}^{-1} \text{ nm}^{-2}$ , although this depends on the order of the bond. A large force constant ensures the instantaneous bond length,  $r_{ij}$ , does not deviate far from the equilibrium values,  $r_0$ . Usually, covalent bonds can neither form nor break during the course of an MD simulation; the connectivity between atoms must be defined before the simulation begins. If simulating chemical reactions, however, covalent bonds can be represented by a sum of exponential potentials to model bond breaking and formation in classical MD simulations.<sup>115</sup>

Bond angle bending is also described by harmonic potentials. However, the force constants,  $k_\theta$ , are smaller than those for bond stretching because it is easier to deform a bond angle from its equilibrium value,  $\theta_0$ . Some force fields include a harmonic Urey-Bradley potential, which is a cross term that accounts for the interdependence between bond stretching and angle bending. This defines an equilibrium distance,  $u_0$ , between the 1,3 atoms in a bond angle. The Urey-Bradley term enables better reproduction of vibrational spectra with little additional computational cost.<sup>114,116</sup>

A bond dihedral angle is defined for each set of four simply connected atoms,  $ijkl$ , in a molecule. Each triplet of atoms,  $ijk$  and  $jkl$ , defines a half-plane and the angle of intersection,  $\phi_{ijkl}$ , of these half-planes is the dihedral angle. The potential energy landscape of bond dihedrals is more complex than that of bond stretching and bond angle bending. Bond dihedrals are typically described by a sum of cosine potentials, with  $N$  minima each



**Figure 2.2:** Hydrogen bond between two water molecules. Hydrogen bonds can be identified based on the donor-hydrogen-acceptor angle,  $\theta_{DHA}$ , and the donor-acceptor distance,  $r_{DA}$ .

at a phase-shift of  $\delta_n$  with a force constant of  $k_{\phi,n}$ . Improper dihedral angles are dihedral angles formed by a set of four non-contiguous atoms, and are useful for maintaining the planarity of aromatic rings or other conjugated systems. Improper dihedral angles are modelled using harmonic potentials, with a large force constant,  $k_{\omega}$ , ensuring the angle,  $\omega_{ijkl}$ , does not deviate far from its equilibrium value,  $\omega_0$ .

**Hydrogen bonds** Historically, force fields included an explicit term for hydrogen bonds. However, modern force fields implicitly represent hydrogen bonds via the non-bonded electrostatic and van der Waals interactions between atoms. In the analysis of simulation trajectories, hydrogen bonding can be detected via simple geometric criteria, although more sophisticated approaches provide greater accuracy.<sup>117</sup> Typically, two atoms are considered to be hydrogen bonded if the donor-hydrogen-acceptor angle,  $\theta_{DHA}$ , is greater than  $150^\circ$  and the donor-acceptor distance,  $r_{DA}$ , is less than  $3.5 \text{ \AA}$  (Figure 2.2).<sup>118</sup>

**Parameterisation** A force field defines both the functional form of the interatomic potentials, described above, and the values of the parameters in the potentials, i.e the force constants, equilibrium values, partial charges,  $\epsilon$  and  $\sigma$ . The parameterisation of a force field is important if reality is to be simulated accurately, although it is highly non-trivial. Numerous force fields for biomolecular simulations have been developed over the past few decades, and development continues with parameters for specific biomolecules being optimised via both quantum mechanical calculations and alignment with experimental data.<sup>96–105</sup> The work in this thesis uses the all-atom CHARMM36 and coarse-grained



MARTINI 2 force fields.<sup>98,102,103</sup>

It should be noted that the Lennard-Jones potential involves a sum over *pairs* of atoms; no many-body terms, which may account for more than 10% percent of the potential energy in the liquid phase,<sup>116</sup> are included. In practice, however, force fields are parameterised such that many-body interactions are implicitly taken into account through the pairwise interactions. The potentials therefore do not correspond to the actual pair potential between atoms — instead, they are *effective* pair potentials. Unlike the true pair potential, these effective pair potentials depend on system variables such as pressure and temperature.<sup>119</sup> This means that a force field parameterised with a specific set of system variables may not generalise well to, for example, higher temperatures or pressures. Nonetheless, empirical force fields for lipids can accurately reproduce the transitions between the gel, ripple, and fluid phases of single- and multi-component mixtures.<sup>120,121</sup>

### 2.1.2 System creation

Careful attention must be paid to the construction of a system to be simulated. Molecular dynamics simulations typically aim to sample configurational space near equilibrium; initialising a system to be near its equilibrated state can save significant simulation time. This may mean using the structure of a protein generated from crystallographic or cryo-EM data, or generating supramolecular structures using Packmol<sup>122</sup> or PolyPly.<sup>123</sup> All systems in this thesis were created using the CHARMM-GUI Membrane Builder<sup>124</sup> or MARTINI Maker.<sup>125</sup> These tools generate planar lipid bilayers of user-specified mixtures. First, spheres are placed in the system in the approximate locations of the lipid headgroups. Each sphere is in turn replaced by a molecule from a library of equilibrated lipid structures. Checks are then made to ensure acyl tails are not threaded through the tetracyclic rings of sterols.

### 2.1.3 Gradient descent

Even with careful preparation, the initialisation of atomic coordinates can lead to artefacts such as overlapping atoms. Performing dynamics with overlapping atoms will lead physical instabilities due to very large interatomic forces. To remove such artefacts, a system can be driven towards a local minimum in potential energy via gradient descent. Gra-

gradient descent involves iteratively updating a differentiable function by taking small steps in the direction opposite the gradient of the function at its current point. In the GRO-MACS<sup>126</sup> implementation of gradient descent minimisation, the coordinates at the next step,  $\mathbf{r}_{n+1}$ , are given by:

$$\mathbf{r}_{n+1} = \mathbf{r}_n + \frac{\mathbf{F}_n}{\max(|\mathbf{F}_n|)} h_n \quad (2.7)$$

where  $\mathbf{r}_n$  are the atomic coordinates at the current step,  $n$ .  $\mathbf{F}_n$  is the net force acting on each atom, which is given by the gradient of the potential energy at the current step.  $\max(|\mathbf{F}_n|)$  is the maximum absolute force acting on any atom and  $h_n$  is the maximum displacement of any atom from step  $n$  to step  $n + 1$ . At each step  $n + 1$ , if the potential energy is decreased compared to the previous step then the new coordinates are accepted and  $h_{n+1} = 1.2h_n$ . Otherwise, the new coordinates are rejected and  $h_{n+1} = 0.2h_n$ . This process continues either for a pre-defined number of iterations or until  $\max(|\mathbf{F}_n|)$  is smaller than a specified value.

#### 2.1.4 Initialise velocities

Before dynamics can be simulated, the atomic velocities must first be initialised. At equilibrium, the speed of atoms in three-dimensions will follow the Maxwell-Boltzmann speed distribution,  $f(v)$ :

$$f(v) = 4\pi \left[ \frac{m}{2\pi k_B T} \right]^{3/2} v^2 \exp\left( \frac{-mv^2}{2k_B T} \right) \quad (2.8)$$

Atomic velocities could be initialised uniformly such that  $mv^2 = \frac{3}{2}k_B T$ , and over the course of the simulation the distribution would converge toward  $f(v)$ . However, it is common to generate component-wise velocities using the Maxwell-Boltzmann distribution for a single spatial dimension  $i$ ,  $P(v_i)$ :

$$P(v_i) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left( \frac{-mv_i^2}{2k_B T} \right) \quad (2.9)$$

To prevent drift of the system the velocities are shifted such that  $\langle v_i \rangle = 0$ , and then scaled to reflect to desired temperature:  $m\langle v_i^2 \rangle = \frac{1}{2}k_B T$ .

### 2.1.5 Numerical integration of the equations of motion

The main part of a MD simulation involves iteratively updating the atomic coordinates and velocities in order to simulate the time-evolution of the system. The velocity,  $\mathbf{v}$ , and position,  $\mathbf{r}$ , of atom  $i$  at a future time  $t + \Delta t$  can be calculated via the following definite integrals:

$$\mathbf{v}_i(t + \Delta t) = \mathbf{v}_i(t) + \int_t^{t+\Delta t} \mathbf{a}_i(t) dt \quad (2.10)$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \int_t^{t+\Delta t} \mathbf{v}_i(t) dt \quad (2.11)$$

where the accelerations,  $\mathbf{a}_i(t)$ , can be calculated via Newton's second law of motion:

$$\mathbf{a}_i(t) = \frac{\mathbf{F}_i(t)}{m} \quad (2.12)$$

This relates the net force,  $\mathbf{F}_i$ , acting on a particle at time  $t$  to its mass,  $m$ , and its instantaneous acceleration. The force itself is calculated as the negative of the gradient of the potential:

$$\mathbf{F}_i = -\nabla U(\mathbf{r}_i) \quad (2.13)$$

Once the force — and thus acceleration — is known, the velocity and position at time  $t + \Delta t$  can be calculated according to Equations 2.10 and 2.11. However, for a system of interacting atoms, comprised of many different chemical species, it is impossible to solve these integrals analytically. Instead, finite difference methods are used to numerically integrate Equations 2.10 and 2.11. A commonly used integrator is the velocity Verlet central difference method:<sup>127</sup>

$$\mathbf{v}(t + \frac{\delta t}{2}) = \mathbf{v}(t) + \frac{\delta t}{2} \mathbf{a}(t) \quad (2.14)$$

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \delta t \mathbf{v}(t + \frac{\delta t}{2}) \quad (2.15)$$

$$\mathbf{v}(t + \delta t) = \mathbf{v}(t + \frac{\delta t}{2}) + \frac{\delta t}{2} \mathbf{a}(t + \delta t) \quad (2.16)$$

where the force calculation is performed between Equations 2.15 and 2.16. This algorithm has several attractive properties — it is numerically stable, time-reversible, permits longer timesteps, and is easy to implement.<sup>116</sup>

For conservation of energy, the velocity Verlet algorithm requires that atomic velocities and accelerations are constant over the timestep used. The timestep is thus chosen to be as large as possible, whilst remaining approximately an order of magnitude smaller than the time taken for the highest frequency motions to occur, which is usually the vibration of C-H bonds.<sup>128</sup> The vibrational frequency of C-H bond stretching is on the order of  $1 \times 10^{14}$  Hz, which puts an upper limit on the timestep of around 1.0 fs. However, these high frequency motions can be constrained to their equilibrium values using a modification to the velocity Verlet integrator. Such constraints allow for the use of a larger timestep without breaking the law of conservation of energy, thus enabling longer simulation times. LINCS and SHAKE are two of the most commonly used constraint algorithms, and can be used to permit a timestep of up to 2 fs in all-atom simulations.<sup>129,130</sup> The work in this thesis uses LINCS, as implemented in GROMACS,<sup>126</sup> to constrain high-frequency motions.

### 2.1.6 Thermostats and barostats

The methodology described thus far will simulate a system in the microcanonical ensemble. However, as experiments are typically performed under constant temperature and pressure, it is useful to introduce the concept of thermostats and barostats for molecular simulation. Thermostats and barostats maintain the temperature and pressure, respectively, at specified values throughout the course of a simulation, either via an extended dynamics or by introducing stochasticity into the equations of motion. In doing so, thermostats and barostats enable MD simulations to sample from the isothermal or

isothermal-isobaric ensembles.

The Berendsen thermostat is often used at the beginning of MD simulations to efficiently equilibrate the temperature of the system.<sup>116</sup> At each timestep, the instantaneous temperature,  $T'(t)$ , is calculated:

$$T'(t) = \frac{\sum_{i=1}^N m_i v_i^2}{N_f k_B} \quad (2.17)$$

where  $k_B$  is Boltzmann's constant,  $v_i$  and  $m_i$  are the velocity and mass of particle  $i$ ,  $N$  is the number of particles, and  $N_f = 3N - 3$  is the number of degrees of freedom for  $N$  particles. The atomic velocities are then linearly rescaled such that  $T'(t)$  will converge to the target temperature of the simulation,  $T$ , within a specified time  $\tau$ . Specifically, the atomic velocities are scaled by a factor  $\lambda$ :

$$\lambda = \sqrt{1 + \frac{\Delta t (T - T'(t))}{\tau T'(t)}} \quad (2.18)$$

where the time constant,  $\tau$ , sets the coupling strength of the thermostat to the heat bath. The Berendsen thermostat is fast and efficient to compute, but it does not reproduce the canonical ensemble.<sup>131</sup> This is because it does not generate the expected energy fluctuations for the canonical ensemble. Therefore, after the temperature has been equilibrated using the Berendsen thermostat, a more sophisticated thermostat is often used to ensure the desired ensemble is sampled from. One option is to add stochasticity to the Berendsen thermostat.<sup>132</sup> Another widely used method is the Nosé-Hoover thermostat,<sup>133,134</sup> which is an extended dynamics with an extra degree of freedom in the Hamiltonian for the heat bath.

The Berendsen barostat is analogous to the Berendsen thermostat: pressure is maintained around an equilibrium value by rescaling the system dimensions at each timestep. This barostat is often used to efficiently equilibrate the system pressure, but it does not correctly reproduce the fluctuations expected for the isothermal-isobaric (NPT) ensemble. In the production stage of MD simulations, a Parrinello-Rahman barostat can be used to generate the NPT ensemble. This barostat is analogous to the Nosé-Hoover thermostat

in its approach to maintaining pressure.

### 2.1.7 Improving computational efficiency

The pairwise calculation of potential energy and interatomic forces is the most computationally expensive part of an MD simulation. Below I will discuss some of the approaches taken to reduce the computational cost of these calculations without sacrificing accuracy.

**Force calculation** As the number of atoms,  $N$ , increases, the number of non-bonded pairwise interactions increases as  $N(N - 1)$ . Even with high performance computers, this scaling is prohibitively expensive for large  $N$ . Therefore, all MD simulation engines make use of Newton's third law of motion:

$$\mathbf{F}_{ij} = -\mathbf{F}_{ji} \quad (2.19)$$

which states that the force acting on atom  $i$  due to atom  $j$  is equal and opposite to the force acting on atom  $j$  due to atom  $i$ . This means that only *unique* pairs of atoms need to be considered in the force calculation (Listing 2.1).

```
for each atom i in {0, ..., N-1}
  for each atom j in {i+1, ..., N}
    calculate Fij
    add Fij to the total force acting on i
    add -Fij to the total force acting on j
```

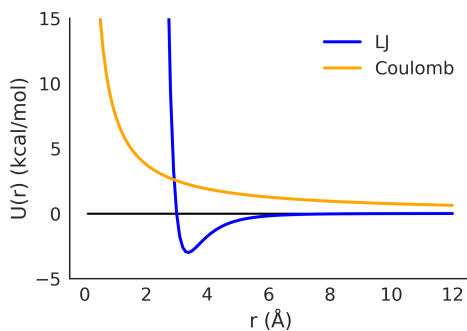
**Listing 2.1:** Pseudocode for the pairwise force calculation in an MD simulation.

This immediately cuts the number of force calculations in half without comprising the accuracy of the simulation. The same approach of making a single pass over unique pairs of atoms is taken for calculating the total potential energy of the system. For force fields that include three-body terms, such as COMPASS,<sup>135</sup> a similar approach can be used by iterating over unique triplets of atoms.<sup>116</sup>

**Potential truncation** Even  $\frac{N(N-1)}{2}$  calculations becomes expensive for large  $N$ . To further reduce the number of pairwise calculations, only pairs within a cutoff distance,  $r_c$ , of one another are considered. Beyond this cutoff distance, the Lennard-Jones potential is taken to be zero whilst the Coulombic interactions are calculated via a summation in reciprocal space. The cutoff distance is defined in the parameterisation of a force

field; CHARMM<sub>36</sub> and MARTINI 2 use cutoff distances of 12 Å and 11 Å, respectively.<sup>102,136</sup>

**Switching functions** The Lennard-Jones 12-6 potential rapidly decays toward zero as



**Figure 2.3:** Coulomb and Lennard-Jones 12-6 potentials used to model non-bonded interactions between water oxygen atoms in CHARMM<sub>36</sub>.<sup>102</sup>

$r \rightarrow \infty$ . At  $r = 2.5\sigma$  the potential energy is approximately equal to  $\epsilon/60$  (Figure 2.3). Therefore, at values of  $r > r_c$ ,  $U_{LJ}^{12,6}(r)$  can be taken to be exactly zero with little loss in accuracy. To avoid a discontinuity in the potential energy — and thus an infinite force — a switching function,  $S(r)$ , can be introduced to ensure the potential goes to zero smoothly at the cutoff distance. This is the approach taken by CHARMM<sub>36</sub>,<sup>98,102</sup> and leads to the following scheme for calculating the Lennard-Jones potential:

$$U_{LJ}^{12,6}(r) = \begin{cases} U_{LJ}^{12,6}(r), & \text{if } r < r_{\text{switch}} \\ U_{LJ}^{12,6}(r)S(r), & \text{if } r_{\text{switch}} \leq r < r_{\text{cut}} \\ 0, & \text{if } r_{\text{cut}} \leq r \end{cases} \quad (2.20)$$

where  $r_s$  is the interatomic distance at which the switching function is applied. In GRO-MACS,<sup>126</sup>  $S(r)$  is a fifth-degree polynomial:

$$S(r) = \frac{1 - 10(r - r_s)^3(r_c - r_s)^2 + 15(r - r_s)^4(r_c - r_s) - 6(r - r_s)}{(r_c - r_s)^5} \quad (2.21)$$

An alternative approach to ensuring the Lennard-Jones potential is equal to zero at the cutoff is to shift  $U_{LJ}^{12,6}(r)$  by its value at the cutoff distance.

**Particle-Mesh Ewald Summation** For the Coulombic interactions, the potential energy and force are also calculated explicitly for  $r < r_c$ . However, electrostatic interactions decay slowly and are non-negligible at  $r = 2.5\sigma$  (Figure 2.3). Therefore, past the cutoff distance, electrostatic interactions are calculated by an Ewald summation in reciprocal space. The smooth particle-mesh Ewald (PME) summation involves discretising the simulation box, calculating the mean partial charge in each grid point, interpolating the grid of partial charges to remove missing values, then transforming into reciprocal space using a fast Fourier transform.<sup>137,138</sup> The pairwise interactions between all grid points can be calculated using a single sum in reciprocal space, then the real-space interactions obtained by performing an inverse Fourier transform. The result is that electrostatic interactions can be calculated using the smooth-PME algorithm with  $O(N \log N)$  time complexity.

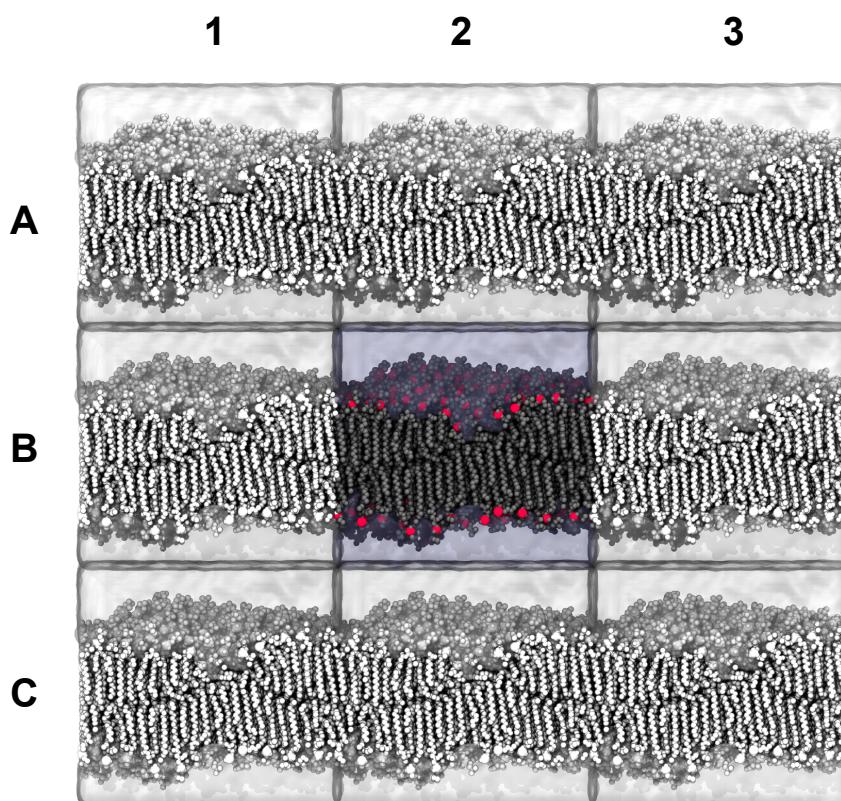
**Verlet neighbour lists** The use of a cutoff value,  $r_c$ , can significantly reduce the number of pairwise calculations performed at each timestep. However, it remains computationally expensive to determine whether each pair of atoms is within this interaction cutoff distance. Rather than calculating all interatomic distances at each timestep, a Verlet neighbour list can be used to keep track of all pairs of atoms that might be within  $r_c$  of one another. At the first timestep, all pairwise distances are calculated. All pairs within a distance of  $r_l$  from one another are added to a neighbour list, where  $r_l > r_c$ . This means that all pairs of atoms within  $r_c$  of one another are present in the list. Then, at each iteration, interatomic distances are calculated only for pairs of atoms on the neighbour list. To ensure all pairs within  $r_c$  of one another are considered, the neighbour list must be updated when the sum of the two largest atomic displacements is greater than  $r_l - r_c$ .<sup>116</sup> In practice, the neighbour list is updated every  $n$  steps, where  $n$  is typically around 20 for simulations of liquids. For gases and solids, respectively,  $n$  should be decreased or increased, whilst still ensuring all interacting pairs are on the neighbour list at all times. In doing so, the expensive distance calculation needs to be performed between all pairs only every  $n$  steps.

### 2.1.8 Simulating bulk behaviour

The central image of Figure 2.4 (row B, column 2) shows a lipid membrane system studied by Smith et al.<sup>19</sup> The system contains 400 DPPC lipids, is fully solvated by 150 mM



NaCl, and is approximately  $1 \times 10^3 \text{ nm}^3$  in size. This is not at the current limit of system sizes accessible to MD simulations, but is typical for all-atom simulations of lipid membranes. It is, however, far from the thermodynamic limit. This means that substantial finite size effects would be present in a system of this size. In order to simulate bulk behaviour without requiring  $1 \times 10^{23}$  molecules, periodic boundary conditions (PBC) can be applied to the system.



**Figure 2.4:** Illustration of the minimum image convention used in molecular dynamics simulations to simulate bulk behaviour.

In the system shown in Figure 2.4, PBCs effectively create an infinite number of stacked bilayers, with each bilayer being infinitely long. As an atom moves out of the central image, it is replaced on the opposite side of the box by a copy of itself. For instance, if in Figure 2.4 an atom moves from cell (B2) to cell (B1), a replica of this atom will simultaneously move from cell (B3) into the central image (B2).

The use of PBCs has important implications for the calculation of interatomic distances. The minimum image convention dictates that, in each spatial dimension, the *shortest* distance between any images of two atoms must be used. For a system of length  $L_d$  along

dimension  $d$ , the minimum image convention states that the distance,  $r_{ij}^d$ , between atom  $i$  and atom  $j$  along dimension  $d$  is given by:

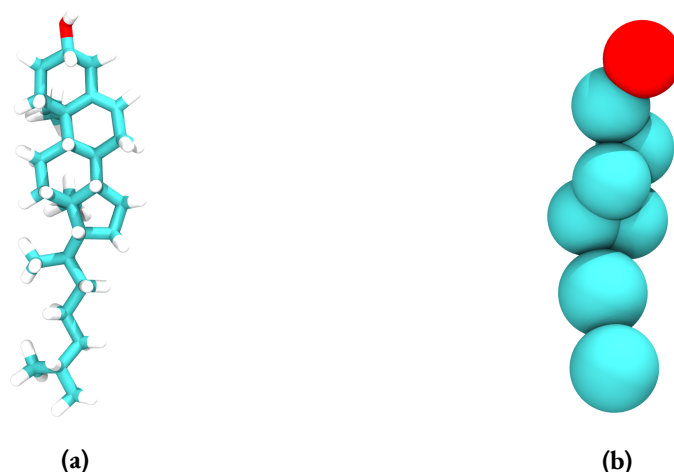
$$r_{ij}^d = \begin{cases} r_{ij}^d + L_d, & \text{if } r_{ij}^d < \frac{-L_d}{2} \\ r_{ij}^d - L_d, & \text{if } r_{ij}^d > \frac{L_d}{2} \\ r_{ij}^d, & \text{otherwise} \end{cases}$$

This ensures that the maximum distance in dimension  $d$  between any two atoms is equal to  $\frac{L_d}{2}$ . It also puts an upper limit on the value of  $r_c$  of  $\frac{\min(L_d)}{2}$ , where  $\min(L_d)$  is the minimum extent of all spatial dimensions. This is to prevent the non-physical behaviour that would arise if an atom were to interact with itself through periodic boundaries.

### 2.1.9 Coarse-grained molecular dynamics

Much of the discussion thus far applies to both all-atom and coarse-grained classical MD simulations. In all-atom MD, each atom is treated explicitly. Coarse-grained MD simulations, on the other hand, use interacting beads that approximate the behaviour of multiple atoms (Figure 2.5). A popular coarse-grained force field for biomolecular simulation MARTINI, which is used for the coarse-grained simulations in this thesis. The MARTINI model groups three to five heavy (non-hydrogen) atoms into each interacting bead. Different bead types are used to represent different combinations of grouped atoms along with different chemical environments. For MARTINI 2, there are four groups of bead type: charged, polar, non-polar, and apolar. Each group has five distinct types that are distinguished by their polar affinity or their tendency to form hydrogen bonds. The strength of interaction between different bead types ranges from 2.0 to 5.6 kJ mol<sup>-1</sup>. This provides a degree of chemical specificity, albeit a reduced one compared to all-atom simulations.

The MARTINI force field uses a similar functional form as in Equations 2.2 and 2.6, but with some important differences in the treatment of both bonded and non-bonded interactions. MARTINI has no Urey-Bradley term nor a term for proper dihedrals, although improper dihedrals can be used to maintain planarity of rigid structures.



**Figure 2.5:** (a) Atomistic and (b) Coarse-grained representations of cholesterol. The mapping shown here is used in the MARTINI coarse-grained force field.<sup>98,139</sup>

MARTINI uses the same  $U_{LJ}^{12,6}(\mathbf{r})$  potential to describe non-bonded dispersion forces. In the original parameterisation of MARTINI 2, a switching function was applied to the Lennard-Jones potential to ensure it went to zero at the interaction cutoff distance of 12 Å.<sup>98</sup> Since 2016, however, the recommendation has been to shift the potential by its value at the cutoff distance, and the cutoff distance has been reduced to 11 Å. This is both to improve computational efficiency and to allow MARTINI simulations to be performed with GROMACS on GPUs.<sup>136</sup>

Electrostatics are also treated differently by MARTINI. First, there are no partial charges — instead, all beads have integer values of  $q_e$ , with most beads (and lipids) having a charge of zero. Second, in the original parameterisation of MARTINI 2, a switching function was used to set electrostatic interactions to zero at a cutoff distance of 12 Å.<sup>98</sup> Whilst this neglects long-range electrostatics, MARTINI systems are typically sparsely charged, with the Coulomb energy being around 1% of the total Lennard-Jones energy.<sup>136</sup> Since 2016, the recommendation has been to use a reaction field method for evaluating electrostatics.<sup>140</sup> By setting the dielectric constant to be infinity at separation distances greater than  $r_c$ , the electrostatic interactions are still set to zero at the cutoff distance. However, the reaction field approach is more computationally efficient than using a switching function.

Coarse-graining allows the study of larger systems over longer time scales. The speed up

of calculations comes from two sources. Firstly, as there are fewer particles in the system, the calculation of pairwise interactions is significantly faster. Secondly, by coarse-graining 3-5 heavy atoms into an interacting bead, the free energy landscape is smoothed. This means that a larger timestep can be used (up to 30 fs) as the highest frequency vibrations are no longer present. It also acts to increase the dynamics by reducing the height of energetic barriers, increasing the rate of exploration of phase space. Taken together, this opens up the ability to simulate systems of hundreds of nm in size for time scales approaching 1.0 ms.<sup>98</sup> This, however, comes at the cost of losing the atomic scale details that all-atom simulations offer. Further, the dynamics of molecular processes are altered in unpredictable ways, and the underlying physics that drive emergent phenomena in these models may not be true to reality.<sup>141</sup>

### 2.1.10 Summary

In summary, classical MD simulations can provide unique insight into condensed matter systems at atomic resolution. Empirical force fields — developed by fitting to quantum-mechanical calculations and experimental data — provide an approximate but accurate representation of physics at the atomic scale. Such force fields can be used to simulate the time-evolution of a system of interacting particles over millions of short timesteps. Numerical integrators are used to update atomic coordinates and velocities at each timestep, and thermostats and barostats can be used to ensure the correct ensemble is sampled from. Various approaches are taken to improve computational efficiency, including decomposing the force calculations into short-range and long-range contributions that are evaluated in real and reciprocal space, respectively. Bulk behaviour can be obtained through the use of periodic boundary conditions, whilst larger systems can be studied over longer timescales using coarse-grained models.

## 2.2 Machine learning

Machine learning is a type of artificial intelligence that allows computers to improve their performance at a task by identifying patterns in data. Machine learning can be broadly categorised into *supervised* and *unsupervised* learning.

Supervised learning refers to algorithms that use labelled examples of input and output data to construct a model that can be used for future predictions. For example, a predictive model of tumour malignancy could be created using data from past patients. The input features for each patient might be the tumour radius, texture, and concavity, whilst the output feature would be whether the tumour is benign or malignant. Once the model has been constructed, it can be used to predict whether a new patient's tumour is benign or malignant based on the tumour's radius, texture, and concavity.

Unsupervised learning is employed when there is no labelled output data. It can be used to partition a set of samples into groups, or clusters, whereby intragroup input features are similar and intergroup features are dissimilar. For instance, Netflix clusters its viewers into distinct groups based on their viewing history. It then provides recommendations to a customer based on other titles watched by members in the same group.

Unsupervised machine learning is now routinely employed in the analysis of MD simulations.<sup>142,143</sup> In this thesis, dimensionality reduction and clustering are used to study cholesterol-sphingomyelin conformations in lipid bilayers, and Hidden Markov models are used to look at the effect of cholesterol oxidation on liquid-liquid phase separation in model membranes.

### 2.2.1 Dimensionality reduction and clustering

The unsupervised clustering of samples is a powerful method for uncovering correlations in complex data sets. Many problems, however, have high-dimensional input spaces, which presents two related problems to clustering. First, as the dimensionality increases, points in the high-dimensional space become extremely sparse. This means that, for the clustering algorithm, most points appear dissimilar to one another. Second, samples may be close in some dimensions but far apart in others, making metrics such as Euclidian distance less useful for identifying nearby data points. Dimensionality reduction techniques aim to overcome these issues by creating low-dimensional representations of high-dimensional spaces, without losing important information contained within the original dataset.

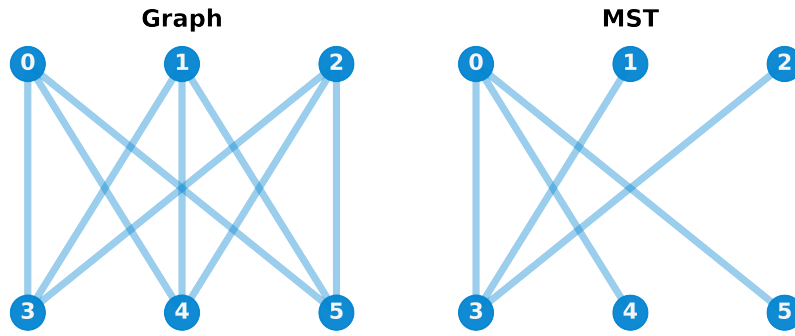
**Linear dimensionality reduction** Linear dimensionality reduction algorithms generate a set of new features based on linear combinations of the many input features. Prin-

Principal Component Analysis (PCA) is probably the most famous technique of this kind. PCA defines a set of orthogonal components through the eigendecomposition of the covariance matrix of the input data. There are  $N$  components in total, where  $N$  is the dimensionality of the input space. The component with the largest eigenvalue will be the one that maximises the variance when the input data is projected onto it. Therefore, by projecting onto the  $n$  components with the largest eigenvalues, the input data can be transformed into an  $n$ -dimensional representation in which the variance amongst samples is maximised.

**Non-linear dimensionality reduction** If the input data does not lie on a hyperplane in the high-dimensional space, linear techniques will be unable to faithfully reproduce the salient features of the data in a low-dimensional representation.<sup>143</sup> In these instances, non-linear, or *manifold*, dimensionality reduction techniques can be used. Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) is a widely used non-linear method for creating low-dimensional embeddings of complex high-dimensional data.<sup>144</sup> UMAP creates a weighted graph of the points in the high-dimensional space, then iteratively optimises a low-dimensional graph to be as similar to it as possible. In doing so, UMAP is able to capture both local and global information in the low-dimensional embedding. UMAP has proven to be particularly useful for analysing MD simulations of biomacromolecules, preserving the biological information present in high-dimensional data better than linear methods such as PCA.<sup>142</sup> UMAP is used in this thesis in the analysis of cholesterol-sphingomelin conformations in lipids bilayers.

**HDBSCAN** Data in a low-dimensional space, of less than around 15 dimensions, can be segregated into distinct groups using unsupervised clustering techniques. HDBSCAN is a density-based hierarchical clustering algorithm that is well-suited to identify clusters of different shapes, sizes, and densities.<sup>145,146</sup>

HDBSCAN first transforms the input space to sparsify regions that are already of low density, whilst leaving regions of high density unchanged. This serves to reduce the tendency of noise to connect regions of high density. To do so, HDBSCAN calculates the



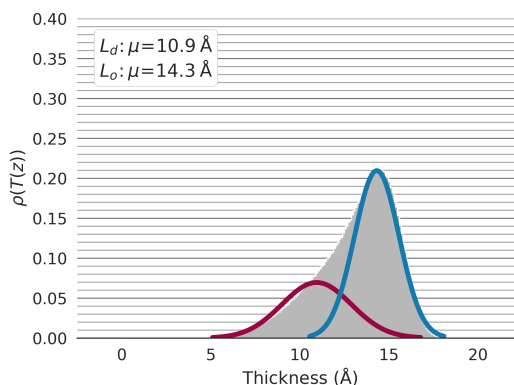
**Figure 2.6:** The minimum spanning tree (MST) of an unweighted graph is defined as the set of edges that contains every node of the graph whilst minimising the number of edges in the tree.

*mutual reachability distance*,  $d_{mreach,k}$ , for each pair of input points:

$$d_{mreach,k}(a,b) = \max\{\text{core}_k(a), \text{core}_k(b), d(a,b)\} \quad (2.22)$$

where  $\text{core}_k(x)$  is the distance from point  $x$  to its  $k$ th nearest neighbour, and  $d(a,b)$  is the distance between points  $a$  and  $b$ . Using the mutual reachability distance, dense points with small  $\text{core}_k(x)$  values remain unchanged. Sparse point with larger  $\text{core}_k(x)$  values, however, are moved such that they are at least  $\text{core}_k(x)$  away from **all** other points. That is, point  $a$  will be moved to increase the distance  $d(a,b)$  if  $\text{core}_k(a)$  is above some threshold and  $d(a,b)$  is less than  $\text{core}_k(a)$ .

The minimum spanning tree (MST) of a weighted graph is then generated from this transformed data, where the weights are the distances between points. The MST of a graph is the set of vertices that generates a fully-connected, acyclic graph whilst minimising the total weight of its vertices (Figure 2.6). A hierarchical clustering of the MST's connected components is then performed. At this stage, the related DBSCAN clustering algorithm uses a single cutoff distance for identifying distinct clusters.<sup>147</sup> HDBSCAN, however, uses several cutoff distances in order to identify the set of most stable clusters, each with a minimum number of points. Using multiple cutoff distances is what makes HDBSCAN so useful for identifying clusters of variable densities. HDBSCAN is used in this thesis to cluster the cholesterol-sphingomyelin conformations in the low dimensional embedding generated with UMAP.



**Figure 2.7:** Acyl tail thickness of DOPC lipids in bilayer with coexisting  $L_d$  and  $L_o$  phases. Assuming the distributions of  $L_d$  and  $L_o$  thicknesses can each be described by a normal distribution, a Gaussian Mixture Model (GMM) can be used to determine the parameters of these underlying. A Gaussian Mixture Model can also perform a soft-clustering of each data point, providing a probability of each lipid being either in the  $L_d$  or the  $L_o$  phase.

**Gaussian mixture models** Mixture models are another form of unsupervised clustering. They can be used to decompose a probability distribution into a set of subpopulations, each of which is described parametrically. For Gaussian Mixture Models (GMM), the subpopulations are assumed to follow Gaussian distributions. Mixture models are useful when the number of subpopulations is known *a priori*. For example, consider a bilayer that has both  $L_o$  and  $L_d$  phase DOPC lipids. The thicknesses of the acyl tails follows a Gaussian distribution with negative skewness (Figure 2.7), whereby  $L_o$  tails are generally thicker than  $L_d$  tails. A GMM can be used to determine the parameters of the two underlying Gaussian distributions — corresponding to  $L_o$  or  $L_d$  lipids — based on the observed total distribution. GMMs can also perform a *soft clustering* of the input data, providing a probability of each sample (lipid molecule) belonging to each subpopulation ( $L_d$  or  $L_o$ ).

To determine the parameters of the underlying distributions mixture models use the Expectation-Maximisation (EM) algorithm, which is a general algorithm that iteratively optimises parameter fitting. For the example of the DOPC lipid bilayer, EM consists the following steps:

1. Initialisation: Generate an initial guess of the parameters (mean and standard deviation) of the two underlying Gaussians. Randomly assign each data point (acyl tail thickness) to one of the two Gaussians.



2. **Expectation (E):** Bayes theorem is used to calculate the probability of each data point belonging to each Gaussian.
3. **Maximisation (M):** The parameters of the two Gaussians are recalculated. For a given Gaussian, the contribution of each data point is weighted by its probability of belonging to the Gaussian.
4. Repeat the EM steps until the parameters of two consecutive iterations converge to within some pre-defined tolerance.

GMMs are particularly useful for determining the parameters of the underlying distributions. However, the soft-clustering of points is less useful when there is a large overlap between the two distributions, such as in Figure 2.7. GMMs are used in this thesis to decompose the distribution of lipid tail thicknesses into two *hidden* states, corresponding to the  $L_o$  and  $L_d$  phases. The parameters of the underlying distributions are then used as the input to Hidden Markov models.

### 2.2.2 Hidden Markov Models

A Hidden Markov model (HMM) relates a timeseries of observations — such as the thickness of DOPC acyl tails — to a timeseries of unobserved hidden states — such as whether each DOPC lipid is in the  $L_d$  or the  $L_o$  phase. HMMs can thus be thought of as a mixture model with a time component. The timeseries of hidden states is assumed to be memoryless; the state at time  $t + 1$  depends only on the state at time  $t$ .

To fit the parameters of a HMM model, a special case of the EM algorithm, known as the Baum-Welch algorithm, is used. The Baum-Welch algorithm is used to find the optimal values of the following parameters:

- the probability of starting in each hidden state.
- the *transition probability* matrix of moving between each hidden state from time  $t$  to time  $t + 1$ .
- the *emission probability* matrix of each hidden state leading to each observed state.

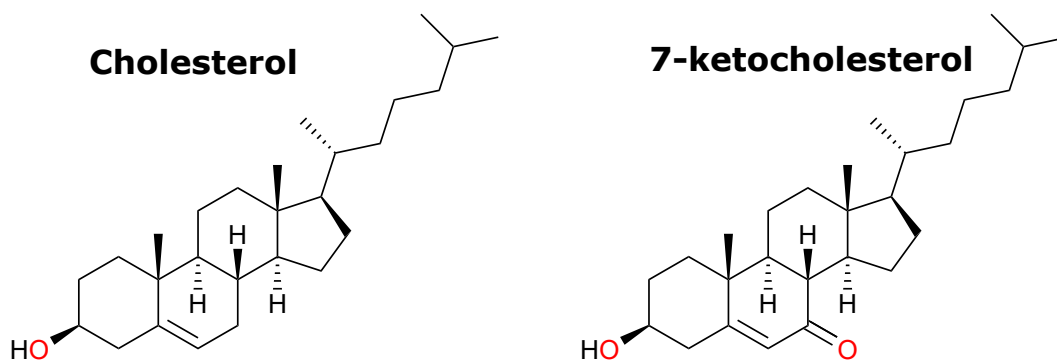
Once these model parameters have been fit, the Viterbi algorithm can be used to determine the most likely sequence of hidden states based on the sequence of observed states.

In this thesis, HMMs are used to study the coexistence of  $L_d$  and  $L_o$  phases in lipid membranes.

## Chapter 3

# Cholesterol Oxidation Modulates the Formation of Liquid-Ordered Domains in Model Membranes

**A note on the reliability of the results presented in this chapter** The results in this chapter were published as a preprint on bioRxiv.<sup>148</sup> However, soon after publication of the preprint, M. Javanainen brought to my attention an issue with the MARTINI model of cholesterol when used with the default GROMACS parameters for the LINCS constraint algorithm. Due to the presence of virtual sites in the cholesterol model, use of the default LINCS parameters leads a lack of conservation of energy in the system. This manifests as an unphysical lateral temperature gradients across the bilayers.<sup>149</sup> It is also what causes the lateral phase separation described in this chapter - if more conservative LINCS parameters are used, the temperature gradient disappears but so too does the phase separation. As such, the results in the section are unlikely to be reliable. Nonetheless, the analysis methods developed in this chapter could be applied to other simulations of phase separating membranes that do not suffer aphysical temperature gradients. In Section 3.5 of this chapter, I will discuss the physical origin of the temperature gradient in more detail as well as approaches to avoiding this issue.



**Figure 3.1:** Chemical structure of cholesterol and 7-ketocholesterol.

### 3.1 Introduction

Since Simons and Ikonen first described lipid rafts,<sup>43</sup> the existence, origin and nature of these structures in cellular membranes has been hotly debated.<sup>27–42</sup> However, there is now direct evidence of microdomains in live yeast cell organelles;<sup>150,151</sup> of nanodomains in live plant cell plasma membranes;<sup>152</sup> of functional membrane microdomains in live bacteria;<sup>153,154</sup> and of nanodomains in isolated mammalian cell plasma membranes.<sup>155</sup> The ubiquitous presence of lipid-raft-like structures across the domains of life suggests they serve some biological function. This is further supported by their suspected roles in many membrane processes: from membrane signaling<sup>156</sup> to membrane trafficking,<sup>157</sup> from membrane deformation<sup>158</sup> to membrane vesiculation,<sup>159</sup> and from sites for oligomerization of peptides<sup>160</sup> to sites for attachment of pathogens.<sup>161</sup>

Given the biological importance of lipid rafts, the disruption of liquid-ordered domains has the potential to impact myriad biological pathways and processes. Elevated levels of ring-oxidised sterols — produced by the autoxidation of cholesterol<sup>162</sup> — are implicated in numerous pathologies,<sup>61,163–174</sup> and have been speculated to prevent liquid-ordered domain formation.<sup>60,175,176</sup> 7-ketocholesterol (KChol; Figure 3.1) is one of the most abundant and cytotoxic oxysterols,<sup>60</sup> and its presence in lipid rafts can induce cell death.<sup>165</sup> KChol causes apoptosis via inactivation of the phosphatidylinositol 3-kinase/Akt signaling pathway<sup>177</sup> — a pathway that depends on lipid rafts as signaling platforms.<sup>156</sup> Further, by excluding KChol from lipid rafts, cell death is avoided.<sup>178</sup> It is therefore possible that KChol induces apoptosis via disrupting the formation of liquid-ordered domains in the plasma membrane.

The concept of lipid rafts originated as an explanation for the dynamic clustering of cholesterol (Chol; Figure 3.1) and sphingolipids in the plasma membrane, and the preferential sorting of certain proteins into these domains.<sup>43</sup> Since then, many different lipid mixtures have been found to be capable of nano- or micro-domain formation.<sup>42,47,179–185</sup> Indeed, the plasma membrane is thought to consist of many different raft-like and non-raft-like regions of varying lipid composition.<sup>31,35,182</sup> These raft-like regions may arise through many different physical processes,<sup>34,186,187</sup> with different physical mechanisms dominating at different stages of domain formation.<sup>188</sup> Given the complexity of the plasma membrane, model membranes are typically employed for the study of domain formation. Membranes consisting of 1,2-dipalmitoyl-sn-3-phosphocholine (DPPC), 1,2-dioleoyl-sn-glycero-3-phosphocholine (DOPC) and Chol were the first phase-separating ternary mixture to have its phase boundaries fully mapped,<sup>47</sup> and has since become the canonical mixture for studying phase separation in lipid membranes. While this mixture produces macroscopic phase separation, nanodomains behave surprisingly like genuine phases and so studying macroscopic phase separation may also inform us about nanodomains and lipid rafts.<sup>48</sup>

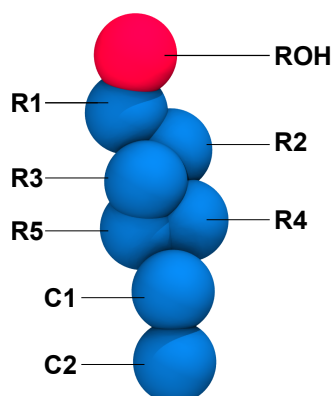
In this chapter, I report on the effect of cholesterol oxidation on domain formation studied by means of coarse-grained molecular dynamics simulations.

## 3.2 **Methods**

### 3.2.1 **Simulation protocol**

I used the CHARMM-GUI MARTINI Maker<sup>124,125</sup> to construct an equimolar mixture of DPPC:DOPC:Chol with 6,000 lipids per leaflet. The system had 89,995 non-polarizable water beads (10% of which were anti-freeze beads), 1,154 Na beads, and 1,154 Cl beads. I used the MARTINI 2 force field along with the Melo et al. parameters for cholesterol.<sup>98,139</sup> To construct the DPPC:DOPC:KChol membrane, I used the DPPC:DOPC:Chol bilayer and replaced all Chol molecules with KChol molecules.

I first performed a steepest descent minimization for 5,000 steps in which the sterol constraints were replaced by harmonic bonds. I then performed a series of short ( $\sim 1$  ns) equilibrations, with increasing timesteps (2, 5, 10, 15, 20 fs), to relax the systems. In



**Figure 3.2:** Structure of the MARTINI model of Chol/KChol. The R2 bead is of type SC<sub>3</sub> and SNo for Chol and KChol respectively. The red ROH bead is polar, representing the hydroxyl group of the sterols, and the blue beads are apolar.

these equilibrations, I applied position restraints with decreasing coefficients (200, 100, 50, 20, 10 kcal mol<sup>-1</sup>) in the z-dimension to the PO<sub>4</sub> and ROH beads. These beads correspond to the phosphate group of the glycerophospholipids and the hydroxyl group of the sterols.

For production simulations, I used a timestep of 25 fs and to suppress large-scale undulations<sup>189</sup> I applied a 2 kcal mol<sup>-1</sup> restraint in the z-dimension to the PO<sub>4</sub> beads of phospholipids in the upper leaflet.

All simulations were performed using the semi-isotropic NPT ensemble at 310 K and 1 bar, and using the new-RF parameter set<sup>136</sup> for performing MARTINI simulations with GROMACS. To perform the second replicas of each mixture I used a different random seed when generating initial velocities. Coordinates were stored every 0.5 ns. All simulations were performed using GROMACS 2018.2.<sup>126</sup>

### 3.2.2 MARTINI parameters for 7-ketocholesterol

In keeping with the modular philosophy of MARTINI,<sup>141</sup> I modelled 7-ketocholesterol by changing the R2 bead type from SC<sub>3</sub> (semi-repulsive to water) to SNo (intermediate with water), which has the same mass as SC<sub>3</sub> but is more polar (Figure 3.2).

SNo is a conservative choice of bead for the ketone group. The SP<sub>1</sub> bead (polar; almost attractive with water) is used for the ROH group in the MARTINI model of cholesterol, as well as in all three ROH groups of the cholate MARTINI model.<sup>139</sup> These hydroxyl groups, however, are more polar than the ketone group of KChol — hydroxyl

moeities can both donate and accept hydrogen bonds. I therefore used the SNo bead (intermediate polar; intermediate with water). The mass of the beads and all bond lengths are unchanged, which is in keeping with the cholate MARTINI model (the ROH beads in cholate keep the same mass as the corresponding beads in the cholesterol model).

I also performed simulations using an No bead type rather than SNo bead type for the R<sub>2</sub> ketone bead. The No bead has stronger Lennard-Jones interactions with other beads compared to the SNo bead. However, when using the No bead I observed freezing of water at the membrane-water interface after around 4  $\mu$ s of simulation time, most likely due to the increased interactions between the ketone group of KChol and the water beads.

The angular orientation of our KChol model is in line with previous atomistic simulations of this ring-oxidized sterol.<sup>60,175,176</sup> This is an important differentiator between Chol and KChol, thus giving confidence in the KChol model.

### 3.2.3 Analysis methods

Analysis was performed using MDAnalysis,<sup>190,191</sup> LiPyphilic,<sup>192</sup> FATSLiM,<sup>193</sup> SciPy,<sup>194</sup> and HMMLearn.<sup>195</sup> Unless stated otherwise, every tenth frame (5 ns) was used in the analysis. The standard errors reported in Table 3.1 were calculated using 50 ns block averages.

**Hidden Markov Model** Lipids were assigned to be either ordered ( $L_o$ ), disordered ( $L_d$ ), or intermediate ( $L_{d/o}$ ) by constructing Hidden Markov Models based on lipid thicknesses. SciPy<sup>194</sup> was used to generate the Gaussian Mixture Model, from which the initial HMM parameters were derived. HMMLearn was then used to refine the model parameters and subsequently decode the most likely sequence of ordered states. Smith et al.<sup>196</sup> describes this procedure in more detail.

**Area per lipid** The area per lipid was calculated via a Voronoi tessellation of the  $x$  and  $y$  coordinates of GL<sub>1</sub>, GL<sub>2</sub>, and ROH beads within each leaflet. These beads correspond to the glycerol moiety of the phospholipids and the hydroxyl group of the sterols. The analysis was performed using LiPyphilic,<sup>192</sup> which uses Freud<sup>197</sup> to perform the tessellation of atomic coordinates.

**Coarse-grained order parameter** The coarse-grained order parameter,  $S_{CC}$ , is given

by:

$$S_{CC} = \frac{\langle 3 \cos^2 \theta \rangle}{2}$$

where  $\theta$  is the angle between the membrane normal (approximated as the z-axis) and the vector connecting two consecutive tail beads. The average is taken over all beads in a molecule. LiPyphilic<sup>192</sup> was used to perform the calculation.

**Membrane thickness** For each phospholipid, a local leaflet patch was defined by all PO<sub>4</sub> (phosphate) beads within 60 Å of the reference lipid's PO<sub>4</sub> bead. The normal to this patch was used to identify a reference lipid for the apposing leaflet, and a local patch defined for this second lipid in a similar manner. The membrane thickness for the original lipid was taken to be the distance between the center of mass of the two leaflet patches. FATSLiM<sup>193</sup> was used to perform the calculation.

**Fractional enrichment** To calculate the fractional enrichment of lipid species, a neighbor matrix,  $A$ , was first constructed. The matrix is 12,000 by 12,000, where each row or column represents a distinct lipid molecule.  $A_{ij} = 1$  if two lipids are neighbors and  $A_{ij} = 0$  otherwise. Two lipids were considered neighbors if they have any of the GL1, GL2, or ROH beads within 15 Å of one another. The neighbor matrix was then used to determine the fractional enrichment of each species over the final 4 μs of simulation time. The fractional enrichment of species  $B$  around species  $A$ ,  $E_{AB}$ , is given by:

$$E_{AB} = \frac{[B]_{\text{Local}}}{[B]_{\text{Bulk}}}$$

where  $[B]_{\text{Bulk}}$  and  $[B]_{\text{Local}}$  are the bulk concentrations and local concentration around species  $A$ , respectively, of species  $B$ .<sup>121</sup>

The same neighbor matrix was used to calculate the fractional enrichment based on lipid order ( $L_d$ ,  $L_{d/o}$ , or  $L_o$ ) of lipids. LiPyphilic<sup>192</sup> was used to construct the neighbor matrix.

**Largest domain** To calculate the largest cluster of  $L_o$  lipids at a given frame, the neighbor matrix described above was used. First, the rows and columns of non- $L_o$  lipids were removed. Then the largest connected component of this new matrix was found, which corresponds to the largest cluster of  $L_o$  lipids. The same approach was used to identify



lipids in the largest  $L_d$  domain at each frame. LiPyphilic<sup>192</sup> was used to find the largest clusters.

**Registration** The interleaflet registration,  $r_{u/l}$ , can be defined as the Pearson correlation coefficient between lateral densities of  $L_o$  lipids in the upper and lower leaflets.<sup>183</sup> Values of  $r_{u/l} = 1$  correspond to perfectly registered domains and values of  $r_{u/l} = -1$  correspond to perfectly anti-registered domains. LiPyphilic<sup>192</sup> was used to perform the calculation.

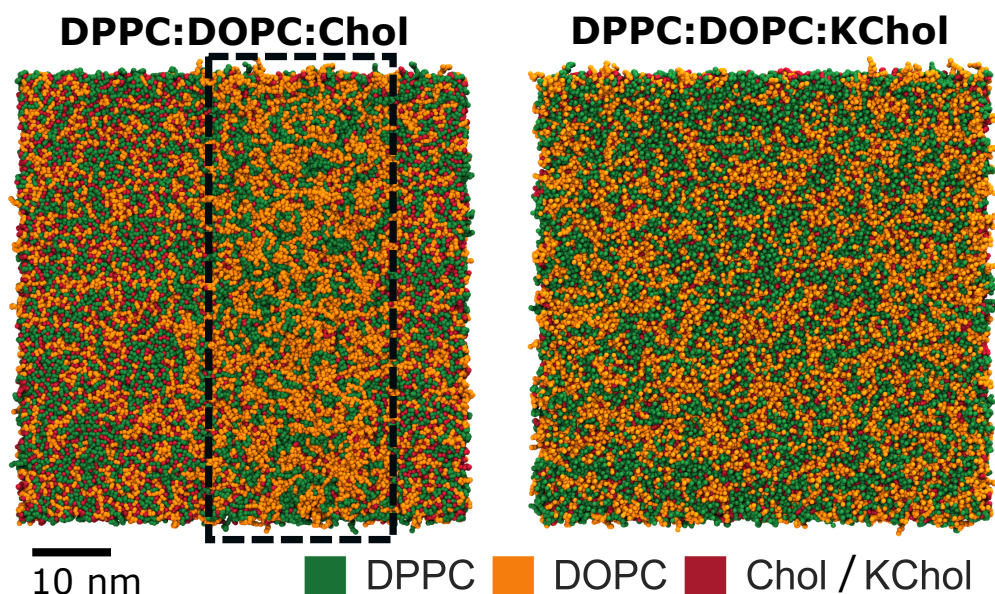
**Flip-flop** To calculate the flip-flop rate of cholesterol, lipids were first assigned to leaflets based on the  $z$  coordinate of their GL1, GL2, and ROH beads using LiPyphilic.<sup>192</sup> A cholesterol molecule with its ROH bead within 10 Å of its local membrane midpoint was classified as being in the midplane. A cholesterol molecule was taken to have flip-flopped if it left one leaflet, passed through the midplane, and then resided in the apposing leaflet for at least 10 ns. Every tenth frame (0.5 ns) from the final 4 μs of each replica was used in the analysis. The rate was calculated by dividing the total number of observed flip-flops by the product of the number of sterol molecules and the total simulation time used for the analysis. LiPyphilic<sup>192</sup> was used to perform the calculation.

**PMF** Sterol height was calculated as the signed distance in  $z$  from the ROH bead to the membrane midpoint. Sterol orientation was defined as the angle between the  $z$ -axis and the vector from bead R5 to R1. The PMF of sterol orientation and height,  $F(z, \theta_z)$ , was then calculated directly from the joint probability distribution,  $P(z, \theta_z)$ . The PMF is given by:

$$F(z, \theta_z) = -k_B T \ln P(z, \theta_z)$$

where  $k_B$  is the Boltzmann constant and  $T$  is the temperature in Kelvin. LiPyphilic<sup>192</sup> was used to plot the PMFs.

**Lateral diffusion** The lateral diffusion coefficient was calculated from the mean-square displacement (MSD) of PO4 and ROH beads via the Einstein relation. The MSD was calculated for lipids in the largest  $L_o$  or  $L_d$  clusters separately. The center of mass motion of the  $L_o$  or  $L_d$  cluster was removed from the MSD of the respective lipids. The MSD and diffusion coefficients were calculated using LiPyphilic,<sup>192</sup> which uses tidynamics<sup>198</sup> to calculate the MSD via the Fast Correlation Algorithm.

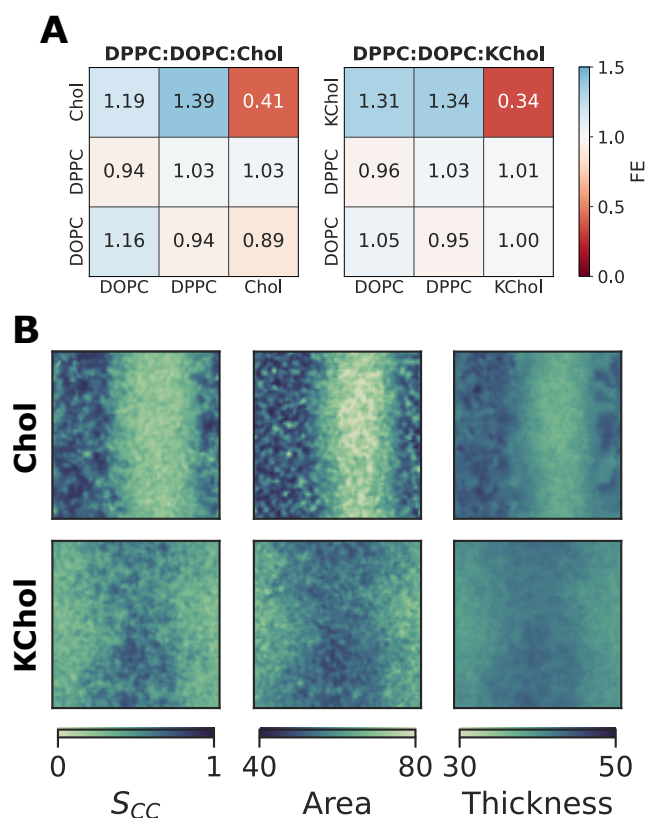


**Figure 3.3:** Lateral lipid distribution at 20  $\mu$ s. The column highlighted in the DPPC:DOPC:Chol mixture is depleted of cholesterol.

### 3.3 Results

There is a lateral demixing of lipids in the DPPC:DOPC:Chol membrane, with clearly defined Chol-poor and Chol-enriched regions (Figure 3.3). On the other hand, lipids in the DPPC:DOPC:KChol membrane appear more uniformly distributed (Figure 3.3). This immediately illustrates the profound impact that a single chemical substitution within one of the lipid constituents has on the lipid mixing within the membrane. To quantify the demixing of lipids in the membranes, I calculated the lipid enrichment/depletion index<sup>121,199</sup> of each species over the final 4  $\mu$ s of simulation time (Figure 3.4A). Chol has a clear preference for DPPC over DOPC, whilst DOPC tends to self-aggregate. This affinity between Chol and DPPC is what drives the macroscopic phase separation in the DPPC:DOPC:Chol membrane.<sup>188,200</sup> Several recent studies have shown that small changes in a phospholipid's chemistry can alter its affinity for Chol, and thus change the size and stability of lateral heterogeneities.<sup>7,181,185,201–204</sup> Here, I find that KChol, an oxidation product of Chol, has significantly less affinity for DPPC over DOPC than Chol — and the result is a disruption of the macroscopic phase separation.

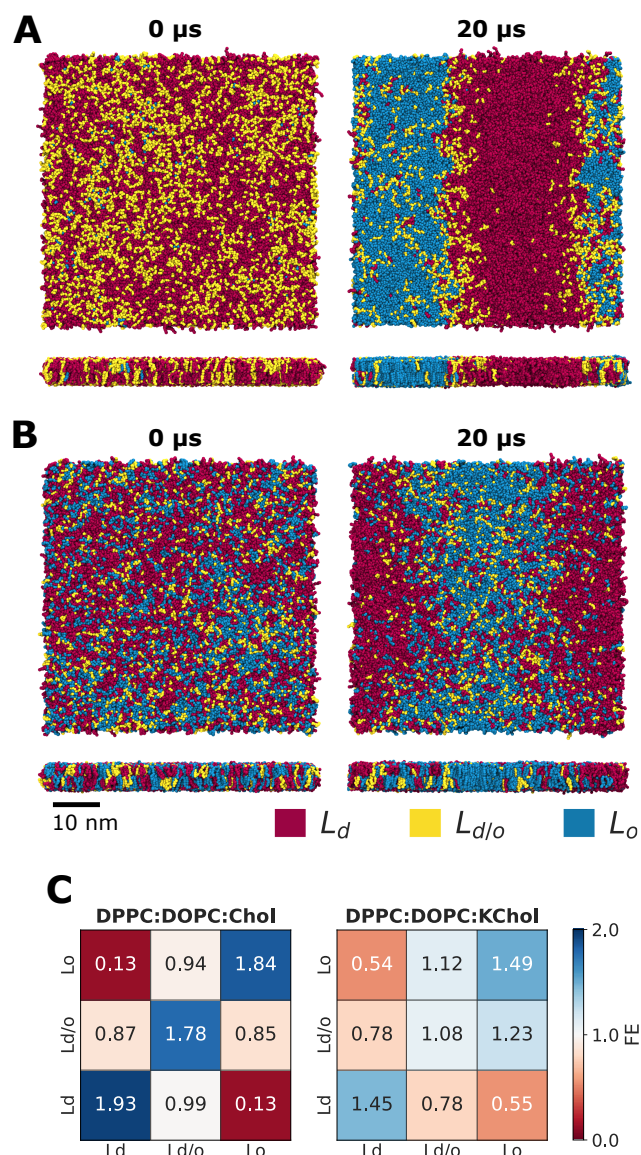
As a result of the phase separation, there is a large order gradient across the DPPC:DOPC:Chol membrane — the Chol-depleted region in Figure 3.3 is signifi-



**Figure 3.4:** (A) Fractional enrichment of lipid species, calculated using the final 4  $\mu\text{s}$  of each simulation. Values above and below 1 indicate enrichment and depletion, respectively. (B) Projection onto the membrane plane of the coarse-grained order parameter ( $S_{CC}$ ), area per lipid ( $\text{\AA}^2$ ), and local membrane thickness ( $\text{\AA}$ ) of the phospholipids.

cantly more disordered than the Chol-enriched region. It has a larger area per lipid, smaller membrane thickness, and more disordered acyl tails than the Chol-enriched region (Figure 3.4B, upper panel). Whilst there is little lateral demixing of lipid species in the DPPC:DOPC:KChol membrane, there is still an order gradient across the membrane (Figure 3.4B, lower panel). There is a large ordered region in the center of this membrane, with disordered regions either side. There is, however, a reduced gradient compared to the one in DPPC:DOPC:Chol membrane, and the boundary between the ordered and disordered domains is more diffuse. The lateral heterogeneity in the DPPC:DOPC:KChol membrane is thus more akin to the nanodomains that form in DPPC:cholesterol binary mixtures<sup>205</sup> than to the phase separated DPPC:DOPC:Chol membrane.

To better understand the affect of cholesterol oxidation on the domain-formation



**Figure 3.5:** Lateral distribution of ordered ( $L_o$ ), disordered ( $L_d$ ), and intermediate ( $L_{d/o}$ ) lipids in the (A) DPPC:DOPC:Chol membrane and (B) DPPC:DOPC:KChol membrane. (C) Fractional enrichment of lipids by their phase ( $L_d$ ,  $L_{d/o}$ ,  $L_o$ ), calculated using the final 4  $\mu$ s of each simulation. Values above and below 1 indicate enrichment and depletion, respectively.

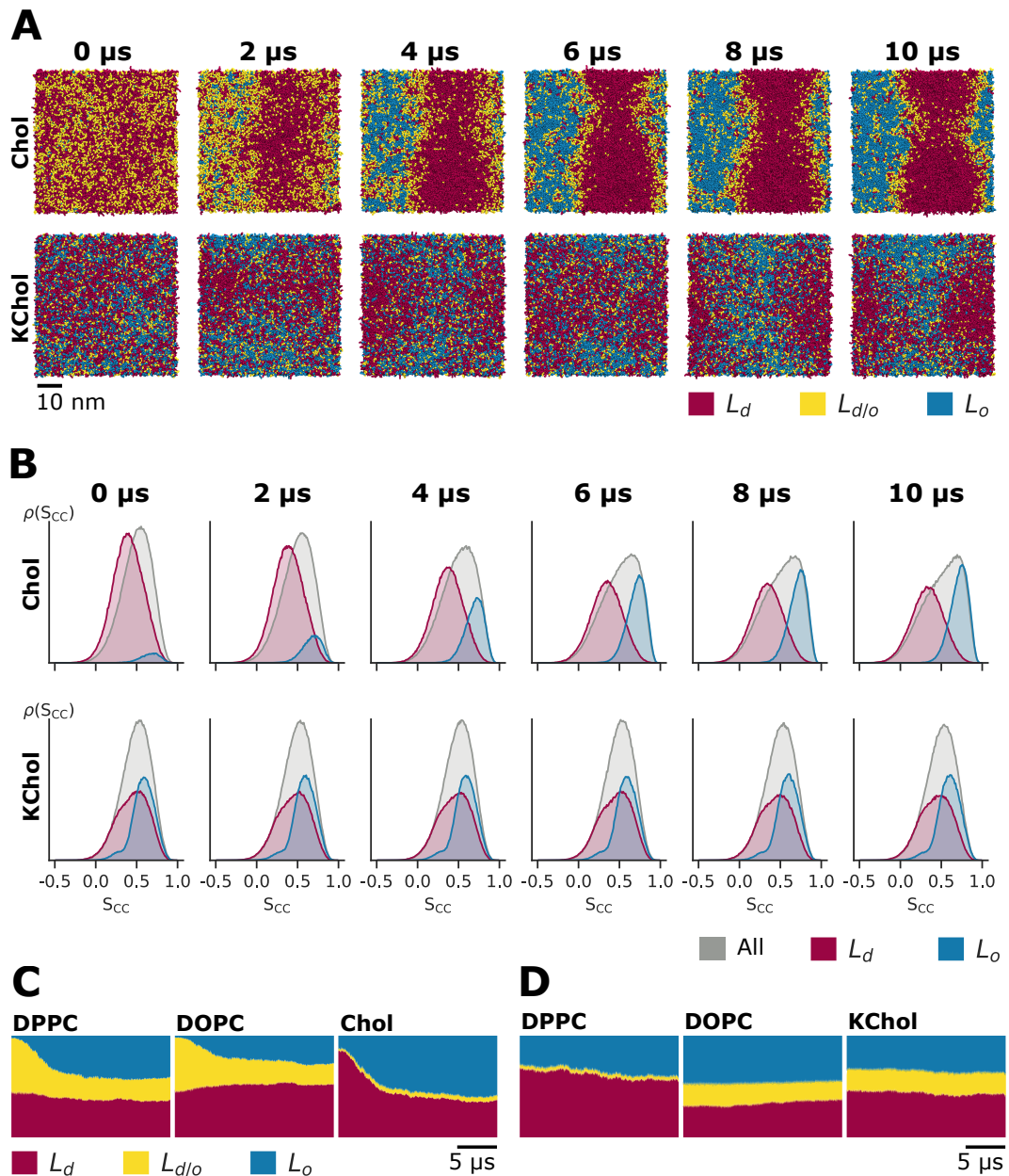
process, I constructed hidden Markov models (HMM) based on lipid thicknesses to assign each lipid molecule at each frame to one of three states: ordered ( $L_o$ ), disordered ( $L_d$ ), or intermediate ( $L_{d/o}$ ). Here, the thickness of a lipid refers to the extent of the lipid in the  $z$  - dimension; tilted lipids and lipids with disordered acyl tails will have smaller thicknesses than non-tilted lipids or those with ordered and extended acyl tails. To construct the HMM, I followed the methodology proposed by Park and Im.<sup>206</sup> Briefly, I first

calculated the thickness of each phospholipid molecule as the mean thickness in  $z$  of its two acyl tails, and the thickness of each sterol as the extent in  $z$  of the entire molecule. Then, for each lipid species, I binned these thicknesses into nine states, which served as the emission states of the model. I used a Gaussian mixture model to initialize the parameters ( $\mu, \sigma$ ) of the hidden Gaussian distributions, before using the Baum-Welch algorithm to fit the model parameters based on the emission states and initial parameters. Finally, I used the Viterbi algorithm to decode the most likely time series of hidden states ( $L_o, L_d$ , or  $L_{d/o}$ ) for each lipid.

The lateral distribution of ordered states can be seen in Figure 3.5. At 20  $\mu\text{s}$ , the  $L_o$  and  $L_d$  regions of the DPPC:DOPC:Chol membrane clearly correspond to the ordered and disordered regions, respectively, seen in Figure 3.4B. Further, the  $L_{o/d}$  lipids are predominantly found at the  $L_o$ - $L_d$  interface, giving us confidence that the HMM has accurately assigned lipids to the correct ordered state.

At 0  $\mu\text{s}$ , very few lipids in the DPPC:DOPC:Chol membrane are in the  $L_o$  state — almost all lipids are either  $L_d$  or  $L_{d/o}$  (Figure 3.5A). In particular, cholesterol is mostly  $L_d$  whereas the phospholipids are predominantly in the intermediate  $L_{d/o}$  state (Figure 3.6C). These  $L_d$  and  $L_{d/o}$  lipids are initially evenly distributed within the bilayer, with no sign of  $L_o$  domains. The domain formation process begins with a demixing of the  $L_d$  and  $L_{d/o}$  lipids (Figure 3.6A).  $L_d$  Chol then proceeds to become more ordered (Figure 3.6C). This in turn causes the DPPC and DOPC molecules in the intermediate state to also transition into the ordered  $L_o$  state. This transition from  $L_d$  and  $L_{d/o}$  to  $L_o$  is almost complete by 5  $\mu\text{s}$ . Over time, the boundary between the  $L_o$  and  $L_d$  regions becomes more well-defined, and the phase separation is nearly complete by 10  $\mu\text{s}$  (Figure 3.6A). By 20  $\mu\text{s}$ , there are two clear phases present in the DPPC:DOPC:Chol membrane, and there is a significant enrichment of  $L_o$  lipids around other  $L_o$  lipids (Figure 3.5C).

Conversely, in the DPPC:DOPC:KChol membrane there is little change in the fraction of lipids in  $L_o$  state over time (Figure 3.6D). Instead, the  $L_d$  and  $L_o$  regions seen in Figure 3.5B form via a lateral demixing of the ordered and disordered lipids. Unlike in the DPPC:DOPC:Chol membrane, this demixing is not followed by an increased ordering of the  $L_o$  acyl tails (Figure 3.6B). In fact, there is very little change in the ordering

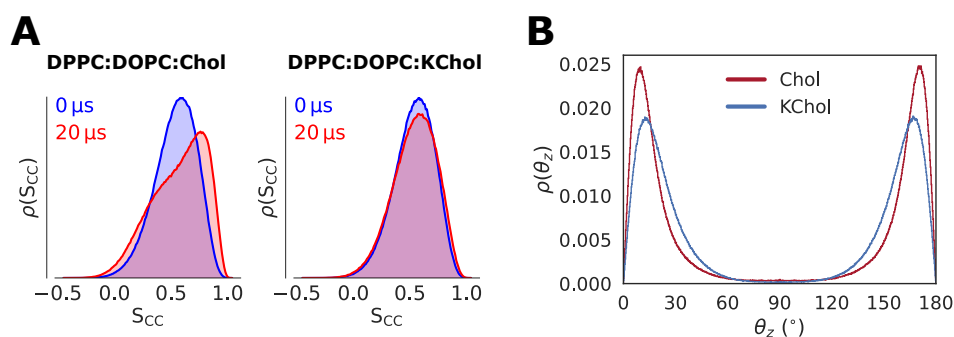


**Figure 3.6:** (A) Lateral distribution of  $L_d$ ,  $L_o$  and intermediate ( $L_{d/o}$ ) lipids throughout the first 10  $\mu$ s of simulation time. (B) Coarse-grained order parameter,  $S_{CC}$ , for phospholipids throughout the first 10  $\mu$ s of simulation time. (C, D) Fraction of each lipid species in  $L_d$ ,  $L_o$  or intermediate states over time.

of acyl tails in the DPPC:DOPC:KChol membrane throughout the simulation (Figure 3.7A). This is in clear contrast to the acyl tails in the DPPC:DOPC:Chol membrane, which become significantly more ordered. The result is two co-existing macroscopic phases in the DPPC:DOPC:Chol membrane, but smaller, less stable nanodomains in the DPPC:DOPC:KChol membrane.

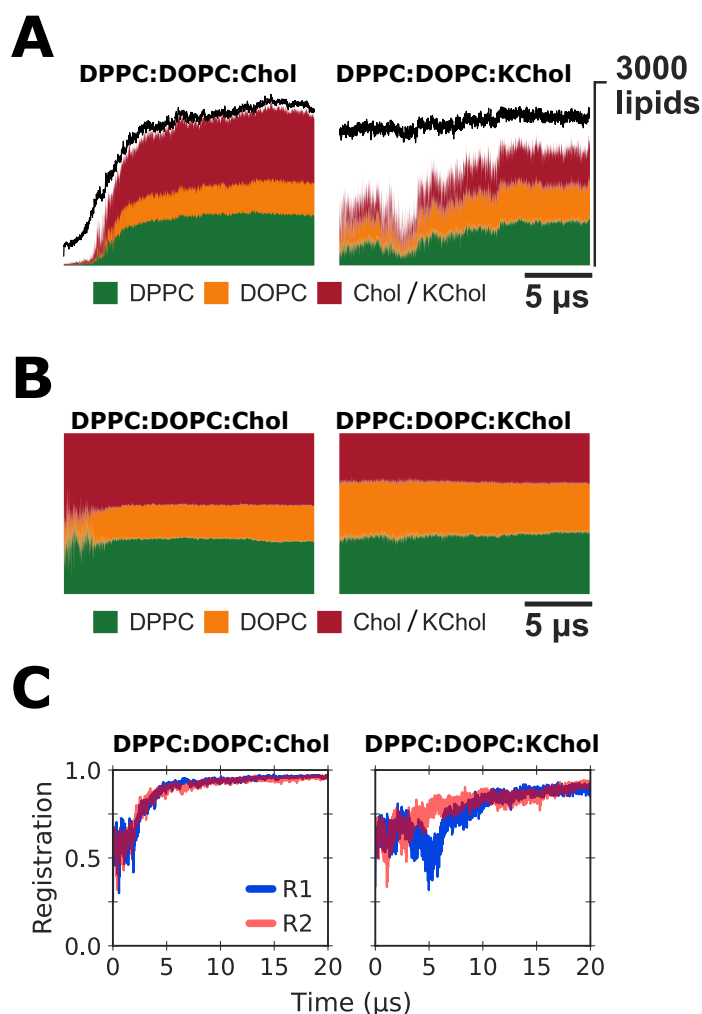
**Table 3.1:** Mean area per lipid, coarse-grained order parameter ( $S_{CC}$ ), and membrane thickness over the final 4  $\mu$ s of simulation time. The standard error is 0.01 or less for all values.

System		Lipid					
		DPPC		DOPC		Chol/KChol	
		$L_d$	$L_o$	$L_d$	$L_o$	$L_d$	$L_o$
Area ( $\text{\AA}^2$ )	DPPC:DOPC:Chol	66.5	50.2	69.3	51.1	35.4	29.8
	DPPC:DOPC:KChol	59.6	52.6	62.8	55.7	35.6	32.0
$S_{CC}$	DPPC:DOPC:Chol	0.38	0.75	0.28	0.65	-	-
	DPPC:DOPC:KChol	0.48	0.68	0.31	0.55	-	-
Thickness ( $\text{\AA}$ )	DPPC:DOPC:Chol	39.4	43.9	39.1	43.7	-	-
	DPPC:DOPC:KChol	40.7	42.6	40.3	41.8	-	-



**Figure 3.7:** (A) Coarse-grained order parameter,  $S_{CC}$ , of the phospholipids. (B) Sterol orientation, defined as the angle between the positive  $z$  axis and the vector from bead  $R_5$  to bead  $R_1$ .

The  $L_o$  and  $L_d$  regions in the DPPC:DOPC:KChol membrane are much more alike than those of the DPPC:DOPC:Chol mixture (Table 3.1). Generally, the  $L_d$  lipids of the KChol membrane are less disordered than those of the Chol membrane, while the  $L_o$  lipids of the KChol membrane are less ordered than those of the Chol membrane. However, the  $L_d$  lipids in the KChol membrane have a larger area per lipid than the  $L_d$  lipid molecules in the Chol membrane, albeit only by 0.2  $\text{\AA}$ . This is likely because KChol adopts a wider range of orientations in the membrane (Figure 3.7B). Chol has a strong tendency to be oriented at around 10° (and 170°), whereas KChol has a broader distribution of orientations with a peak at around 15° (and 165°). KChol adopts a wider range of orientations in the membrane so that its hydrophilic ketone group can be exposed to the solvent. This increased orientational freedom of KChol will likely lead to an increased area per lipid. An implication of this is that KChol will disrupt the local



**Figure 3.8:** (A) Number of each lipid species in the largest cluster of  $L_o$  lipids. The black curve shows the total number of  $L_o$  lipids present. (B) Fractional composition of the largest  $L_o$  domain over time. (C) Domain registration.

packing of lipids in the  $L_o$  phase.<sup>60,176</sup> This therefore explains why the order gradient in the DPPC:DOPC:KChol membrane does not increase after the demixing of  $L_o$  and  $L_d$  lipids, unlike in the DPPC:DOPC:Chol membrane.

Within the two mixtures, there is a difference in the lipid composition of their respective  $L_o$  domains (Figure 3.8B). The  $L_o$  domain of the DPPC:DOPC:KChol membrane is enriched in DPPC (with a DPPC:DOPC:KChol ratio of 0.38 : 0.29 : 0.33), and there is no significant change in its composition over the course of 20  $\mu\text{s}$ . On the other hand, in the DPPC:DOPC:Chol membrane, small ordered clusters enriched in Chol form at the beginning of the simulation. Then, the onset of nanodomain formation is associated with an increase in other lipid species, especially DPPC, in the Chol-enriched



ordered clusters. Despite the other species joining the  $L_o$  domain, it remains enriched in Chol even at 20  $\mu\text{s}$  (with a DPPC:DOPC:Chol ratio of 0.34 : 0.22 : 0.44).

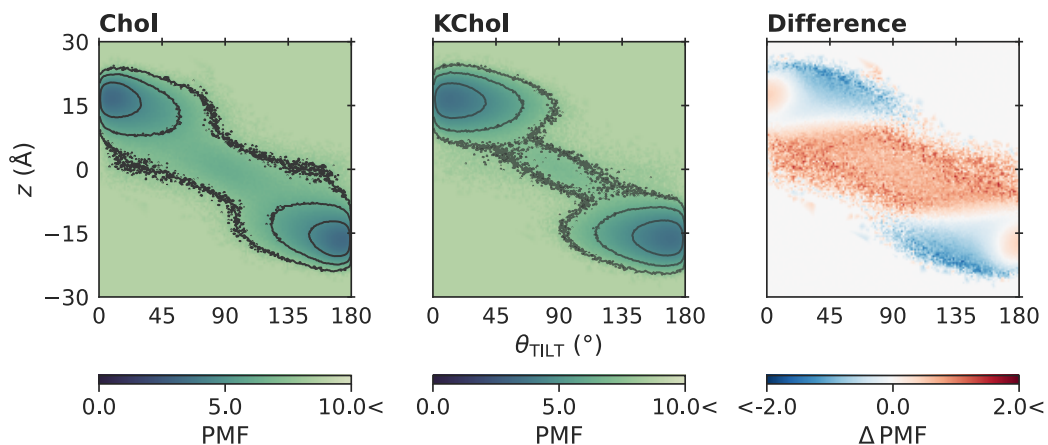
The resulting  $L_o$  domain of DPPC:DOPC:Chol is not only more ordered than that of DPPC:DOPC:KChol, it is also larger and more stable. Figure 3.8A shows the largest cluster of  $L_o$  lipids in the upper leaflet of each mixture over time. From around 2  $\mu\text{s}$  onward, almost all  $L_o$  lipids in the DPPC:DOPC:Chol mixture are part of the  $L_o$  domain. Conversely, the largest  $L_o$  cluster in the DPPC:DOPC:KChol membrane dissociates at around 5  $\mu\text{s}$  before reforming, and even at 20  $\mu\text{s}$  no more than 80% of  $L_o$  lipids are in the  $L_o$  domain. This dissociation of the  $L_o$  domain coincides with a decrease in the interleaflet registration of the  $L_o$  domains (Figure 3.8C). Domain registration in the Chol membrane, however, also equilibrates faster and is more stable.

**Table 3.2:** Flip-flop rate,  $k$  ( $\times 10^6 \text{ s}^{-1}$ ), of cholesterol and 7-ketocholesterol, and lateral diffusion coefficient,  $D_{xy}$  ( $\times 10^{-7} \text{ cm}^2 \text{ s}^{-1}$ ), of lipids in the  $L_o$  and  $L_d$  domains. Values calculated using the final 4  $\mu\text{s}$  of each trajectory.

	$k$	$D_{xy}$	
		$L_o$	$L_d$
Chol	1.80	3.4	5.1
KChol	0.64	3.2	3.7

It is not only the structure of the  $L_o$  domains that changes upon oxidation, but also the dynamics of the molecules within the domains. In the DPPC:DOPC:Chol mixture, the lateral diffusion of  $L_o$ -domain lipids is 1.5 times slower than those in the largest cluster of  $L_d$  lipids (Table 3.2), which is in line with atomistic simulations and experimental measurements.<sup>121</sup> This difference is significantly reduced in the DPPC:DOPC:KChol membrane (Table 3.2) — a result of the fact that this membrane forms a nanodomain, with a smaller order gradient, rather than a microdomain, with a larger order gradient.<sup>207</sup>

I also find a substantial affect on interleaflet dynamics upon cholesterol oxidation. The rate of cholesterol flip-flop in the DPPC:DOPC:Chol membrane is around 3 times faster than in the DPPC:DOPC:KChol membrane (Table 3.2). The reduced flip-flop rate for KChol is the result of an increased free energy barrier to translocation compared to Chol. The potentials of mean force (PMF) for the height and orientation of Chol and KChol within the membranes are shown in Figure 3.9. For Chol, there is a barrier



**Figure 3.9:** Potential of Mean Force (PMF) of sterol orientation ( $\theta_{\text{TILT}}$ ) and height ( $z$ ). For Chol, there is a free energy barrier of around  $5 \text{ kcal mol}^{-1}$  in the region  $-12 \text{ \AA} < z < 12 \text{ \AA}$ ,  $65^\circ < \theta_{\text{TILT}} < 115^\circ$ . The difference plot shows  $PMF_{\text{KChol}} - PMF_{\text{Chol}}$ ; red regions are less favorable for KChol and blue regions more favorable.

to flip-flop of around  $5 \text{ kcal mol}^{-1}$ . This is due to the unfavorable desolvation of the hydroxyl group during the flip-flop process, which occurs as the sterol crosses through the hydrophobic core of the bilayer ( $-12 \text{ \AA} < z < 12 \text{ \AA}$ ) and rotates to align with lipids in the apposing leaflet ( $65^\circ < \theta_{\text{TILT}} < 115^\circ$ ). The ring-oxidation of cholesterol into 7-ketocholesterol further increases this barrier by another  $2 \text{ kcal mol}^{-1}$  (Figure 3.9). This is because both the ketone and hydroxyl groups must be desolvated for flip-flop to occur. The result is that KChol is less likely to move to the midplane and thus there is a reduced rate of flip-flop in the DPPC:DOPC:KChol mixture.

### 3.4 Conclusion

I have shown that the macroscopic phase separation seen in a DPPC:DOPC:Chol membrane is disrupted by the autoxidation of cholesterol into 7-ketocholesterol. In a DPPC:DOPC:KChol membrane, there is instead nanodomain formation that is more akin to that expected in the plasma membrane.<sup>31,35,40,182</sup> This disruption arises from the hydrophilicity of the ketone group of KChol, which has two effects on the domain formation. First, to allow for the hydration of the ketone group, KChol adopts a broader distribution of orientations in the membrane. This disrupts the local packing of lipids, inducing disorder in  $L_o$  regions.<sup>60,175,176</sup> Second, the reason Chol prefers to interact with DPPC over DOPC is because DPPC is better at shielding the hydrophobic rings of cholesterol

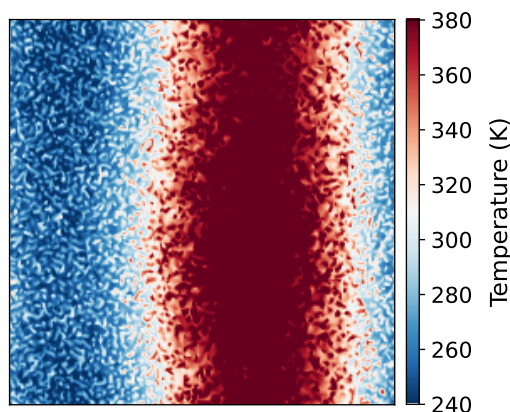
from the surrounding solvent. The tetracyclic rings of KChol, however, are less hydrophobic due to the presence of the ketone group; KChol tends to expose this moiety to the solvent rather than seeking refuge in the hydrophobic core of the bilayer, meaning KChol has less of a preference for DPPC over DOPC. This reduced preference for DPPC suppresses the lateral demixing of lipids, which in turn disrupts the liquid-liquid phase separation seen in the DPPC:DOPC:Chol mixture. In addition, the hydrophilicity of the KChol ketone group suppresses translocation. The reduced rate of translocation has little effect on domain registration at in the equilibrated mixture studied here, but in more physiologically-relevant mixtures sterol flip-flop is required for interleaflet domain registration.<sup>183</sup>

Chol preferentially mixes with sphingolipids over glycerophospholipids for the same reason it prefers DPPC over DOPC - sphingolipids are better at shielding cholesterol from the surrounding solvent.<sup>196</sup> Therefore, it is expected that the increased hydrophilicity of KChol will diminish its affinity for sphingolipids compared to cholesterol. This would either disrupt the formation of Chol-sphingolipid nanodomains in biological membranes, or at least reduce the lateral order gradient as seen here. The reduced order gradient would have implications lipid-raft protein-sorting due to the hydrophobic mismatch between raft regions and their embedded proteins. Such implications include the disruption of cell-signaling pathways, via which KChol is known to induce apoptosis.

### 3.5 Addendum

The simulations in this chapter were performed using the recommended new-RF parameter set for running Martini simulations with GROMACS.<sup>136</sup> This parameter set, however, uses the standard LINCS parameters (`lincs_iter=1`, `lincs_order=4`) rather than the more conservative parameters that the cholesterol model was parameterised with (`lincs_iter=2`, `lincs_order=8`).<sup>139</sup> This results in an unphysical temperature gradient of over 100 K across the membrane. (Figure 3.10).

**Physical origin of the temperature gradient** The MARTINI model of cholesterol acts as a heat sink when simulated using the default LINCS parameters. This is down to the high-frequency motion of the virtual sites in the cholesterol model. Due to these high-



**Figure 3.10:** A lateral temperature gradient of over 100 K is present across the ordered and disordered regions of the DPPC:DOPC:Chol membrane simulated in this Chapter. This is due to use of the Melo et al.<sup>139</sup> cholesterol model in conjunction with the default GROMACS parameters for the LINCS constraint algorithm.<sup>149</sup>

frequency motions, the LINCS algorithm does not converge within the default number of iterations. This lack of convergence means that the total energy of the system is not conserved. Energy is drained from the system via cholesterol, then pumped back in by the thermostat via other molecules. This causes a lateral temperature gradient across the system, and an *artificial* phase separation.<sup>149</sup>

**Approaches to avoid the temperature gradient** Energy is conserved and no lateral temperature gradient is observed when either:

- more conservative LINCS parameters are used, which ensure the algorithm converges each timestep
- each lipid species is coupled to a separate thermostat, meaning energy cannot be drained via cholesterol as the thermostat will keep the cholesterol molecules at the desired temperature
- a significantly smaller timestep (10 fs) is used, which decreases the atomic displacements at each timestep and thus ensures the LINCS algorithm converges even when using the default parameters

However, when the artificial temperature gradient disappears, the  $L_o/L_d$  phase separation disappears with it. This means that the formation of  $L_o$  domains is an artefact of the aphysical temperature gradient. Therefore, whilst the results in this Chapter are in line

with previous experimental<sup>178</sup> and simulation<sup>60,175,176</sup> studies, the findings are unreliable. In light of this, I withdrew the preprint and have not submitted the article for peer review.

## Chapter 4

# Two Coexisting Membrane Structures Are Defined By Lateral and Transbilayer Interactions Between Sphingomyelin and Cholesterol

This Chapter was published in *Langmuir* in 2020 and is reproduced here with permission from: Smith, P; Quinn, P.J.; Lorenz, C.D., ‘Two Coexisting Membrane Structures Are Defined By Lateral and Transbilayer Interactions between Sphingomyelin and Cholesterol’, *Langmuir*, DOI:10.1021/acs.langmuir.0c01237. Copyright American Chemical Society 2020.

**Summary of the work** Cholesterol and sphingomyelin are two of the most important lipids in mammalian plasma membranes. In this chapter, I present my work on equimolar mixtures of cholesterol and palmitoylsphingomyelin. I show that this equimolar mixture forms two coexisting bilayer structures that are primarily distinguished by their transbilayer thickness, as detected via both small-angle x-ray scattering (SAXS) and wide-angle x-ray scattering (WAXS) measurements. These are two **coexisting** bilayer structures that may, for example, correspond to the liquid-ordered and liquid-disordered regions of a phase-separated membrane. Using Hidden Markov Models based on local phosphate-phosphate distances, I uncover the interatomic interactions that give rise to the two

distinct transbilayer thicknesses observed via SAXS. Then, through an unsupervised clustering of cholesterol-sphingomyelin conformations sampled in MD simulations, I identify four distinct modes of interaction between these lipid species. I argue that one mode in particular may explain why cholesterol preferentially mixes with sphingomyelin over glycerophospholipids, which is a long-standing observation without a clear answer.

**Author contributions** I performed all MD simulations and analysis of the trajectories, and produced all figures in the main text and the Supporting Information. Peter Quinn performed the SAXS and WAXS measurements and analysed the data. Peter Quinn wrote the first draft of the introduction, as well as the first paragraph of the results section. I wrote the first draft of the remainder of the manuscript. All authors contributed to and approved the manuscript in its final form.

The Supporting Information for the article is available in Appendix A.

## Two Coexisting Membrane Structures Are Defined by Lateral and Transbilayer Interactions between Sphingomyelin and Cholesterol

Paul Smith, Peter J. Quinn,\* and Christian D. Lorenz\*



Cite This: *Langmuir* 2020, 36, 9786–9799



Read Online

ACCESS |



Metrics & More

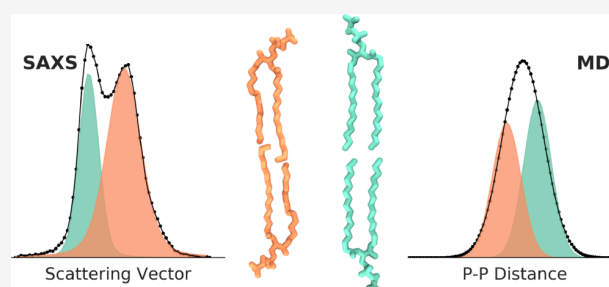


Article Recommendations



Supporting Information

**ABSTRACT:** The structure of fully hydrated bilayers composed of equimolar proportions of palmitoylsphingomyelin (PSM) and cholesterol has been examined by synchrotron X-ray powder diffraction and atomistic molecular dynamics (MD) simulations. Two coexisting bilayer structures, which are distinguished by the transbilayer phosphate–phosphate distance of coupled PSM molecules, are observed by diffraction at 37 °C. The MD simulations reveal that PSM molecules in the thicker membrane are characterized by more ordered, more extended, and less interdigitated hydrocarbon tails compared to those in the thinner membrane. Intermolecular hydrogen bonds further distinguish the two bilayer structures, and we observe the disruption of a sphingomyelin intermolecular hydrogen bond network induced by the proximity of cholesterol. Through an unsupervised clustering of interatomic distances, we show for the first time that the asymmetry of phospholipids is important in driving their interactions with cholesterol. We identify four distinct modes of interaction, two of which lead to the dehydration of cholesterol. These two modes of interaction provide the first description of precise physical mechanisms underlying the umbrella model, which itself explains how phospholipids may shield cholesterol from water. The most dehydrating mode of interaction is particular to the *N*-acylated fatty acid moiety of PSM and thus may explain the long-held observation that cholesterol preferentially mixes with sphingomyelins over glycerophospholipids.



### INTRODUCTION

Sphingomyelin and cholesterol are prominent lipids of the plasma membrane of animal cells. Their colocalization and enrichment in the plasma membrane occur despite different sites of synthesis; cholesterol is synthesized in the endoplasmic reticulum whereas sphingolipids are synthesized in the Golgi. The plasma membrane contains approximately 80% of the total cellular cholesterol, which represents about 45% of the total lipids present in the membrane.<sup>1</sup> This means that the molar proportion of cholesterol is at least equal to or exceeds that of sphingomyelin in plasma membranes. The ratio is initially established by nonvesicular transport of precursors mediated by ceramide transfer proteins and oxysterol binding proteins which serve to couple the metabolism of the two lipids and regulate their distribution in subcellular membranes.<sup>2</sup> Enrichment of the two lipids proceeds in membrane transformations that take place along the secretory pathway culminating in a differentiated plasma membrane.<sup>3</sup> The particular mode of interaction between the sterol and phospholipid is the subject of considerable interest because they interact to form a liquid-ordered phase said to organize a range of membrane-mediated physiological processes. A significant factor underlying the formation of this liquid-ordered phase is the favorable interactions between sphingomyelin and cholesterol, which are mediated by hydrogen bonds originating from the amide and hydroxyl groups acting as both

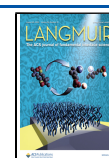
hydrogen bond donors and acceptors and the hydrogen bond acceptor of the amide carbonyl group.<sup>4</sup> These hydrogen bonding properties are the primary feature that distinguishes interactions between sphingomyelin and glycerophospholipids. Furthermore, molecular species of sphingomyelin tend to be more saturated than their glycerolipid counterparts, which additionally contributes to preferential interactions between sphingomyelin and cholesterol in biological membranes. While sphingomyelin predominates in the outer leaflet of the plasma membrane, the transbilayer distribution of cholesterol appears to vary from one membrane to another.<sup>5</sup>

The lateral association of condensed structures in biological membranes forms so-called lipid rafts. The existence of lipid rafts in plasma membranes is conjectural, and their role in the creation of platforms for the assembly of membrane components responsible for the reception of extracellular ligands and the transduction of the signals to effectors residing at the cytoplasmic surface of the membrane has yet to be

Received: April 27, 2020

Revised: July 20, 2020

Published: July 23, 2020





convincingly demonstrated. Recent reviews<sup>6,7</sup> consider the evidence for lipid rafts and give an optimistic perspective of the tools now available to shed light on lipid domain formation in biological membranes. To that end, a recent neutron scattering study provided evidence of lateral lipid domain formation in a bacterial cell membrane.<sup>8</sup> The properties of condensed structures formed by the interaction between sphingomyelin and cholesterol result in their phase separation from liquid-disordered regions by processes that involve differences in line tensions that develop between domains.<sup>9</sup> The thickness of bilayers of liquid-ordered structures and the surface area of lipid–sterol complexes increases linearly with increasing proportions of cholesterol when formed in mixtures of glycerophospholipids.<sup>10</sup> The lateral domain size of the liquid-ordered structure also increases with decreasing thickness of the liquid-disordered phase.<sup>11</sup> Furthermore, the phase-separated domains of thicker liquid-ordered regions from thinner liquid-disordered structures in ternary mixtures of glycerophospholipids and cholesterol were found to be in register across the bilayer, with relative domain sizes that depended on the bending moduli of the respective domains<sup>12</sup> and the lipid density mismatch between the phases.<sup>13</sup> The transbilayer distribution of cholesterol is said to be determined by the asymmetric distribution of molecular species of phospholipids across the bilayer, while the resulting bending modulus is thought to depend on the interaction of the particular membrane lipids with cholesterol.<sup>14</sup>

Computer simulations of equimolar complexes of stearyl-sphingomyelin and cholesterol in bilayers of fluid dioleoyl-phosphatidylcholine have also showed that the liquid-ordered structure was approximately 4.5 Å thicker than the surrounding phospholipid bilayer.<sup>15</sup> Greater differences in bilayer thickness between the liquid-ordered phase of egg sphingomyelin–cholesterol (51.9 Å) and surrounding glycerophospholipid (39 Å) have been reported.<sup>16</sup> These differences between the relative thickness of liquid-ordered and liquid condensed structures are the basis of theories of hydrophobic mismatch when sorting raft-associating membrane components.<sup>17</sup> Electron density calculations from simulations<sup>18</sup> and X-ray diffraction methods<sup>19</sup> of binary mixtures of different sphingomyelins and cholesterol, however, showed no significant differences in bilayer thickness between pure sphingomyelin bilayers and sphingomyelin bilayers containing cholesterol, even when complexes of sphingomyelin and cholesterol coexisted in bilayers of fluid glycerophospholipid.<sup>20</sup> The clear differences between the way cholesterol interacts with glycerophospholipids on the one hand and sphingomyelin on the other indicate that models of membrane rafts based on purely glycerophospholipid–cholesterol mixtures are of doubtful biological significance.

Studies of the atomistic interactions that underlie the observed properties of mixtures of phospholipids and cholesterol have focused mainly on glycerophospholipids rather than sphingomyelin and have been performed at temperatures remote from those relevant to the purported physiological functions of the structures that are created. In this study, we have characterized the structure of fully hydrated bilayers consisting of an equimolar proportion of palmitoyl-sphingomyelin and cholesterol using synchrotron X-ray diffraction methods. Simulations of this mixture have been performed to extract parameters related to the two coexisting bilayers observed by diffraction as well as the modes of interaction between the lipids and their resulting mutual

orientations. A hidden Markov model, based on local phosphate to phosphate distances, was constructed to assign each lipid to one of the two coexisting bilayer structures, which are defined by their respective thicknesses. The preponderance of intermolecular hydrogen bonds between the cholesterol oxygen and the phosphate and sphingosine hydroxyl moieties further distinguishes the two bilayer structures. An unsupervised clustering of sphingomyelin–cholesterol pairs has been performed on the basis of the distances between heavy (non-hydrogen) atoms of the two lipids, resulting in the identification of four distinct modes of interaction between sphingomyelin and cholesterol. The hydration of cholesterol in the different conformations is considered in the context of the umbrella model,<sup>21</sup> which describes hydration forces at the bilayer interface and mixing of the constituent lipids.

## METHODS

**X-ray Diffraction.** *N*-Palmitoyl-D-erythro-sphingosylphosphorylcholine (PSM) was purchased from Avanti Polar Lipids (Alabaster, AL). Cholesterol was purchased from Sigma-Aldrich (U.K.). The lipids dissolved in warm (45 °C) chloroform/methanol (2:1, vol/vol) were mixed in an equimolar proportion. The solvent was subsequently evaporated under a stream of oxygen-free dry nitrogen at 45 °C, and any remaining traces of solvent were removed by storage under high vacuum for 2 days at 20 °C. The dry lipids were hydrated with deionized water to give a dispersion of 25 wt % lipid. The dispersion was sealed under argon and annealed by 50 thermal cycles between –20 and 90 °C to ensure complete mixing. A sample of dispersed lipid (20 μL), sandwiched between thin mica windows 0.5 mm apart, was equilibrated at 37 °C and examined by synchrotron X-ray methods on Station 2.1 of the Daresbury SRS (U.K.).<sup>22</sup> The SAXS intensity profiles were subjected to analysis using PeakFit (v4.12; Systat Software Inc.) software. The first four orders of small-angle X-ray scattering (SAXS) reflection and the wide-angle X-ray scattering (WAXS) profile could all be fitted by Voigt area functions with fitting coefficients greater than  $R^2 = 0.99$ . We observe two SAXS unit cells, SAXS<sub>1</sub> and SAXS<sub>2</sub>, and two WAXS unit cells, WAXS<sub>1</sub> and WAXS<sub>2</sub>. Background subtraction was carried out on each diffraction band; angular correction of the scattering intensity was performed to yield the relative mass by the method described elsewhere.<sup>23</sup>

**Simulation Protocol.** We have performed all-atom molecular dynamics (MD) simulations of a lipid bilayer consisting of an equimolar mixture of PSM and cholesterol. The bilayer, which contains 100 PSM molecules and 100 cholesterol molecules in each leaflet, was built using the CHARMM-GUI membrane builder.<sup>24–26</sup> This bilayer was surrounded by at least 6000 water molecules (Table 1).

**Table 1. List of Simulations Detailing the Number of Water Molecules in the System ( $N_w$ ) and the Method Used to Encourage Lipid Mixing, Either a Constant Elevated Temperature or a Cyclical Process of Heating and Cooling**

replica	$N_w$	mixing method
1	16 768	cyclical
2	6000	constant
3	6000	constant
4	6000	constant

The MD simulations reported in this article were performed using the GROMACS simulation package.<sup>27,28</sup> For these simulations, we used the CHARMM36 force field to model the interactions of the PSM<sup>29</sup> and cholesterol.<sup>30</sup> Water was treated with the CHARMM-modified TIP3P water model.<sup>31</sup> The Lennard-Jones and electrostatic nonbonded interactions were both cut off at 1.2 nm, while the long-range electrostatic interactions were calculated using the fast, smooth particle-mesh Ewald algorithm.

We used the simulation protocol and the corresponding input files provided by CHARMM-GUI<sup>26</sup> to minimize the potential energy, equilibrate the temperature, and then equilibrate the density of the bilayer. Then, for one replica, we ran an initial production simulation for 50 ns at a temperature of 310 K and a pressure of 1 bar. After this initial production simulation, we used simulated annealing in order to enhance the mixing of the two components of our membrane. The simulated annealing stage consisted of a series of temperature cycles where the bilayer and water were heated from 310 to 400 K over 50 ps and then they were kept at a constant temperature of 400 K for 900 ps and finally cooled back down from 400 to 310 K over another 50 ps. In total, this simulated annealing stage was performed for 250 ns such that we performed 250 of the 1 ns thermal cycles. During the simulated annealing stage, we employed the Berendsen thermostat and the semi-isotropic Berendsen barostat to control the temperature and pressure, respectively. For three other replicas, after equilibration we simulated each system at a constant, elevated temperature of 400 K for 200 ns before cooling the systems to 310 K over a 50 ns period. We then ran each of the four replicas for at least 500 ns at a temperature of 310 K and a pressure of 1 bar. In these production simulations, we employed the Nosé–Hoover thermostat and the semi-isotropic Parrinello–Rahman barostat to control the temperature and pressure, respectively.

**Analysis Methods.** Unless stated otherwise, we use the final 150 ns of the production simulations for analysis, with coordinates stored every 100 ps. Analysis scripts were written in Python with the use of MDAnalysis,<sup>32,33</sup> Scikit-learn,<sup>34</sup> HMMLearn,<sup>35</sup> UMAP,<sup>35</sup> HDBSCAN,<sup>36,37</sup> and Scipy.<sup>38</sup>

**Identification of Two Bilayers.** Hidden Markov models (HMMs), based on the local lipid composition<sup>39,40</sup> or the lipid thickness and area,<sup>41</sup> have previously been used to identify the lateral phase separation of lipids in MD simulations. Here, we use an HMM to directly compare our simulation data with the bilayer thickness of the two unit cells observed via small-angle X-ray scattering. Our approach to constructing the HMM is based on the methodology described by Park and Im.<sup>41</sup> In the construction of the HMM, we assume that there are two hidden states (bilayers), as revealed by the SAXS measurements. We construct an HMM, based on local transbilayer phosphate–phosphate distances, to assign each PSM molecule to a thick ( $B_1$ ) or thin ( $B_2$ ) bilayer at each frame. We first identify the transbilayer couple of each PSM molecule. We do so by finding the shortest distance in the  $xy$  plane between the reference PSM C2S atom and the C2S atom on any PSM in the opposing leaflet (Figure S1 for CHARMM atom names). Then we use a Gaussian mixture model (GMM) to decompose the distribution of P–P distances of coupled molecules into two states. The phosphate–phosphate distances are binned into nine states, which serve as the emission state signals for training the HMM. To determine the initial emission probabilities, we integrate each Gaussian obtained from the GMM over the nine emission states. We use the Baum–Welch algorithm to determine the transition matrix for hidden states and then the Viterbi algorithm to decode the most likely sequence of hidden states ( $B_1$  or  $B_2$ ) for each lipid. The difference in mean thickness between  $B_1$  and  $B_2$  is on the order of 1 Å. To determine whether the membrane thickness distributions of  $B_1$  and  $B_2$  are statistically distinct, we therefore calculate the 95% confidence intervals of the mean thicknesses via bootstrapping with 1000 resamples. (See Grossfield et al.<sup>42</sup> for an excellent discussion on uncertainty quantification in molecular dynamics simulations.)

**Microscopic Origin of the Two Bilayers.** We measured the physical properties of PSM and cholesterol to understand the microscopic origin of the difference between  $B_1$  (or SAXS<sub>1</sub>) and  $B_2$  (or SAXS<sub>2</sub>). In doing so, we have determined the following structural properties of the bilayer. We measure the membrane thickness on the basis of C2S–C2S, P–P, and N–N distances of coupled PSM molecules. We also measure the area per lipid via a Voronoi tessellation<sup>43</sup> of atom positions, along with the lipid order parameter and the lipid tail thickness of the sphingosine (SPH) and N-linked fatty acid (FA) tails. The latter is calculated as the greatest difference in  $z$  between any two atoms of a given hydrocarbon tail. We calculate

the radial distribution functions of the oxygen atoms of water around the PSM phosphorus and cholesterol oxygen atom, from which the hydration is calculated by integrating over  $r$  up to the first minimum in  $g(r)$ . The extent of PSM interdigitation is calculated by constructing an intrinsic surface of the tails in each leaflet and determining the maximum amount of penetration into this surface by each PSM tail.

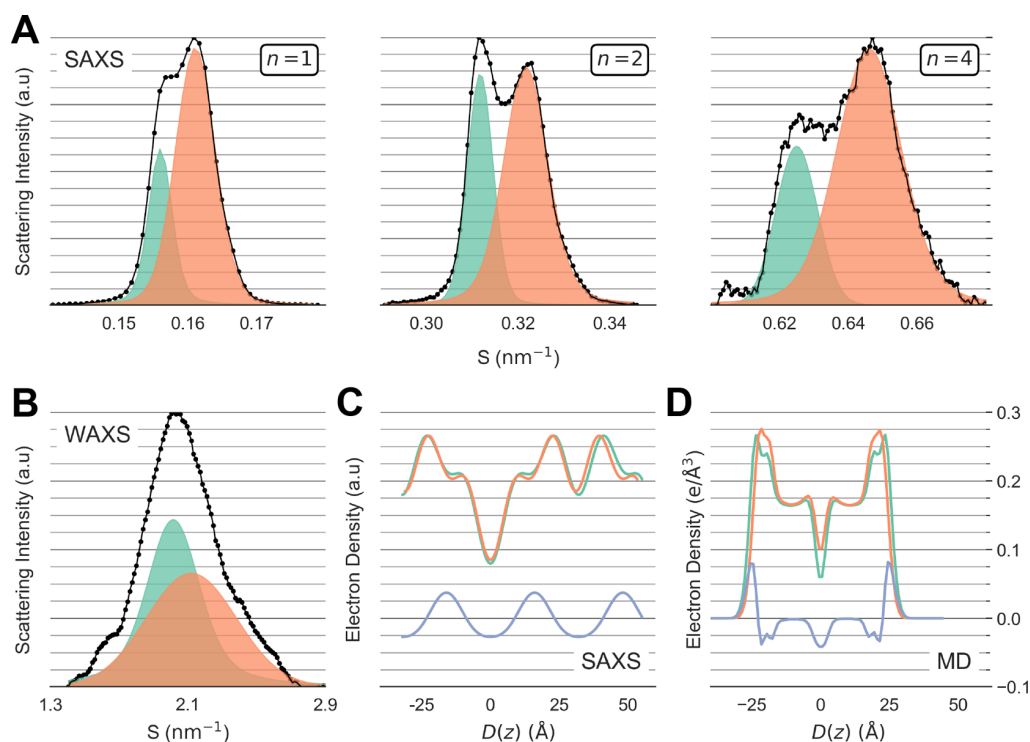
We have defined various angles, depicted in Figure S2, to describe the orientation of the different lipids and of different parts of the lipids within the bilayer. The orientation of the PSM headgroup is characterized by calculating the angle between the P–N vector ( $\theta_{\text{pnz}}$ ) or the C2S–P ( $\theta_{\text{cpz}}$ ) vector and the  $z$  axis. We consider the splay of the lipid tails ( $\theta_{\text{tails}}$ ) by calculating the angle made by the terminal methyl carbon atoms in the hydrocarbon tails and the C2S atom of PSM. The height at which PSM sits in a leaflet is calculated as the difference between the  $z$  coordinate of a PSM C2S atom and the mean  $z$  coordinate of all PSM C2S atoms in a given leaflet. The height at which cholesterol sits in a leaflet is calculated as the difference between the  $z$  coordinate of a cholesterol O3 atom and the mean  $z$  coordinate of all PSM C2S atoms in a given leaflet. The tilt of cholesterol is defined as the angle made between the vector formed between the C17 and C3 atoms of cholesterol and the  $z$  axis. We determine the hydrogen bonds formed between all polar groups of the PSM and cholesterol, using the hydrogen bond analysis tool<sup>44</sup> of MDAnalysis.<sup>32,33</sup>

**Comparison with SAXS and WAXS.** We compare the membrane thicknesses of  $B_1$  and  $B_2$  with those of SAXS<sub>1</sub> and SAXS<sub>2</sub>, respectively. The peaks in the relative electron densities of the bilayer profiles are due to the phosphate groups of PSM, and thus we compare the membrane  $d$  spacings with the P–P distances calculated via simulation. SAXS is unable to provide information on the lateral distribution of SAXS<sub>1</sub> and SAXS<sub>2</sub>, but we determine the local lipid composition of  $B_1$  and  $B_2$  via a Voronoi tessellation of atomic positions.

For comparison with the WAXS data, we assume that the distinct  $d$  spacings of WAXS<sub>1</sub> and WAXS<sub>2</sub> arise from the distance between the center of mass of cholesterol and the center of mass of the hydrocarbon tails of a neighboring PSM molecule. We calculate these distances from the simulation trajectory and then find mean values for PSM–cholesterol interactions within  $B_1$  and  $B_2$ . We also find the mean values of these distances for PSM interacting with the  $\alpha$  face of cholesterol and PSM interacting with the  $\beta$  face of cholesterol. Since the differences reported by WAXS are very small (on the order of  $1 \times 10^{-1}$  Å), we report distances calculated from simulation of up to  $1 \times 10^{-1}$  Å and measure 95% confidence intervals using bootstrapping with 1000 resamples (Table S2).

**Lateral Distribution of PSM around Cholesterol.** To understand the distribution of the PSM around cholesterol, we first identified all cholesterol–PSM pairs via a Voronoi tessellation at each snapshot of our simulated trajectory. Lipids sharing at least one edge of the Voronoi diagram are considered to be neighbors. We translated and rotated the cholesterol–PSM pairs such that the principal axes of the tetracyclic rings of cholesterol are centered at the origin and then aligned in the  $x$ ,  $y$ , and  $z$  dimensions. We then produce 2D density maps of select atoms of PSM around the cholesterol. We consider the distribution of PSM around cholesterol in terms of the two faces of the cholesterol: (i) the smooth  $\alpha$  face and (ii) rough  $\beta$  face, which is characterized by the protrusion of methyl groups from the tetracyclic rings of cholesterol (Figure S3).<sup>45–47</sup>

**Conformational Clustering of Cholesterol–PSM Pairs.** We use distances between selected heavy (non-hydrogen) atoms of PSM and cholesterol in each neighboring pair to identify distinct modes of interaction between them. To use every heavy atom in the clustering would result in a prohibitively high dimensional space ( $\sim 1344$  dimensions). We therefore use 11 atoms from each molecule of a given pair (Figure 6), chosen such that they capture the geometry of the lipids as well as the functional groups involved in hydrogen bonding. We also use every 10th frame (1 ns) to further reduce the computational cost. In total, we use 376 257 PSM–cholesterol pairs for the conformational clustering. We use the uniform manifold



**Figure 1.** Scattering intensity profile showing the first four orders of reflection in the small-angle X-ray scattering (SAXS, A) region and the wide-angle X-ray scattering (WAXS, B) region. The first, second, and fourth ( $n = 1, 2,$  and  $4$ ) orders in SAXS are shown. Each of the Bragg reflections were deconvolved, using Voigt area ( $G + L$ ) functions, into two components designated as SAXS<sub>1</sub>/WAXS<sub>1</sub> (green) and SAXS<sub>2</sub>/WAXS<sub>2</sub> (orange). (C) Electron density profiles of SAXS<sub>1</sub> (green) and SAXS<sub>2</sub> (orange) and the difference between them (SAXS<sub>1</sub>–SAXS<sub>2</sub>, purple) multiplied by 5 for emphasis. (D) Electron density profiles of PSM in the two bilayers identified via simulation,  $B_1$  (green) and  $B_2$  (orange), and the difference between them ( $B_1 - B_2$ , purple).

approximation and projection for dimension reduction (UMAP<sup>35</sup>) algorithm to embed the 121 distances of each pair into a 2D space. UMAP constructs a graph of the points in the high-dimensional space and then optimizes a low-dimensional representation such that the topological distance is preserved in the embedding.<sup>35</sup> Therefore, similar conformations that are close in the high-dimensional space will also be close in the reduced space. We can thus cluster the points in the embedded space to identify distinct modes of interaction between PSM and cholesterol. We set the  $n\_neighbors$  and  $min\_dist$  hyperparameters to 10 and 0.0, respectively, with the latter being a requirement if the points in the reduced space will later be clustered.<sup>35,48</sup> We use HDBSCAN<sup>36,37</sup> to cluster the PSM–cholesterol pairs in the embedded space. From this, we identify six conformations taken by neighboring PSM–cholesterol pairs. We compare the characteristics of each conformation through a consideration of the physical properties of PSM and cholesterol. We measure all properties of PSM and cholesterol that we use to compare  $B_1$  and  $B_2$  but focus our discussion on the importance of intermolecular hydrogen bonds in defining the six conformations. We also characterize the lateral distribution of PSM around cholesterol for each conformation. Along with the previously described  $\alpha$  and  $\beta$  faces of cholesterol,<sup>45–47</sup> we define analogous  $\alpha$  and  $\beta$  faces of PSM (Figure S3) and consider the distribution of cholesterol around PSM.

## RESULTS

**Identification of Two Bilayers.** The structure and properties of hydrated bilayers composed of equimolar proportions of PSM and cholesterol were examined by X-ray diffraction and all-atom molecular dynamics simulations. A multibilayer dispersion was characterized by synchrotron X-ray diffraction methods in a sample equilibrated at 37 °C. Bragg reflections from the first four orders of diffraction in the SAXS

region and the reflection in the WAXS region are shown in Figure 1.

Each of the SAXS reflections can be deconvolved into two unit cells, SAXS<sub>1</sub> and SAXS<sub>2</sub>, each of which consists of a lipid bilayer and a layer of water. Deconvolving the SAXS reflections into three unit cells does not improve the fitting, indicating the presence of two structures only. On the basis of the SAXS data alone, one may reasonably intuit that the two distinct bilayer structures, of SAXS<sub>1</sub> and SAXS<sub>2</sub>, arise from there being two distinct states of PSM. That is, there exist two conformational states of PSM that are distinguished by their thickness, and the transbilayer coupling of these states thus determines the membrane thickness. However, while we do find two distinct bilayers from the analysis of the MD trajectories, we do not see significant transbilayer coupling of PSM properties (Figure S4). This is because lipids do not take on only two conformational states.<sup>49</sup> PSM may adopt many conformations, some of which will be more prevalent in SAXS<sub>1</sub> and others in SAXS<sub>2</sub>. Furthermore, these states are not discrete; there will be continuous transitions between them. On average, the properties of the two monolayers of SAXS<sub>1</sub> or SAXS<sub>2</sub> will be the same, but this does not necessitate the properties of any two coupled PSM molecules being correlated. The only property to show mild transbilayer (anti)correlation is the degree of interdigitation of the sphingosine (SPH) and N-linked fatty acid (FA) tails, with interdigitation increasing in one leaflet as it decreases in the other (Figure S4). This arises from the disordered terminal methyl and methylene groups of one PSM molecule creating space that the more ordered tails

of the coupled PSM then occupy (Figure S5). It is unsurprising that interdigitation is correlated given this is the only observable we measure that is determined by interactions at the interface of the two leaflets. There is, however, generally little interdigitation between coupled PSM molecules, and when it is present the hydrocarbon tails never protrude more than 5 Å into the opposing leaflet. Nonetheless, this indicates that interdigitation is an important mechanism by which forces are transmitted across leaflets, even when the extent of interdigitation is small. This provides a second account, after Nagle et al.,<sup>50</sup> of how mini-interdigitation may occur in lipid membranes of relatively short-chained phospholipids.

The difference in bilayer thickness, measured as the distance between the peaks of relative electron densities in SAXS<sub>1</sub> and SAXS<sub>2</sub>, is consistent with those of the bilayers identified by simulation, B<sub>1</sub> and B<sub>2</sub> (Table 2). However, these latter bilayers

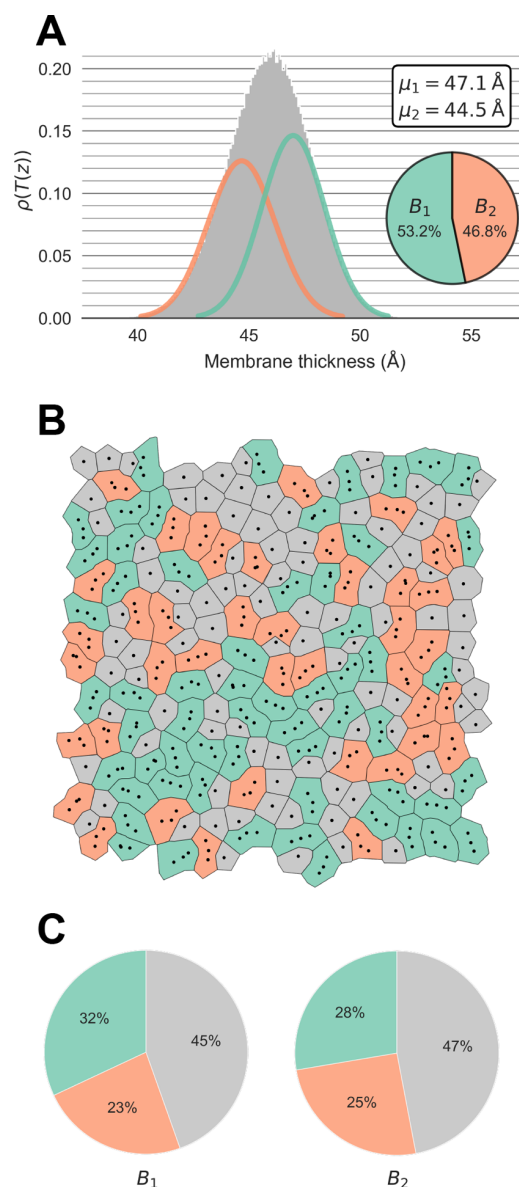
**Table 2. Structural Parameters of PSM–Cholesterol Bilayers Derived from X-ray Diffraction (SAXS<sub>1</sub> and SAXS<sub>2</sub>; WAXS<sub>1</sub> and WAXS<sub>2</sub>) and Simulation (B<sub>1</sub> and B<sub>2</sub>)<sup>a</sup>**

	<i>d</i>	<i>d<sub>w</sub></i>	<i>d<sub>C2S</sub></i>	<i>d<sub>p</sub></i>	<i>d<sub>N</sub></i>	relative mass
SAXS <sub>1</sub>	64.1	17.9		46.2		31.7
SAXS <sub>2</sub>	62.1	17.3		44.8		68.3
B <sub>1</sub>			40.3	47.1	49.5	53.2
B <sub>2</sub>			38.5	44.5	47.6	46.8
WAXS <sub>1</sub>	5.0					51.0
WAXS <sub>2</sub>	4.7					49.0

<sup>a</sup>Values (in angstroms) for the *d* spacings (*d*), water layer thickness (*d<sub>w</sub>*), and membrane thickness (*d<sub>p</sub>*) were obtained from electron density distributions calculated from the first four orders of Bragg reflection in the SAXS region (Figure 1). Values for the mean membrane thickness for B<sub>1</sub> and B<sub>2</sub> are defined as the C2S–C2S (*d<sub>C2S</sub>*), P–P (*d<sub>p</sub>*), and N–N (*d<sub>N</sub>*) distances in *z* between coupled PSM molecules. (See Figure S1 for the atom names of PSM.)

have notably more distinct mean values compared to SAXS<sub>1</sub> and SAXS<sub>2</sub>. The difference in the distribution of mass between the unit cells identified in the SAXS diffraction region and the bilayers characterized by simulation may be a reflection of the difference in the degree of equilibration in the two systems. Nonetheless, an analysis of B<sub>1</sub> and B<sub>2</sub> provides insight into the intraleaflet interactions that characterize the distinct bilayers of SAXS<sub>1</sub> and SAXS<sub>2</sub>. It should be noted that while the difference in the thicknesses of B<sub>1</sub> and B<sub>2</sub> is on the order of 1 Å, the P–P, C2S–C2S, and N–N distributions of B<sub>1</sub> and B<sub>2</sub> are statistically distinct on the basis of their 95% confidence intervals. (See Table S1 for the confidence intervals and Figure S1 for the atom names of PSM.)

SAXS provides no information on the lateral distribution of PSM in SAXS<sub>1</sub> and SAXS<sub>2</sub>. From the MD simulations, however, an analysis of the local lipid environment shows that B<sub>1</sub> lipids have a local neighborhood enriched in PSM, specifically other B<sub>1</sub> PSM, compared to B<sub>2</sub> (Figure 2). B<sub>2</sub> lipids, conversely, have local neighborhoods enriched in cholesterol compared to B<sub>1</sub>. That PSM molecules in the thinner bilayer, B<sub>2</sub>, are more likely to be neighbors to cholesterol is contrary to the view that cholesterol has an ordering, and thus thickening, effect on the hydrocarbon tails of PSM. However, it is consistent with findings that cholesterol disrupts networks of intermolecularly hydrogen bonded sphingomyelin (SM) molecules,<sup>51</sup> fluidizing an SM-enriched gel phase.<sup>52,53</sup>



**Figure 2.** (A) Distribution of P–P distances (gray) and the decomposition of this distribution into a thick (B<sub>1</sub>, green) and a thin (B<sub>2</sub>, orange) bilayer through a Gaussian mixture model. A hidden Markov model (HMM) was constructed to determine the most likely sequence of bilayer states (B<sub>1</sub> or B<sub>2</sub>) for each PSM molecule over time. Mean peak values and relative masses (inset) are those obtained from the HMM. (B) A Voronoi diagram that illustrates the lateral distribution of B<sub>1</sub> (green), B<sub>2</sub> (orange) PSM, and cholesterol (gray) within the bilayer. (C) Local lipid composition composed of B<sub>1</sub> (green), B<sub>2</sub> (orange), and cholesterol (gray) neighbors for B<sub>1</sub> (left) and B<sub>2</sub> (right) PSM. The local lipid composition is determined by the mean number of B<sub>1</sub> PSM, B<sub>2</sub> PSM, and cholesterol molecules directly neighboring a reference B<sub>1</sub> or B<sub>2</sub> PSM in the Voronoi tessellation.

From the WAXS measurements, we find two states regarding the intraleaflet lateral distribution of hydrocarbon tails, with *d* spacings of 5.0 and 4.7 Å for WAXS<sub>1</sub> and WAXS<sub>2</sub>, respectively (Figure 1). These *d* spacings are related to the lateral distances between the center of mass of the PSM hydrocarbon tails and the center of mass of neighboring cholesterol molecules. Specifically, this distance can be calculated via  $2r = d/\sin(\theta)$ ,

where  $\theta = 45^\circ$ . One may intuitively identify these two states, WAXS<sub>1</sub> and WAXS<sub>2</sub>, with the two bilayers identified via simulation, B<sub>1</sub> and B<sub>2</sub>, as B<sub>1</sub> lipids may be more tightly packed than those of B<sub>2</sub>. However, this is not the case (Table 3).

**Table 3. Values of  $2r = d/\sin(\theta)$  (Å); Area per Diffracting Unit ( $A = \pi r^2$ , Å<sup>2</sup>); and Relative Mass, as Calculated via WAXS (WAXS<sub>1</sub> or WAXS<sub>2</sub>) and Simulation (B<sub>1</sub> or B<sub>2</sub>; Chol $\alpha$  or Chol $\beta$ )<sup>a</sup>**

	$2r$	area	relative mass
WAXS <sub>1</sub>	7.1	39.6	51.0
WAXS <sub>2</sub>	6.7	35.3	49.0
B <sub>1</sub>	7.3	41.9	52.8
B <sub>2</sub>	7.2	40.7	47.2
Chol $\alpha$	7.4	43.0	49.7
Chol $\beta$	7.1	39.6	50.3

<sup>a</sup>In the simulations, distance  $2r$  corresponds to the mean distance from the center of mass of PSM hydrocarbon tails to the center of mass of neighboring cholesterol. We assume that the two distinct mean distances arise via either (i) PSM–cholesterol interactions in the two distinct bilayers identified by simulations (B<sub>1</sub> and B<sub>2</sub>) or (ii) PSM interacting with the  $\alpha$  or  $\beta$  face of cholesterol (Chol $\alpha$  or Chol $\beta$ ). The latter assumption (ii) provides better agreement with WAXS<sub>1</sub> and WAXS<sub>2</sub>, given that there is only a 0.1 Å difference in  $2r$  between B<sub>1</sub> and B<sub>2</sub>. See Table S2 for confidence intervals.

Instead, the simulation results suggest that WAXS<sub>1</sub> and WAXS<sub>2</sub> correspond to PSM interacting with the  $\alpha$  and  $\beta$  faces of cholesterol, respectively (Figure S3). It may seem counter-intuitive that the interactions of PSM with the smooth  $\alpha$  face of cholesterol have a larger  $d$  spacing than those with the rough  $\beta$  face. However, the methyl groups protruding from the  $\beta$  face of cholesterol push its center of mass toward the tails of the PSM interacting with this face, thus decreasing the distance between their respective centers of mass.

That WAXS<sub>1</sub> and WAXS<sub>2</sub> do not correspond to the two unit cells identified via SAXS may explain why the WAXS<sub>1</sub>/WAXS<sub>2</sub> ratio is almost 1:1 even though the SAXS<sub>1</sub>/SAXS<sub>2</sub> ratio is approximately 1:2. That is, because the SAXS unit cells are distinct from the WAXS unit cells, we should not necessarily expect their respective relative masses to be equal. Alternatively, since it is not possible to relate the WAXS reflections to any particular arrangement of the PSM hydrocarbon tails associated with either SAXS<sub>1</sub> or SAXS<sub>2</sub>, a possible explanation could be that the proportion of the two packing arrangements identified in the WAXS is different in SAXS<sub>1</sub> and SAXS<sub>2</sub>. In

either scenario, SAXS<sub>1</sub> and SAXS<sub>2</sub> cannot be described purely by the face of cholesterol with which PSM interacts.

**Bilayers Identified by Simulation.** We now provide a descriptive account of the structural differences between B<sub>1</sub> and B<sub>2</sub>. There are differences of 1.8, 2.6, and 1.9 Å among the C2S–C2S, P–P, and N–N thicknesses, respectively, of B<sub>1</sub> and B<sub>2</sub> (Figure 3 and Table 2). The broader distributions of N–N thicknesses compared to those of C2S–C2S and P–P are indicative of the degree of conformational flexibility afforded to the quaternary N of the choline moiety. The SPH tails of PSM are 0.3 Å thicker (Figure S6), 0.9 Å less interdigitated (Figure S7), and 7% more ordered (Figure S8) in B<sub>1</sub> as compared to those in B<sub>2</sub>, while the total headgroup thickness shows less than a 0.1 Å difference between the leaflets (Figure S9). This demonstrates that the major difference in membrane thickness is due to the thickness of the hydrophobic core as opposed to the thickness of the headgroup region. This is further borne out by the positive correlation between the tail and membrane thickness, but there is a lack of correlation between the headgroup and membrane thickness (Figure S10). There is little difference in the mean area per lipid of B<sub>1</sub> and B<sub>2</sub> PSM molecules (Figure S11), with the thinner membrane actually having a smaller area per lipid by 0.8 Å<sup>2</sup>. The smaller area of PSM in B<sub>2</sub> corroborates our above finding that the lateral  $d$  spacing of B<sub>2</sub> is slightly smaller than that of B<sub>1</sub> (Table 3).

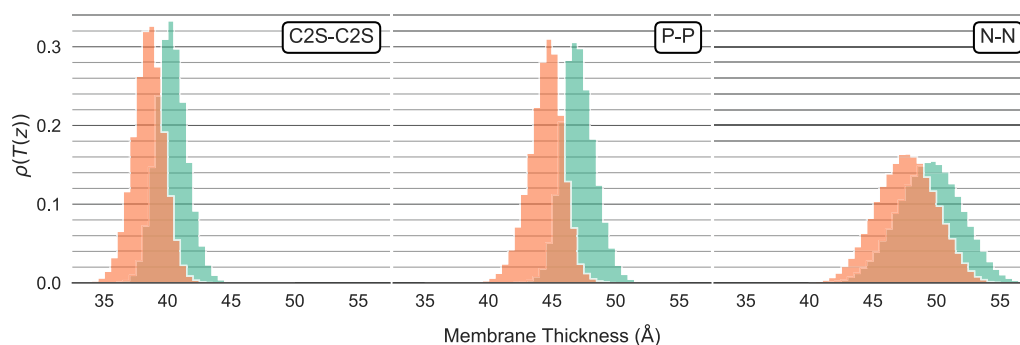
There is a 12% difference in the mean angle that the P–N vectors make with the  $z$  axis ( $\theta_{\text{pnz}}$ ) in B<sub>1</sub> and B<sub>2</sub>. We see a more significant difference of 29% in the orientation of PSM's ceramide plane ( $\theta_{\text{cpz}}$ ) in the two bilayers, with the C2S–P vector in B<sub>1</sub> aligning more with the membrane normal (Table 4).

**Table 4. Mean Values of  $\theta_{\text{cpz}}$ ,  $\theta_{\text{pnz}}$ , and  $\theta_{\text{tails}}$  for B<sub>1</sub> and B<sub>2</sub> (in Degrees)<sup>a</sup>**

structure	$\theta_{\text{cpz}}$	$\theta_{\text{pnz}}$	$\theta_{\text{tails}}$
B <sub>1</sub>	25.9	72.0	21.4
B <sub>2</sub>	36.4	65.8	22.7

<sup>a</sup> $\theta_{\text{tails}}$  is a measure of the degree of splay of the PSM tails. (See Figure S2 for a definition of these angles.)

We see a small difference in the splay of the SPH and FA tails ( $\theta_{\text{tails}}$ ) between the two bilayers, with B<sub>1</sub> lipids more likely to have less-splayed tails (Table 4). The differences in the distributions of  $\theta_{\text{pnz}}$ ,  $\theta_{\text{cpz}}$ , and  $\theta_{\text{tails}}$  and the differences in SPH thickness and interdigitation are correlated with the different hydrogen bonding propensities of PSM in B<sub>1</sub> and B<sub>2</sub>. These are



**Figure 3.** Distribution of membrane thicknesses for B<sub>1</sub> (green) and B<sub>2</sub> (orange) lipids.

important interrelated features that distinguish the two bilayers.

**Hydrogen Bonding in  $B_1$  and  $B_2$ .** The hydrogen bond donating groups of the SPH and FA tails of PSM partake in both intramolecular and intermolecular hydrogen bonding. This is an important feature that distinguishes PSM from phosphatidylcholines (PC) and other glycerophospholipids.<sup>4</sup> We found that PSM has an intramolecular hydrogen bond from its hydroxyl to phosphate group 97% of the time (Table 5). The prevalence of this hydrogen bond is similar to that

**Table 5. Percentage of PSM Molecules Partaking in Each Type of Hydrogen Bond for  $B_1$  and  $B_2$ <sup>a</sup>**

	HP	AE	AH	CE	AC	CP	CH	CP or CH	any PSM-CHOL
$B_1$	97	23	11	15	10	2	1	2	25
$B_2$	97	25	16	23	8	14	3	16	45

<sup>a</sup>PSM hydroxyl-phosphate (HP) hydrogen bonds are intramolecular. All other hydrogen bonds are intermolecular. Hydrogen bond types are defined by two letters (DA, donor/acceptor), with H corresponding to the PSM hydroxyl group; P corresponding to the PSM phosphate group; A corresponding to the PSM amide group; E corresponding to the PSM carbonyl group; C corresponding to the cholesterol hydroxyl group, and "any PSM-CHOL" representing the presence of any type of hydrogen bond between PSM and cholesterol.

reported by Venable et al.<sup>29</sup> but around twice that reported in Wang et al.<sup>51</sup> This is due to the different definitions of a hydrogen bond used in each analysis. Venable et al. did not use an angle cutoff, while Wang et al. used a stricter donor-acceptor distance (3.0 Å) and donor-hydrogen-acceptor angle (150°) cutoff than ours (3.5 Å, 120°). Our less strict definition reveals that almost all (97%) PSM molecules have their hydroxyl and phosphate moieties oriented within these limits and that interactions with cholesterol do not affect this (Table 6).

**Table 6. Percentage of PSM Partaking in Intermolecular PSM-PSM Hydrogen Bonds<sup>a</sup>**

all	Chol	not Chol	CE	not CE	AC	not AC	CP	not CP	CH	not CH
35	28	39	26	37	15	37	44	34	34	35

<sup>a</sup>all, all PSM; Chol, PSM hydrogen bonded to cholesterol; not Chol, PSM not hydrogen bonded to cholesterol; CE/not CE, AC/Not AC, CP/not CP, and CH/not CH are for PSM partaking/not partaking in a given type of hydrogen bond. Hydrogen bond types are defined in Table 5.

The idiosyncratic hydrogen bond properties of sphingomyelin distinguish both its lipid-lipid and lipid-cholesterol interactions from those of other phospholipids. Sphingomyelin bilayers are characterized by a network of intermolecular hydrogen bonds, while cholesterol may preferentially interact with sphingomyelins over other phospholipids because of their increased ability to hydrogen bond.<sup>4,29,51</sup> Below we discuss the prevalence of PSM-cholesterol hydrogen bonds in  $B_1$  and  $B_2$  as well as their effect on the intermolecular PSM hydrogen bond network and the physical properties of the two bilayers.

**Cholesterol-PSM Hydrogen Bonds.** We found that PSM in  $B_2$  is almost twice as likely to be hydrogen bonded to cholesterol compared to PSM in  $B_1$  (Table 5). Furthermore, the  $B_2$  PSM hydroxyl and phosphate oxygen atoms are more

than 3 times and 7 times as likely, respectively, to accept a hydrogen bond from cholesterol compared to those in  $B_1$ . These differences are correlated with the changes in conformation of PSM that are largely responsible for the different thicknesses of  $B_1$  and  $B_2$ .

The hydrogen bonding of cholesterol to the phosphate or hydroxyl groups of PSM is associated with a thinning of the membrane (Figure S12). This occurs both through reorienting the C2S-P vector of PSM to sit laterally along the  $xy$  plane (Figure S13) and through an increase in the interdigitation of the SPH tail (Figure S14) into the opposing leaflet.

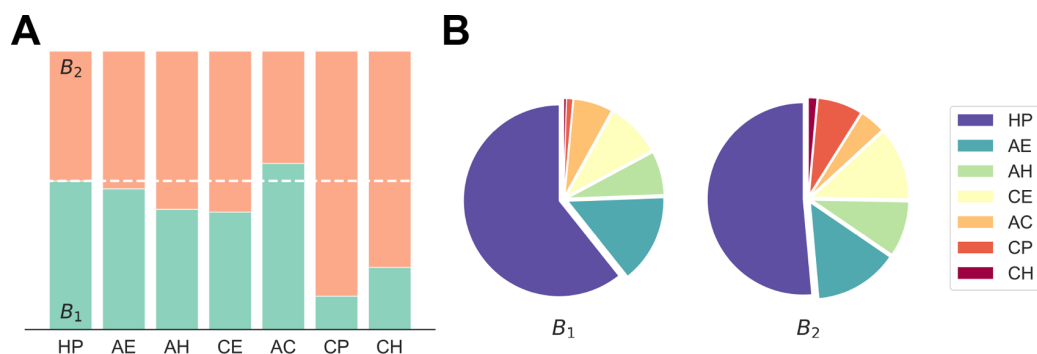
PSM-cholesterol phosphate hydrogen bonds are also correlated with cholesterol sitting higher in the membrane while PSM is pulled toward the membrane core. PSM-cholesterol phosphate hydrogen bonds also result in the P-N vector aligning with the membrane normal (Figure S13). This is possibly due to the steric hindrance caused by the C2S-P vector sitting flat on the surface, and PSM being pulled toward the membrane core when its phosphate group accepts a hydrogen bond from cholesterol.

Unlike PSM-cholesterol phosphate hydrogen bonds, PSM-cholesterol hydroxyl hydrogen bonds result in both PSM and cholesterol sitting deeper in the membrane, although to a lesser extent than does PSM as a result of PSM-cholesterol phosphate hydrogen bonds (Figure S15). Hydrogen bonds between cholesterol and the hydroxyl group of PSM are also associated with a small increase in SPH thickness and a more significant increase in SPH interdigitation, along with a substantial thinning of the FA tail (Figure S14). The differing effects on each PSM tail arise by virtue of the hydroxyl group being on the SPH tail, which becomes elongated when hydrogen bonding with cholesterol. Therefore, while both PSM-cholesterol phosphate and PSM-cholesterol hydroxyl hydrogen bonds are characteristic of  $B_2$  and both are associated with membrane thinning via PSM sitting deeper in the membrane, the latter also result in the thickening and interdigitation of the SPH tail.

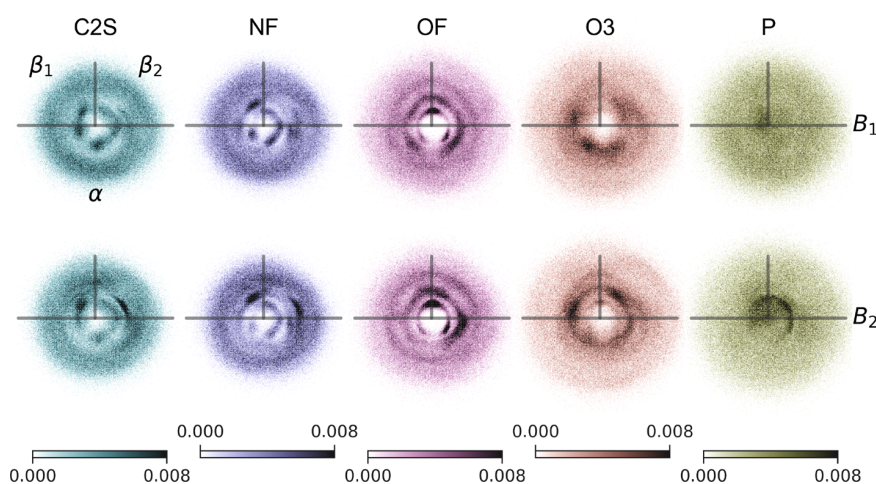
PSM amide-cholesterol hydrogen bonding, on the other hand, is associated with neither membrane thinning nor membrane thickening (Figure S12), as evidenced by their almost even distribution between  $B_1$  and  $B_2$ . PSM amide-cholesterol hydrogen bonding does, however, result in a bimodal distribution of the splay of the PSM tails ( $\theta_{\text{tails}}$ , Figure S16), with the second peak being more prominent. There is no clear angular (Figure S17) or distance (Figure S18) dependence on  $\theta_{\text{tails}}$ . Instead, the bimodal distribution is due to the tilt of the PSM tails with respect to the cholesterol backbone ( $\theta_{\text{tilt}}$  in Figure S2; Figure S19). When the PSM tails and the cholesterol ring structures are parallel to one another, the PSM tails become more splayed. In contrast, when cholesterol and PSM are oriented such that there is a larger tilt angle between them, PSM is able to reduce the splay of its tails.

PSM-cholesterol hydrogen bonding is strongly associated with  $B_2$  rather than  $B_1$ . Below we see that this is because cholesterol disrupts the intermolecular hydrogen bond network of highly ordered PSM molecules, which causes a thinning of the membrane.

**Disruption of the PSM Hydrogen Bond Network.** Sphingomyelins are known to form a network of intermolecular hydrogen bonds.<sup>4,51,53</sup> We see that PSM molecules hydrogen bonded to cholesterol have 35% fewer hydrogen bonds with other PSM molecules (Figure 4 and Table 6), indicating a disruption of this PSM-PSM hydrogen bond



**Figure 4.** (A) Percentage of lipids with a given type of hydrogen bond belonging to  $B_1$  (green) and  $B_2$  (orange). The white dotted line is at 53.2%  $B_1$ , which is the prevalence of  $B_1$  lipids in the membrane (Figure 1). (B) For each bilayer, the probability of a hydrogen bond being of a given type. Hydrogen bond types are defined in Table 5.



**Figure 5.** Distribution of PSM atoms around a neighboring cholesterol molecule for  $B_1$  (upper row) and  $B_2$  (lower row). Horizontal gray bars represent 25 Å; vertical gray bars represent 10 Å. See Figure S3 for a definition of the  $\alpha$  and  $\beta$  faces of cholesterol, and see Figure S1 for the atom names of PSM.

network. This disruption of the hydrogen bond network by the cholesterol in equimolar mixtures is supported by recent simulation<sup>51</sup> and experimental findings.<sup>52</sup> In addition, cholesterol has also been found to disrupt the intermolecular hydrogen bond network of sphingomyelin in more complex mixtures that also contain glycerolipids.<sup>53</sup>

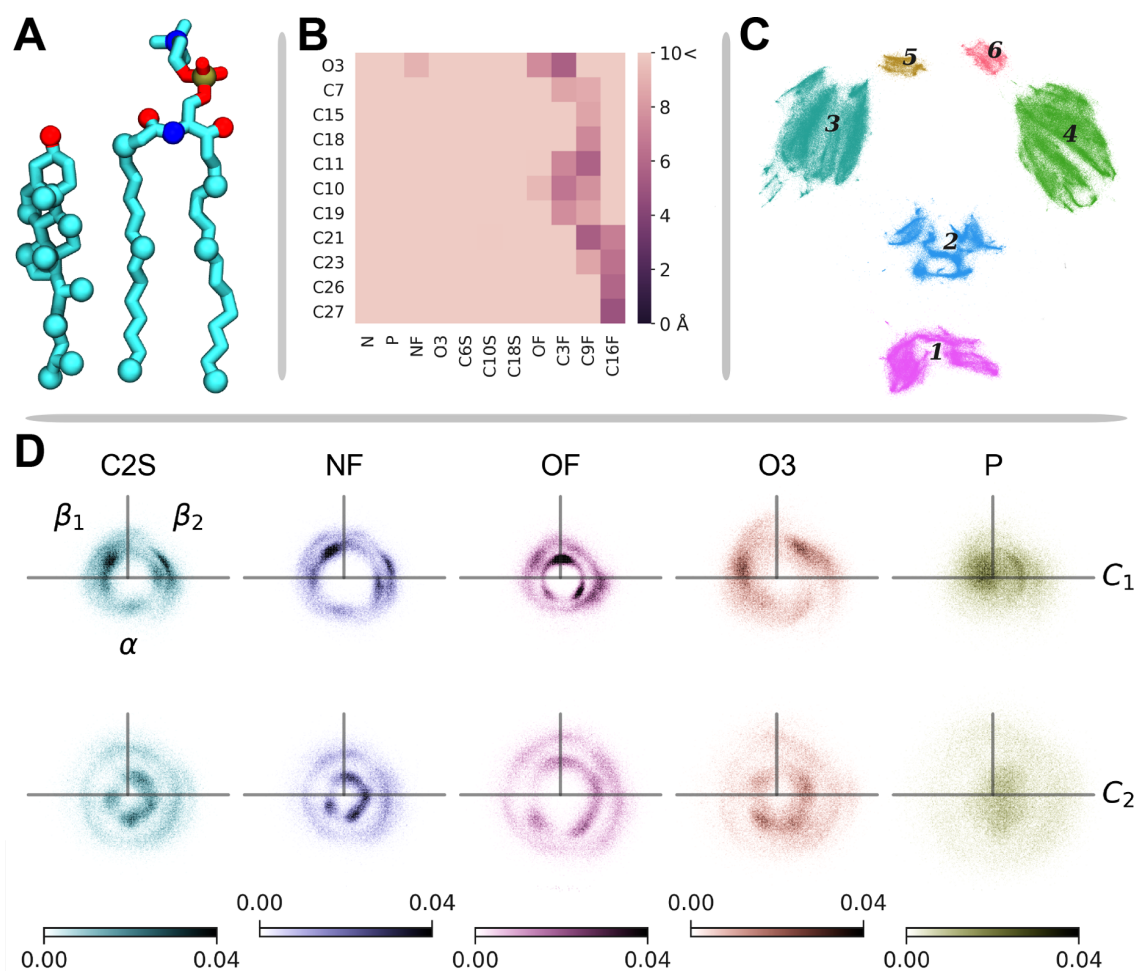
The largest disruption is seen with PSM amide–cholesterol hydrogen bonding, which results in a 59% decrease in the number of intermolecular PSM–PSM hydrogen bonds. This may be explained by the increase in  $\theta_{\text{tails}}$  brought about by PSM amide–cholesterol hydrogen bonding (Figure S16). The increased splay of the hydrocarbon tails will increase lateral pressure in the membrane and therefore push neighboring PSM away from one another. In contrast to our results, Sodt et al. found that PSM amide–cholesterol hydrogen bonds actually encourage the formation of further PSM–PSM intermolecular hydrogen bonding.<sup>39</sup> However, in the Sodt et al. simulation with a lipid composition closest to our equimolar mixture, a composition of 0.64/0.03/0.33 PSM/dioleoylphosphatidylcholine/cholesterol and a temperature of 295 K were used. The PSM–cholesterol interactions, and thus hydrogen bonding, in such mixtures and at such temperatures will be different from those in the equimolar mixture studied here.

PSM–cholesterol phosphate hydrogen bonds are unique in that they result in the formation of 29% more PSM–PSM

intermolecular hydrogen bonds (Table 6). This can be explained by PSM sitting deeper in the membrane when its phosphate group accepts a hydrogen bond from cholesterol, as described above. This in turn will desolvate the carbonyl and hydroxyl groups of PSM, thus encouraging the formation of PSM–PSM hydrogen bonds to avoid the free-energy cost associated with this dehydration.

**Lateral Distribution of PSM around Cholesterol.** The asymmetry of cholesterol is important in driving its interactions with the hydrocarbon tails of phospholipids in bilayers.<sup>46</sup> Saturated phospholipid tails preferentially interact with the smooth  $\alpha$  face of cholesterol, whereas unsaturated tails show no such preference.<sup>45–47</sup> The  $\beta$  face is characterized by the protrusion of two methyl groups from its tetracyclic ring structure. We show that  $\beta$ -face interactions induce disorder in saturated hydrocarbon tails and that the polar groups of PSM also have preferences for the face of cholesterol with which they interact. We interpret the effect of the asymmetry of cholesterol in terms of its influence on membrane thickness and hydrogen bonding with PSM.

The hydrocarbon tails of  $B_1$  and  $B_2$  lipids interact preferentially with the  $\alpha$  and  $\beta$  faces of cholesterol, respectively (Figure 5; Figure S20, C2S). Cholesterol molecules neighboring only  $B_1$  PSM are more aligned with the membrane normal compared to those neighboring only  $B_2$  PSM (Figure S21).



**Figure 6.** (A) Atoms used in the clustering of PSM–cholesterol conformations are rendered as large spheres. (B) Distance matrix (in angstroms) for the atoms of the PSM–cholesterol pair shown in panel A. Atom labels are the CHARMM atom names. Distance matrices were calculated for every PSM–cholesterol pair at 1 ns intervals, giving a total of 376 257 pairs. UMAP<sup>35</sup> was used to embed these 121D matrices into a 2D space. (C) PSM–cholesterol pairs in the 2D embedded space, clustered using HDBSCAN.<sup>36,37</sup> (D) Distribution of PSM atoms around a neighboring cholesterol molecule for clusters  $C_1$  and  $C_2$ . Horizontal gray bars represent 25 Å; vertical gray bars represent 10 Å. See Figure S3 for a definition of the  $\alpha$  and  $\beta$  faces of cholesterol, and see Figure S1 for the atom names of PSM.

This indicates that interactions of the saturated hydrocarbon tails with the  $\alpha$  face of cholesterol simultaneously order the tails and encourage cholesterol to sit aligned with the membrane normal, possibly to maximize the apolar contacts between the neighboring lipids. This has the effect of increasing the P–P distance in the bilayer and is characteristic of  $B_1$ .

For both  $B_1$  and  $B_2$ , there are concentric shells around cholesterol in which PSM is preferentially located (Figure 5: C2S, NF, and OF). Note that these are not necessarily equivalent to solvation shells of PSM around cholesterol, as we consider only PSM–cholesterol pairs that share an edge in the Voronoi tessellation of atomic positions.

The innermost shell occupied by NF atoms shows increased density on the  $\alpha$  face of cholesterol, especially in  $B_1$ . The second shell, conversely, is characterized by increased density on the  $\beta$  face of cholesterol, especially in  $B_2$ . The second shell is also less well defined than the innermost. The interaction shells of OF atoms show an inverse relationship to those of NF: the first and second shells have increased density on the  $\beta$  and  $\alpha$  faces of cholesterol, respectively. This inverse relation-

ship is explained by the stereochemistry of the N-linked fatty acid of the PSM: the NF and OF atoms of the amide group are in a planar trans configuration. We therefore suggest that the innermost shell of NF or OF atoms is due to direct interactions with cholesterol, whereas the second shell results from the configuration of the peptide bond. That is, when the NF atom of PSM interacts with cholesterol, the planar nature of the peptide bond prohibits the OF atom of the same PSM molecule from interacting with the same cholesterol molecule. Moreover, the amide NF and OF atoms prefer to interact with the cholesterol  $\alpha$  and  $\beta$  faces, respectively.

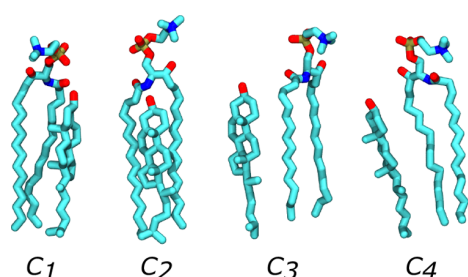
The phosphate group of  $B_2$  PSM sits preferentially on top of the center of mass of neighboring cholesterol, specifically, on top of its  $\beta_2$  face (Figure 5, P). In contrast,  $B_1$  PSM shows no such preference. We identify this increased density of PSM around the  $\beta_2$  face of cholesterol in  $B_2$  with PSM–cholesterol phosphate hydrogen bonds, which are significantly more prevalent in  $B_2$ .

**Sphingomyelin–Cholesterol Conformations.** Through the above consideration of the lateral distribution of PSM around cholesterol, we see the effect of the asymmetry of



cholesterol on membrane thickness and PSM hydrogen bonding. However, by considering instead conformations formed by PSM–cholesterol pairs, we can probe the effect of the mutual orientation of cholesterol and PSM on membrane properties. This reveals that the asymmetry of PSM, arising from the protrusion of its carbonyl and hydroxyl groups on its  $\beta$  face (Figure S3), is as important as the asymmetry of cholesterol in driving their interaction.

Through a clustering of the interatomic distances between representative atoms of interacting PSM and cholesterol molecules, we have identified six modes of interaction between them (Figure 6). The first two modes,  $C_1$  and  $C_2$ , are distinguished from the others by virtue of cholesterol interacting with both the SPH and FA tails of PSM simultaneously (Figure 7).  $C_1$  is characterized by a cholesterol



**Figure 7.** Representative conformations from clusters  $C_1$  to  $C_4$ .  $C_1$  is characterized by interactions between the  $\beta$  face of cholesterol and the  $\beta$  face of PSM.  $C_2$  is characterized by interactions between the  $\alpha$  face of cholesterol and the  $\alpha$  face of PSM. Cholesterol in  $C_3$  interacts only with the N-linked fatty acid of PSM. Cholesterol in  $C_4$  interacts only with the sphingosine base of PSM.

interacting only with the  $\beta$  face of PSM (Figure S22), primarily via its own  $\beta$  face (Figure S23). Conversely,  $C_2$  is characterized by cholesterol interacting only with the  $\alpha$  face of PSM (Figure S22), primarily via its own  $\alpha$  face (Figure S23).

$C_3$  and  $C_4$  are characterized by PSM interacting with cholesterol via its FA and its SPH tail, respectively. Neither  $C_3$  nor  $C_4$  shows any preference for either face of cholesterol.  $C_5$  and  $C_6$  show the same pattern of interaction as  $C_3$  and  $C_4$ , respectively, except that the tails of PSM are at least 7 Å from the center of mass of the tetracyclic core of cholesterol. This relatively large distance between PSM and cholesterol leads to a substantial increase in the area (Figure S24) and the hydration (Figure 8) of the cholesterol hydroxyl headgroup

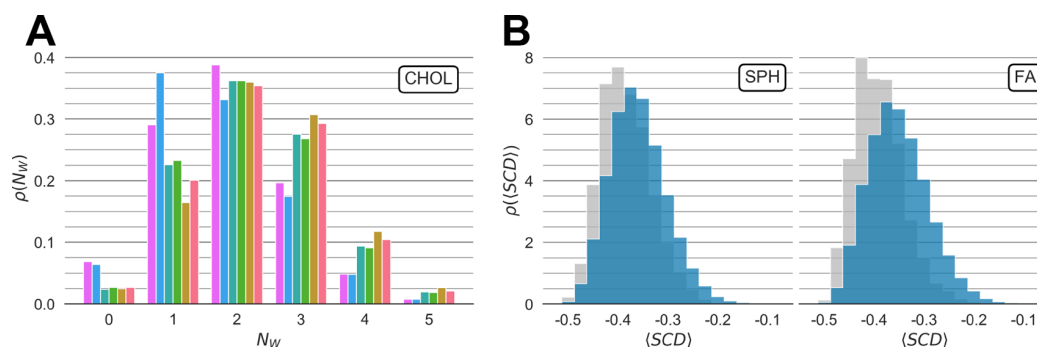
and tetracyclic ring structures. Conformations  $C_5$  and  $C_6$  will therefore incur a large free-energy cost, and as such are observed only 2.6 and 2.4% of the time, respectively.

**Structural and Interfacial Properties of PSM–Cholesterol Conformations.** PSM does not show any difference in physical properties across the six conformations. This may be because each PSM molecule in the bilayer will likely belong to more than one cluster if it is a neighbor of more than one cholesterol molecule. We do, however, see noticeable differences between PSM molecules that neighbor cholesterol and those that neighbor only other PSM molecules. In particular, PSM that neighbors cholesterol is thinner (Figure S25) and has more disordered SPH and FA tails (Figure 8). These observations are in agreement with the local lipid environment of  $B_2$ , the thinner bilayer, being enriched in cholesterol (Figure 2).

Cholesterol molecules belonging to  $C_1$  or  $C_2$  are desolvated (Figure 8) and have a smaller area per lipid (Figure S24) in comparison to those not in either conformation. Molecules belonging to  $C_2$  have a greater effect on both the area and hydration of cholesterol.

In  $C_1$ , the desolvation of cholesterol is due to the phosphate group of PSM sitting on top of cholesterol's hydroxyl group (Figure S22), which we found to be a prominent feature in  $B_2$ . Indeed, more than 80% of PSM–cholesterol phosphate hydrogen bonds occur in either  $C_1$  or  $C_4$  (Figure S26). These two conformations are characterized by the interactions of cholesterol with the  $\beta$  face and the FA tail of PSM, respectively. Unlike  $C_1$ , belonging to  $C_4$  is not correlated with cholesterol desolvation (Figure 8). This is likely because the P atom in the phosphate group does not sit preferentially above cholesterol in the  $C_4$  conformation (Figure S23). We therefore envision a scenario whereby the  $\beta$  face of cholesterol interacts with the  $\beta$  face of PSM. This disorders the PSM tails and allows the C2S–P vector to sit laterally along the membrane surface. In turn, the P atom in the phosphate group sits directly on top of the cholesterol molecule and encourages the formation of a hydrogen bond from cholesterol to the phosphate oxygen atoms of PSM (Figure 7,  $C_1$ ).

The PSM–cholesterol phosphate hydrogen bonds in  $C_4$  most likely arise through these polar groups being brought into close proximity by a combination of intramolecular and intermolecular hydrogen bonds. All PSM–cholesterol hydroxyl hydrogen bonds occur in  $C_4$  (Figure S26), in which the cholesterol interacts only with the N-linked fatty acid of PSM.



**Figure 8.** (A) Probability of the number of water molecules ( $N_w$ ) in the first hydration shell of cholesterol's hydroxyl group for clusters  $C_1$  to  $C_6$ . Colors correspond to those in Figure 6. (B) Distribution of the mean deuterium order parameter of the SPH and FA tails for PSM neighboring cholesterol (blue) and PSM having no cholesterol neighbors (gray).

At the same time, over 96% of PSMs have an intramolecular hydrogen bond between their hydroxyl and phosphate groups. The result is that the phosphate group of PSM is close enough to the hydroxyl group of a neighboring cholesterol to allow the formation of a hydrogen bond between the two.

In  $C_2$ , the desolvation of cholesterol is due to the sterol sitting almost directly underneath the C2S atom of PSM (Figure 6 and Figure S27). This is enabled through the mutual orientation of PSM and cholesterol, with the  $\alpha$  face– $\alpha$  face interactions that characterize  $C_2$  allowing cholesterol to sit underneath the C2S atom of PSM without disrupting the order of the latter's hydrocarbon tails. These interactions are stabilized by PSM amide–cholesterol hydrogen bonds, 100% of which occur in  $C_2$  (Figure S26).

## DISCUSSION AND CONCLUSIONS

From small- and wide-angle X-ray scattering measurements, we have identified the presence of two distinct bilayer structures in equimolar mixtures of PSM and cholesterol at biologically relevant temperatures (37 °C). We subsequently performed all-atom molecular dynamics simulations to understand the microscopic origin of these unit cells in terms of the distinct lateral interactions between cholesterol and PSM. Specifically, we constructed a hidden Markov model (HMM), based on phosphate–phosphate distances, to identify two coexisting bilayers in the simulation. We found that the bilayers identified via simulation,  $B_1$  and  $B_2$ , correspond very well to those identified via SAXS, SAXS<sub>1</sub>, and SAXS<sub>2</sub>. These bilayers, however, cannot be identified with those observed via WAXS. Our simulation results suggest that the WAXS unit cells, WAXS<sub>1</sub> and WAXS<sub>2</sub>, arise from PSM interacting with the  $\alpha$  and  $\beta$  faces of cholesterol, respectively. This therefore leads us to conclude that the unit cells identified via SAXS are distinct from those identified via WAXS.

The thicker bilayer,  $B_1$ , is characterized by PSM molecules with more extended, more ordered, less interdigitated hydrocarbon tails than those of  $B_2$ . We also observe significantly fewer PSM–cholesterol hydrogen bonds but significantly more intermolecular PSM–PSM hydrogen bonds in  $B_1$ . The result is the disruption of the network of intermolecular hydrogen bonds for  $B_2$  PSM, which has been observed in previous simulations<sup>51</sup> and experiments.<sup>52,53</sup> In contrast, Sodt et al.<sup>39</sup> found that hydrogen bonding between cholesterol and sphingomyelin actually encourages further intermolecular SM–SM hydrogen bonding. We suggest that this discrepancy may be due to our examination of an equimolar mixture at physiological temperature (310 K) whereas the simulations performed by Sodt et al. were at 295 K. Temperature is a critical factor in the phase separation of sphingomyelin, cholesterol, and phosphatidylcholine ternary mixtures, as are the proportions of sphingomyelin and cholesterol.<sup>19</sup> We refer the reader to Wang and Klauda<sup>51</sup> for further discussion on the biological significance of this discrepancy.

From the HMM, we found that the face of cholesterol with which PSM interacts impacts the physical properties of PSM, as has been reported elsewhere.<sup>45–47</sup> On the other hand, from a clustering of the interatomic distances of PSM–cholesterol pairs, we found that the face of PSM with which cholesterol interacts impacts the physical properties of cholesterol. Furthermore, we found that the mutual orientation of interacting cholesterol and PSM molecules tends to adopt one of two forms: smooth face–smooth face ( $\alpha$ – $\alpha$ ) contacts

or rough face–rough face ( $\beta$ – $\beta$ ) contacts. The mutual orientation therefore affects the physical properties of both molecules and entirely determines which types of intermolecular hydrogen bonds they are able to form. To the best of our knowledge, this is the first time the asymmetry of PSM, or of any phospholipid, has been shown to be important in driving its interactions with cholesterol.

In the clustering of PSM–cholesterol distance matrices, we consider only pairwise interactions between a single PSM molecule and a single cholesterol molecule. However, the properties of any given cholesterol or PSM molecule will depend on the cooperative activity of multiple cholesterol and PSM neighbors as well as the solvent at the interface. Nonetheless, we have shown that PSM–cholesterol interactions can be broadly categorized into four modes and that PSM is able to shield cholesterol from the surrounding solvent via two mechanisms: (1) the interactions between the  $\beta$  face of cholesterol and the  $\beta$  face of PSM encourage the formation of PSM–cholesterol phosphate hydrogen bonds with a PSM phosphate group sitting directly on top of cholesterol and (2) the interactions between the  $\alpha$  face of cholesterol and the  $\alpha$  face of PSM encourage the formation of PSM amide–cholesterol hydrogen bonds, with PSM straddling the sterol.

The shielding of cholesterol by phospholipid headgroups is a central tenet of the umbrella model of phospholipid–cholesterol mixing. Our findings, however, are contrary to the umbrella model in its original formulation.<sup>21</sup> Huang et al.<sup>21</sup> proposed that cholesterol mixes with phospholipids in order to be shielded by large lipid headgroups, thus avoiding the free-energy cost of its hydrophobic core being exposed to the surrounding solvent. The authors emphasized that favorable cholesterol–phospholipid interactions are not responsible for their mixing behavior. We see, however, that PSM phosphate–cholesterol and PSM amide–cholesterol hydrogen bonding drives interactions between SM and cholesterol molecules and that such hydrogen bonding is an effective means of desolvating cholesterol.

This shielding of cholesterol, however, comes with the cost of disrupting the PSM–PSM hydrogen bond network and thinning the membrane. Thus there is an interplay between maintaining this hydrogen bond network and reducing the free-energy cost of exposing cholesterol molecules to the solvent. This results in the formation of highly ordered, PSM-enriched regions of the membrane surrounded by a mixture of cholesterol and PSM. This is similar to the SM-enriched gel-like nanodomains found both in equimolar cholesterol–SM mixtures<sup>52</sup> and in stearyl–SM mixtures with cholesterol and phosphatidylcholine.<sup>53</sup>

All phospholipids may shield cholesterol from the surrounding solvent via the formation of hydrogen bonds between their phosphate groups and the cholesterol headgroup. However, only sphingomyelin has an N-linked fatty acid tail, which also allows this phospholipid to form hydrogen bonds between the amine group and cholesterol molecules. This latter mechanism of desolvating cholesterol, which we have found to be the most effective, will thus be unavailable to glycerophospholipids. This therefore explains why cholesterol preferentially partitions into SM-rich regions in mixtures of SM, glycerophospholipids, and cholesterol.<sup>54–56</sup> Furthermore, this mechanism provides insight into why cholesterol molecules have been observed to sit deeper in SM membranes as compared to in PC membranes.<sup>4,57,58</sup> It has previously been suggested that cholesterol sits deeper in SM membranes

because it must sit beneath the hydrogen bond network of SM.<sup>4</sup> However, we see that it is instead the preferential interaction between cholesterol and the amide group of PSM that results in the lipid straddling the sterol.

Cell membranes have an asymmetric distribution of lipids, which is necessary for proper cell functioning.<sup>59–62</sup> The plasma membrane lipids are comprised of nearly 50% cholesterol and around 25% SM. The SM is located almost entirely in the extracellular leaflet, while there is some debate around the transbilayer distribution of cholesterol. Assuming an even distribution of cholesterol across the leaflets and given the preferential mixing of cholesterol with SM, there will likely be regions of the extracellular matrix that can be approximated by the equimolar mixtures of cholesterol and SM studied here. More importantly, the intraleaflet interactions through which PSM and cholesterol form  $B_1$  and  $B_2$  may provide insight into the formation of biological nanodomains, particularly since the formation of  $L_o$  nanodomains in the extracellular leaflet, in which PSM and cholesterol colocalize, has been found to induce order in non-raft-forming lipids in the cytoplasmic leaflet.<sup>6,59,63–65</sup>

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.langmuir.0c01237>.

CHARMM atom names; angles characterizing lipids; definition of the two faces of cholesterol and PSM; tail thicknesses; tail interdigitation; tail order parameter; lipid headgroup thickness; lipid area; cholesterol tilt angle; effect of hydrogen bonds on membrane properties; 2D lateral density maps of cholesterol and PSM; and properties of lipids in each conformational cluster (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Peter J. Quinn** – Institute of Pharmaceutical Science, King's College London, London SE1 9NH, U.K.; Email: [p.quinn@kcl.ac.uk](mailto:p.quinn@kcl.ac.uk)

**Christian D. Lorenz** – Department of Physics, King's College London, London WC2R 2LS, U.K.; [orcid.org/0000-0003-1028-4804](https://orcid.org/0000-0003-1028-4804); Email: [chris.lorenz@kcl.ac.uk](mailto:chris.lorenz@kcl.ac.uk)

### Author

**Paul Smith** – Department of Physics, King's College London, London WC2R 2LS, U.K.

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.langmuir.0c01237>

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

Via our membership in the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202/1, EP/R029431/1), this work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>) and the UK Materials and Molecular Modelling Hub (MMM Hub) for computational resources, which is partially funded by the EPSRC (EP/P020194/1) to carry out the MD simulations reported in this article. P.S. acknowledges the funding provided

by the EPSRC DTP Studentship Block Grant (EP/NS09498/1). C.D.L. acknowledges the supportive research environment of the EPSRC Center for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, no. EP/L015854/1).

## ■ REFERENCES

- (1) Das, A.; Goldstein, J. L.; Anderson, D. D.; Brown, M. S.; Radhakrishnan, A. Use of mutant 125I-Perfringolysin O to probe transport and organization of cholesterol in membranes of animal cells. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 10580–10585.
- (2) Funato, K.; Riezman, H.; Muñoz, M. Vesicular and non-vesicular lipid export from the ER to the secretory pathway. *Biochim. Biophys. Acta, Mol. Cell Biol. Lipids* **2020**, *1865*, 158453.
- (3) Infante, R. E.; Radhakrishnan, A. Continuous transport of a small fraction of plasma membrane cholesterol to endoplasmic reticulum regulates total cellular cholesterol. *eLife* **2017**, *6*, No. e25466.
- (4) Slotte, J. P. The importance of hydrogen bonding in sphingomyelin's membrane interactions with co-lipids. *Biochim. Biophys. Acta, Biomembr.* **2016**, *1858*, 304–310.
- (5) Murate, M.; Kobayashi, T. Revisiting transbilayer distribution of lipids in the plasma membrane. *Chem. Phys. Lipids* **2016**, *194*, 58–71.
- (6) Cheng, X.; Smith, J. C. Biological Membrane Organization and Cellular Signaling. *Chem. Rev.* **2019**, *119*, 5849–5880.
- (7) Levental, I.; Levental, K. R.; Heberle, F. A. Lipid Rafts: Controversies Resolved, Mysteries Remain. *Trends Cell Biol.* **2020**, *30*, 341.
- (8) Nickels, J. D.; Chatterjee, S.; Stanley, C. B.; Qian, S.; Cheng, X.; Myles, D. A. A.; Standaert, R. F.; Elkins, J. G.; Katsaras, J. The in vivo structure of biological membranes and evidence for lipid domains. *PLoS Biol.* **2017**, *15*, No. e2002214.
- (9) Stottrup, B. L.; TigreLazo, J.; Bagonza, V. B.; Kunz, J. C.; Zasadzinski, J. A. Comparison of Line Tension Measurement Methods for Lipid Monolayers at Liquid-Liquid Coexistence. *Langmuir* **2019**, *35*, 16053–16061.
- (10) Gallová, J.; Uhríková, D.; Kučerka, N.; Svorková, M.; Funari, S. S.; Murugova, T. N.; Almásy, L.; Mazúr, M.; Balgavý, P. Influence of cholesterol and  $\beta$ -sitosterol on the structure of EYPC bilayers. *J. Membr. Biol.* **2011**, *243*, 1–13.
- (11) Heberle, F. A.; Petruziello, R. S.; Pan, J.; Drazba, P.; Kučerka, N.; Standaert, R. F.; Feigenson, G. W.; Katsaras, J. Bilayer Thickness Mismatch Controls Domain Size in Model Membranes. *J. Am. Chem. Soc.* **2013**, *135*, 6853–6859.
- (12) Nickels, J. D.; Cheng, X.; Mostofian, B.; Stanley, C.; Lindner, B.; Heberle, F. A.; Perticaroli, S.; Feygenon, M.; Egami, T.; Standaert, R. F.; Smith, J. C.; Myles, D. A. A.; Ohl, M.; Katsaras, J. Mechanical Properties of Nanoscopic Lipid Domains. *J. Am. Chem. Soc.* **2015**, *137*, 15772–15780.
- (13) Belička, M.; Weitzer, A.; Pabst, G. High-resolution structure of coexisting nanoscopic and microscopic lipid domains. *Soft Matter* **2017**, *13*, 1823–1833.
- (14) Allender, D. W.; Sodt, A. J.; Schick, M. Cholesterol-Dependent Bending Energy Is Important in Cholesterol Distribution of the Plasma Membrane. *Biophys. J.* **2019**, *116*, 2356–2366.
- (15) Pandit, S. A.; Vasudevan, S.; Chiu, S. W.; Mashl, R. J.; Jakobsson, E.; Scott, H. L. Sphingomyelin-cholesterol domains in phospholipid membranes: Atomistic simulation. *Biophys. J.* **2004**, *87*, 1092–1100.
- (16) Regan, D.; Williams, J.; Borri, P.; Langbein, W. Lipid Bilayer Thickness Measured by Quantitative DIC Reveals Phase Transitions and Effects of Substrate Hydrophilicity. *Langmuir* **2019**, *35*, 13805–13814.
- (17) Munro, S. An investigation of the role of transmembrane domains in Golgi protein retention. *EMBO J.* **1995**, *14*, 4695–4704.
- (18) Khelashvili, G. A.; Scott, H. L. Combined Monte Carlo and molecular dynamics simulation of hydrated 18:0 sphingomyelin-cholesterol lipid bilayers. *J. Chem. Phys.* **2004**, *120*, 9841–9847.

- (19) Quinn, P. J.; Wolf, C. Egg-Sphingomyelin and Cholesterol Form a Stoichiometric Molecular Complex in Bilayers of Egg-Phosphatidylcholine. *J. Phys. Chem. B* **2010**, *114*, 15536–15545.
- (20) Quinn, P. J. Structure of Sphingomyelin Bilayers and Complexes with Cholesterol Forming Membrane Rafts. *Langmuir* **2013**, *29*, 9447–9456.
- (21) Huang, J.; Feigenson, G. W. A microscopic interaction model of maximum solubility of cholesterol in lipid bilayers. *Biophys. J.* **1999**, *76*, 2142–2157.
- (22) Towns-Andrews, E.; Berry, A.; Bordas, J.; Mant, G. R.; Murray, P. K.; Roberts, K.; Sumner, I.; Worgan, J. S.; Lewis, R.; Gabriel, A. Time-resolved x-ray diffraction station: X-ray optics, detectors, and data acquisition. *Rev. Sci. Instrum.* **1989**, *60*, 2346–2349.
- (23) Quinn, P. J. Sphingolipid symmetry governs membrane lipid raft structure. *Biochim. Biophys. Acta, Biomembr.* **2014**, *1838*, 1922–1930.
- (24) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859.
- (25) 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.; Klauda, J. B.; Im, W. CHARMM-GUI membrane builder toward realistic biological membrane simulations. *J. Comput. Chem.* **2014**, *35*, 1997.
- (26) Lee, J. CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405.
- (27) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. GROMACS: Fast, flexible, and free. *J. Comput. Chem.* **2005**, *26*, 1701.
- (28) 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.
- (29) Venable, R. M.; Sodt, A. J.; Rogaski, B.; Rui, H.; Hatcher, E.; MacKerell, A. D., Jr.; W, R.; Pastor; Klauda, J. B. CHARMM All-Atom Additive Force Field for Sphingomyelin: Elucidation of Hydrogen Bonding and of Positive Curvature. *Biophys. J.* **2014**, *107*, 134–135.
- (30) 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. *J. Phys. Chem. B* **2010**, *114*, 7830.
- (31) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (32) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (33) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotsen, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proceedings of the 15th Python in Science Conference*, 2016; pp 98–105.
- (34) Pedregosa, F.; et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **2011**, *12*, 2825–2830.
- (35) McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2018**.
- (36) McInnes, L.; Healy, J.; Astels, S. hdbSCAN: Hierarchical density based clustering. *Journal of Open Source Software* **2017**, *2*, 205.
- (37) Malzer, C.; Baum, M. A Hybrid Approach To Hierarchical Density-based Cluster Selection. *arXiv* **2019**.
- (38) Virtanen, P.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261272.
- (39) Sodt, A. J.; Pastor, R. W.; Lyman, E. Hexagonal Substructure and Hydrogen Bonding in Liquid-Ordered Phases Containing Palmitoyl Sphingomyelin. *Biophys. J.* **2015**, *109*, 948–955.
- (40) Gu, R.-X.; Baoukina, S.; Tieleman, D. P. Phase Separation in Atomistic Simulations of Model Membranes. *J. Am. Chem. Soc.* **2020**, *142*, 2844–2856.
- (41) Park, S.; Im, W. Analysis of Lipid Order States and Domains in Lipid Bilayer Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 688–697.
- (42) Grossfield, A.; Patrone, P. N.; Roe, D. R.; Schultz, A. J.; Siderius, D.; Zuckerman, D. M. Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2019**, *1*. DOI: 10.33011/livecoms.1.1.5067.
- (43) Rycroft, C. H. VORO++: a three-dimensional voronoi cell library in c++. *Chaos* **2009**, *19*, 041111.
- (44) Smith, P.; Ziolk, R. M.; Gazzarini, E.; Owen, D. M.; Lorenz, C. D. On the interaction of hyaluronic acid with synovial fluid lipid membranes. *Phys. Chem. Chem. Phys.* **2019**, *21*, 9845–9857.
- (45) Zhang, Z.; Bhide, S. Y.; Berkowitz, M. L. Molecular dynamics simulations of bilayers containing mixtures of sphingomyelin with cholesterol and phosphatidylcholine with cholesterol. *J. Phys. Chem. B* **2007**, *111*, 12888–12897.
- (46) Martínez-Seara, H.; Róg, T.; Karttunen, M.; Vattulainen, I.; Reigada, R. Cholesterol Induces Specific Spatial and Orientational Order in Cholesterol/Phospholipid Membranes. *PLoS One* **2010**, *5*, No. e11162.
- (47) Leeb, F.; Maibaum, L. Spatially Resolving the Condensing Effect of Cholesterol in Lipid Bilayers. *Biophys. J.* **2018**, *115*, 2179–2188.
- (48) McInnes, L.; Healy, J.; Melville, J. Using UMAP for Clustering. <https://umap-learn.readthedocs.io/en/latest/clustering.html>.
- (49) Buslaev, P.; Gordeliy, V.; Grudinin, S.; Gushchin, I. Principal Component Analysis of Lipid Molecule Conformational Changes in Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1019–1028.
- (50) Nagle, J. F.; Cognet, P.; Dupuy, F. G.; Tristram-Nagle, S. Structure of gel phase DPPC determined by X-ray diffraction. *Chem. Phys. Lipids* **2019**, *218*, 168–177.
- (51) Wang, E.; Klauda, J. B. Examination of Mixtures Containing Sphingomyelin and Cholesterol by Molecular Dynamics Simulations. *J. Phys. Chem. B* **2017**, *121*, 4833–4844.
- (52) Yasuda, T.; Matsumori, N.; Tsuchikawa, H.; Lönnfors, M.; Nyholm, T. K.; Slotte, J. P.; Murata, M. Formation of Gel-like Nanodomains in Cholesterol-Containing Sphingomyelin or Phosphatidylcholine Binary Membrane As Examined by Fluorescence Lifetimes and 2H NMR Spectra. *Langmuir* **2015**, *31*, 13783–13792.
- (53) Yano, Y.; Hanashima, S.; Yasuda, T.; Tsuchikawa, H.; Matsumori, N.; Kinoshita, M.; Al Sazzad, M. A.; Slotte, J. P.; Murata, M. Sphingomyelin Stereoisomers Reveal That Homophilic Interactions Cause Nanodomain Formation. *Biophys. J.* **2018**, *115*, 1530–1540.
- (54) Quinn, P. J.; Wolf, C. An X-ray diffraction study of model membrane raft structures. *FEBS J.* **2010**, *277*, 4685–4698.
- (55) Lönnfors, M.; Doux, J. P.; Killian, J. A.; Nyholm, T. K.; Slotte, J. P. Sterols have higher affinity for sphingomyelin than for phosphatidylcholine bilayers even at equal Acyl-chain order. *Biophys. J.* **2011**, *100*, 2633–2641.
- (56) Nyholm, T. K.; Engberg, O.; Hautala, V.; Tsuchikawa, H.; Lin, K. L.; Murata, M.; Slotte, J. P. Impact of Acyl Chain Mismatch on the Formation and Properties of Sphingomyelin-Cholesterol Domains. *Biophys. J.* **2019**, *117*, 1577–1588.
- (57) Róg, T.; Pasenkiewicz-Gierula, M. Cholesterol-sphingomyelin interactions: A molecular dynamics simulation study. *Biophys. J.* **2006**, *91*, 3756–3767.
- (58) Hanashima, S.; Murakami, K.; Yura, M.; Yano, Y.; Umegawa, Y.; Tsuchikawa, H.; Matsumori, N.; Seo, S.; Shinoda, W.; Murata, M. Cholesterol-Induced Conformational Change in the Sphingomyelin Headgroup. *Biophys. J.* **2019**, *117*, 307–318.
- (59) Marquardt, D.; Geier, B.; Pabst, G. Asymmetric Lipid Membranes: Towards More Realistic Model Systems. *Membranes* **2015**, *5*, 180–196.

(60) Fujimoto, T.; Parmryd, I. Interleaflet coupling, pinning, and leaflet asymmetry-major players in plasma membrane nanodomain formation. *Front. Cell Dev. Biol.* **2017**, *4*, 155.

(61) Rivel, T.; Ramseyer, C.; Yesylevskyy, S. The asymmetry of plasma membranes and their cholesterol content influence the uptake of cisplatin. *Sci. Rep.* **2019**, *9*, 1–14.

(62) Lorent, J.; Ganesan, L.; Rivera-Longworth, G.; Sezgin, E.; Levental, K.; Lyman, E.; Levental, I. The Molecular and Structural Asymmetry of the Plasma Membrane. *bioRxiv* **2019**, 698837.

(63) Nickels, J. D.; Smith, J. C.; Cheng, X. Lateral organization, bilayer asymmetry, and inter-leaflet coupling of biological membranes. *Chem. Phys. Lipids* **2015**, *192*, 87–99.

(64) Thallmair, S.; Ingólfsson, H. I.; Marrink, S. J.; Ingoifsson, H. I.; Marrink, S. J. Cholesterol Flip-Flop Impacts Domain Registration in Plasma Membrane Models. *J. Phys. Chem. Lett.* **2018**, *9*, 5527–5533.

(65) Enkavi, G.; Javanainen, M.; Kulig, W.; Róg, T.; Vattulainen, I. Multiscale Simulations of Biological Membranes: The Challenge To Understand Biological Phenomena in a Living Substance. *Chem. Rev.* **2019**, *119*, 5607–5774.

## Chapter 5

# LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations

This Chapter was published in the Journal of Chemical Theory and Computation in 2021 and is reproduced here with permission from Smith, P; Lorenz, C.D., 'LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations', *J. Chem. Theory Comput.*, DOI: 10.1021/acs.jctc.1c00447. Copyright American Chemical Society 2021.

**Summary of the work** Whilst there exist various tools for the analysis of MD simulations of lipid membranes,<sup>193,256-260</sup> there remain some common but non-trivial analyses that can only be performed with user-written scripts. This is problematic for two reasons: i) many researchers or research groups are expending time and effort to write similar scripts and ii) in-house and user-written scripts for scientific analysis tend not to be tested. This first point means that scientific progress is slowed due to this reproduction of effort. The second point is more problematic - even with proper testing, there will be bugs in any software; without testing these bugs will likely go unnoticed and may lead to erroneous results. This is not the fault of the researcher - it is not feasible to fully test every script. However, this does highlight the usefulness of incorporating common analysis tools into a stable and fully-tested framework.

In this chapter, I present LiPyphilic - a fast, fully-tested, and easy to install Python

package for the analysis of lipid membrane simulations. LiPyphilic can perform analyses that are commonly employed yet are unavailable in other packages. It is built on top of MDAnalysis — a widely-used Python package for the analysis of MD simulations — and so learning to use LiPyphilic is straightforward for those that already use MDAnalysis. LiPyphilic was created following current best practices in software engineering, it is fully documented and interactive tutorials are available online.

**Author contributions** I created LiPyphilic, performed the benchmark simulations, created all figures, and wrote the initial draft of the manuscript. All authors contributed to and approved the manuscript in its final form.

The Supporting Information for the article is available in Appendix B.

# LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations

Paul Smith\* and Christian D. Lorenz\*

Cite This: <https://doi.org/10.1021/acs.jctc.1c00447>

Read Online

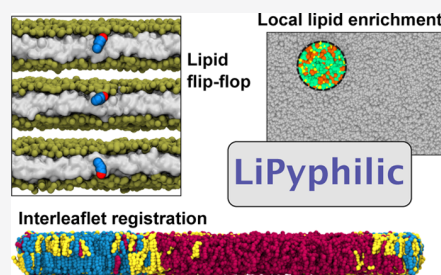
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Molecular dynamics simulations are now widely used to study emergent phenomena in lipid membranes with complex compositions. Here, we present LiPyphilic—a fast, fully tested, and easy-to-install Python package for analyzing such simulations. Analysis tools in LiPyphilic include the identification of cholesterol flip-flop events, the classification of local lipid environments, and the degree of interleaflet registration. LiPyphilic is both force field- and resolution-agnostic, and by using the powerful atom selection language of MDAnalysis, it can handle membranes with highly complex compositions. LiPyphilic also offers two on-the-fly trajectory transformations to (i) fix membranes split across periodic boundaries and (ii) perform nojump coordinate unwrapping. Our implementation of nojump unwrapping accounts for fluctuations in the box volume under the *NPT* ensemble—an issue that most current implementations have overlooked. The full documentation of LiPyphilic, including installation instructions and links to interactive online tutorials, is available at <https://lipyphilic.readthedocs.io/en/latest>.



## 1. INTRODUCTION

The plasma membrane was once thought to be a passive divider between a cell and its external environment. We now understand that it is in fact a dynamic interface upon which many cellular processes, from cell-signaling to membrane transport, depend.<sup>1,2</sup> These processes are emergent phenomena that arise from a complex interplay between the molecular species that comprise the plasma membrane. As such, there is a great interest in understanding how the lipids, proteins, and carbohydrates of the plasma membrane interact with one another.

Molecular dynamics (MD) simulations are routinely used to study lipid–lipid and lipid–protein interactions at a molecular level, and there exist many excellent tools for analyzing the trajectories of such simulations. Both FATSliM<sup>3</sup> and MemSurfer,<sup>4</sup> for example, specialize in the analysis of non-planar membranes such as buckled bilayers or vesicles. PyLipID<sup>5</sup> and ProLint<sup>6</sup> are designed for the easy and efficient analysis of lipid–protein interactions. MLLPA is a recently developed Python package that employs various machine learning algorithms to identify the phase— $L_o$  or  $L_d$ —of lipids in a bilayer.<sup>7</sup> LOOS, on the other hand, is a C++ library with a Python interface for analyzing MD simulations.<sup>8,9</sup> Unlike the above packages, LOOS handles the trajectory reading internally while also offering a large set of analysis tools, some of which are for lipid membranes. Between them, these software packages provide an extensive analysis suite for MD simulations of lipid membranes.

There are, however, some non-trivial analyses that are frequently employed but are not yet available in any analysis software we are aware of. These include the identification of cholesterol flip-flop events,<sup>10–28</sup> the classification of local lipid

environments,<sup>19,29–37</sup> and calculating the degree of interleaflet registration.<sup>13,37–47</sup> These analyses provide important information about the structure and dynamics of lipid membranes, but they currently require the writing of in-house scripts. Here, we present LiPyphilic—a fast, fully tested, and easy to install Python package that can perform these analyses, among others. See Table S1 for a comparison of tools available in LiPyphilic and other software for lipid membrane analysis.

## 2. LIPYPHILIC

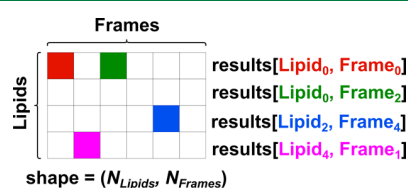
LiPyphilic is an object-oriented Python package for analyzing MD simulations of lipid membranes. It is built directly on top of MDAnalysis and makes use of NumPy<sup>48</sup> and SciPy<sup>49</sup> for efficient computation. It is force field-agnostic and can handle all-atom, united-atom, and coarse-grained systems; LiPyphilic can work with any file format that MDAnalysis can load so long as the topology contains residue names. All analysis tools in LiPyphilic inherit from the MDAnalysis base analysis class, meaning the workflow for running analysis is the same in MDAnalysis and LiPyphilic. This shared workflow makes it simple for users of MDAnalysis to learn how to use LiPyphilic.

At its core, LiPyphilic is designed to easily integrate with the wider scientific Python stack. Results are typically stored in a

Received: May 6, 2021



two-dimensional NumPy array of shape  $N_{\text{lipids}}$  by  $N_{\text{frames}}$  (Figure 1), making it simple to post-process the results for further



**Figure 1.** In LiPyphilic, analysis results are typically stored in a NumPy array of shape  $(N_{\text{lipids}}, N_{\text{frames}})$ .

analysis. Some analysis tools also take a two-dimensional NumPy array of the same shape as input. This input array may contain information about each lipid, such as which leaflet they belong to or their phase ( $L_o$  or  $L_d$ ). Alternatively, the input array may be a boolean mask—an array of True and False values specifying which lipids to include in the analysis. As these inputs are generic NumPy arrays instead of types specific to LiPyphilic, it is possible to use the output from other membrane analysis tools as input to LiPyphilic. For example, you may assign lipids to leaflets using FATSliM,<sup>3</sup> determine their phase state using MLLPA,<sup>7</sup> or calculate local membrane normals using MemSurfer,<sup>4</sup> and extract the results to perform further analysis with LiPyphilic.

The workflow for using LiPyphilic generally involves the following steps:

- 1 import MDAnalysis along with the required LiPyphilic analysis modules
- 2 load a topology and trajectory as an MDAnalysis universe
- 3 create an analysis object using the MDAnalysis universe and specifying the relevant input options
- 4 use the `run()` method to perform the analysis
- 5 store the results either by serializing the analysis object itself or by saving the results data as a NumPy array

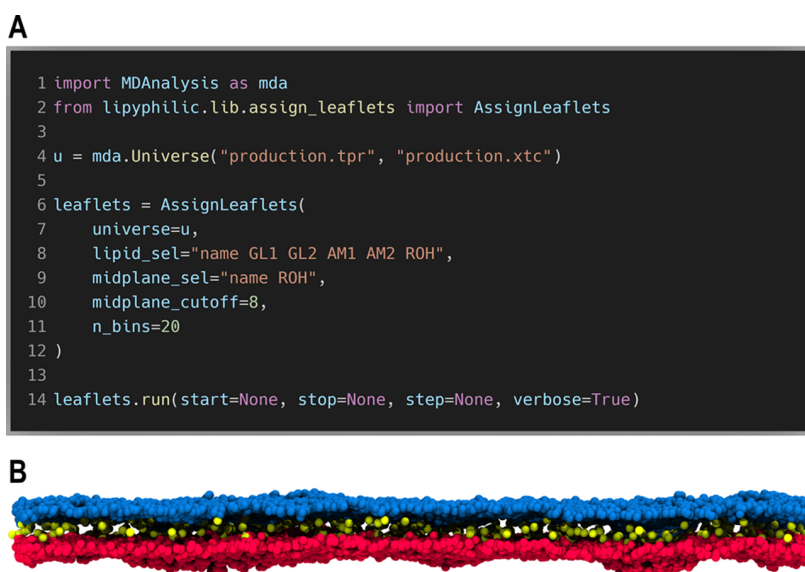
Below, we discuss the implementation and usage of some of the analysis tools currently available in LiPyphilic. We will then

discuss the on-the-fly transformations that LiPyphilic can perform on MDAnalysis trajectories. We then provide benchmarks of the analysis tools and transformations. Finally, we will briefly discuss the software engineering best practices used in developing LiPyphilic. If you are more interested in learning how to use LiPyphilic, rather than how LiPyphilic works per se, we recommend working through the online interactive tutorials, which are accessible via the documentation at <https://lipyphilic.readthedocs.io/en/latest/reference/tutorials.html>.

**2.1. Assign Leaflets.** For many analyses, such as calculating the area per lipid, it is necessary to know the leaflet within which a lipid is found. LiPyphilic has two tools for assigning lipids to leaflets. The class `lipyphilic.lib.assign_leaflets.AssignLeaflets` assigns each lipid to a leaflet based on the distance in  $z$  to its local membrane midpoint. This is suitable only for planar bilayers. On the other hand, the class `lipyphilic.lib.assign_leaflets.AssignCurvedLeaflets` can be used to identify leaflets in a buckled bilayer or a micelle. This uses the MDAnalysis leaflet finder<sup>50,51</sup> to assign non-translocating lipids to leaflets and then at each frame assigns the remaining lipids based on their minimum distance to each leaflet. `AssignLeaflets` remains useful for planar bilayers, especially if the rate of cholesterol translocation is of interest, which is typically measured by assigning lipids to leaflets based on their  $z$ -coordinate.

LiPyphilic can assign molecules not just to the upper or lower leaflet but also to the midplane. This is useful for studying, for example, the local lipid environment of midplane cholesterol<sup>26</sup> or its role in the registration of nanodomains.<sup>52</sup> Assigning cholesterol to the midplane also creates a buffer zone for determining whether a flip-flop event was successful (i.e., crossed the buffer zone) or not.<sup>27</sup>

`AssignLeaflets` and `AssignCurvedLeaflets` have opposing approaches to assigning molecules to the midplane. The former considers a molecule's distance to the midplane, whereas the latter considers its distance to each leaflet. There is naturally an inverse relationship between these two measures—the further a molecule is from the midplane, the closer it is to a leaflet. However, the distance to the midplane is typically



**Figure 2.** LiPyphilic can assign lipids to the upper leaflet, lower leaflet, or midplane. (A) Workflow for assigning leaflets. (B) Lipids in the neuronal plasma membrane studied by Ingólfsson et al.<sup>34</sup> are assigned to the upper leaflet (blue), lower leaflet (red), or midplane (yellow).

**A**

```
1 from liplyphilic.lib.flip_flop import FlipFlop
2
3 flip_flop = FlipFlop(
4     universe=u,
5     lipid_sel="name ROH",
6     leaflets=leaflets.filter_leaflets("name ROH"),
7     frame_cutoff=2
8 )
9
10 flip_flop.run(start=None, stop=None, step=None, verbose=True)
```

**B**

```
1 n_flip_flops = sum(flip_flop.flip_flop_success == "Success")
2
3 flip_flop_rate = n_flip_flops / (n_cholesterol * total_simulation_time)
```

**Figure 3.** (A) Identifying all cholesterol flip-flop events based on the leaflet membership of the cholesterol at each frame. (B) Determining the cholesterol flip-flop rate directly from the number of flip-flop events.

employed when studying flip-flop in planar bilayers,<sup>19,23–27,34</sup> whereas the distance to each leaflet is used for studying flip-flop in undulating bilayers.<sup>52</sup>

As with all analysis tools in LiPyphilic, the assigning of lipids to leaflets is both resolution- and force field-agnostic. Instead of reading atom selections hard coded into the package, the analysis tools rely on the powerful selection language of MDAnalysis. Figure 2A shows how `AssignLeaflets` may be used to determine the leaflet membership of all lipids in the 58-component neuronal plasma membrane studied by Ingólfsson et al.<sup>34</sup> First, an MDAnalysis Universe must be created. Lipids are then assigned to leaflets by passing this Universe to `AssignLeaflets` along with an atom selection of lipids in the bilayer, using the `universe` and `lipid_sel` arguments. Optionally, to allow molecules to be in the midplane, we can use the `midplane_sel` and `midplane_cutoff` arguments. In the example shown in Figure 2A, cholesterol will be assigned to the midplane if its ROH (hydroxyl group) bead is within 8 Å of its local midpoints. Local midpoints are computed by first splitting the membrane into an  $n$  by  $n$  grid in  $xy$ , where  $n$  is specified using the `n_bins` argument. The local midpoint of a grid cell is then given by the center of mass of all atoms selected by `lipid_sel` that are in the grid cell. Through calculating local membrane midpoints, this algorithm can account for small undulations in a bilayer. However, for bilayers with large undulations or for non-bilayer membranes, `AssignCurvedLeaflets` should be used.

After creating leaflets as described above, the analysis is performed by calling the `run` method. Here, the `start`, `stop`, and `step` arguments are used to specify which frames of the trajectories to use, and a progress bar can be displayed on the screen by setting `verbose=True`. Leaflet data are then stored in the `leaflets.leaflets` attribute as a two-dimensional NumPy array. Each row in the results array corresponds to an individual lipid and each column to an individual frame. For example, `leaflets.leaflets[i, j]` contains the leaflet membership of lipid  $i$  at frame  $j$ . `leaflets.leaflets[i, j]` is equal to 1 if the lipid is in the upper leaflet, -1 if the lipid is in the lower leaflet, or 0 if the lipid is in the midplane.

**2.2. Flip-Flop.** Cholesterol is unevenly distributed across the plasma membrane although the precise distribution is still under debate.<sup>53</sup> This uneven distribution plays an important role in numerous cellular processes and is maintained through the ultrafast spontaneous translocation, or flip-flop, of cholesterol across leaflets.

With recent advances in computing power, sterol flip-flop can now be studied directly using coarse-grained, united-atom, or all-atom simulations. Such simulations can be used to study the flip-flop process itself or to extract rates directly from the number of observed flip-flop events. Below, we describe a general analysis tool in LiPyphilic for identifying such flip-flop events in MD simulations.

The class `liplyphilic.lib.flip_flop.FlipFlop` can be used to identify successful and aborted flip-flop events. Figure 3A illustrates how to do so using the output from `AssignLeaflets`. The same MDAnalysis Universe that is used for assigning leaflets is passed to `FlipFlop`. An atom selection that specifies which molecules to consider when identifying flip-flop events is passed to the `lipid_sel` argument. The leaflet membership of each lipid selected by `lipid_sel` is passed to the `leaflets` argument. In this example, this is achieved by filtering the results array of `AssignLeaflets` to include only the leaflet membership of cholesterol molecules. The leaflet membership must be a NumPy array of shape  $(N_{\text{lipids}}, N_{\text{frames}})$  in which each element is equal to:

- 1 if the lipid is in the upper/outer leaflet
- -1 if the lipid is in the lower/inner leaflet
- 0 if the lipid is in the midplane

In this example, the `frame_cutoff` argument is used to specify that a molecule must remain in its new leaflet for at least two consecutive frames in order for the flip-flop to be considered successful.

We again call the `run` method to perform the analysis. For each molecule, LiPyphilic will then identify the frames at which it leaves one leaflet and enters another for at least `frame_cutoff` frames. If the new leaflet is different to the previous leaflet, the flip-flop was successful. If, on the other hand, the molecule left one leaflet, entered the midplane, and returned to the same leaflet as before, then the flip-flop failed.

```

1 from lipophilic.lib.registration import Registration
2
3 registration = Registration(
4     universe=u,
5     upper_sel="name ROH",
6     lower_sel="name ROH",
7     leaflets=leaflets.filter_leaflets("name ROH")
8 )
9
10 registration.run(start=None, stop=None, step=None, verbose=True)

```

**Figure 4.** Calculating the interleaflet registration of cholesterol at each frame.

The success or failure of each flip-flop event is stored in a one-dimensional NumPy array accessible via the `flip_flop.flip_flop_success` attribute. Elements in this array are strings, equal to either “Success” or “Failure”. From this results array, it is easy to calculate the flip-flop rate based on the number of observed events, the number of cholesterol molecules in the membrane, and the total simulation time (Figure 3B).

The example in Figure 3A shows how to use the results from `AssignLeaflets` to identify flip-flop events. However, leaflets need not be identified using `LiPyphilic` in order to use the `FlipFlop` analysis tool. `FlipFlop` expects a NumPy array of leaflet membership as described above. Flip-flop events can thus be found even if leaflets were assigned using, for example, `FATSLiM`<sup>3</sup> or user-written scripts<sup>54</sup> based on the `MDAnalysis LeafletFinder` tool.

Gu et al. showed that translocation is highly influenced by the local lipid environment of a sterol.<sup>26</sup> `FlipFlop` therefore returns not only the success or failure of each event but also the frame at which the flip-flop process begins and ends along with the residue index of the flip-flopping molecule. This information is stored as a two-dimensional NumPy array in the `flip_flop.flip_flops` attribute, where each row corresponds to an individual event and each column contains the

- residue index of the flip-flopping molecule
- frame at which the molecule left its original leaflet
- frame at which the molecule entered its new leaflet
- numerical identifier of its new leaflet: 1 for the upper leaflet and -1 for the lower leaflet

This information enables further analysis, such as a consideration of the local lipid environment before and after translocation.

Cholesterol is typically the only molecule to flip-flop during an MD simulation. However, ceramides and diacylglycerols and fatty acids such as docosahexaenoic acid also have fast flip-flop rates. To find flip-flop events of a molecule other than cholesterol, simply change the lipid selection passed to `FlipFlop`. For example, to find all flip-flop events for ceramides in the neuronal plasma membrane, change `lipid_sel="resname ROH"` to `lipid_sel="resname ??CE"` and pass the corresponding leaflet membership NumPy array to the `leaflets` argument. This again makes use of the powerful selection language of `MDAnalysis` and the fact that all ceramides in the MARTINI force field have residue names that are four characters long and end in “CE”.

**2.3. Registration.** The translocation of cholesterol across leaflets is thought to be important in several cellular processes, including the modulation of the lateral heterogeneity of the membrane.<sup>52</sup> Recently, transient nanodomains of  $L_o$  phase lipids were observed in live mammalian cell plasma membranes.<sup>55</sup>

These nanodomains are thought to be enriched in sphingomyelin and cholesterol and to act as functional platforms for cell signaling. However, their nature, formation, and roles in cellular processes are still not fully understood.

There is particular interest in understanding under what conditions nanodomains in apposing leaflets are spatially aligned. Such an alignment is known as interleaflet registration. This has been the subject of several MD simulations, which have revealed registration to be a complex process.<sup>13,37–47</sup> Registration is modulated by many factors, including the length and saturation of lipid tails as well as the relative affinity of cholesterol for the different lipid species in a domain-forming mixture.

The class `lipophilic.lib.registration.Registration` can be used to quantify the degree of interleaflet registration in a planar bilayer. `Registration` is an implementation of the registration analysis described by Thallmair et al.<sup>44</sup> The degree of registration is calculated as the Pearson correlation coefficient of molecular densities in the upper and lower leaflets. First, the two-dimensional density of each leaflet is calculated

$$\rho(x, y)_L = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma^2} \exp\left[-\frac{1}{2}\left[\left(\frac{x' - x}{\sigma}\right)^2 + \left(\frac{y' - y}{\sigma}\right)^2\right]\right] dx' dy'$$

where the  $(x, y)$  positions of lipid atoms in leaflet  $L$  are binned into two-dimensional histograms with bin lengths of 1 Å.  $L$  is either the upper (u) or lower (l) leaflet. The two-dimensional density is then convolved with a circular Gaussian density of standard deviation  $\sigma$ . The registration between the two leaflets,  $r_{u/l}$ , is then calculated as the Pearson correlation coefficient between  $\rho(x, y)_u$  and  $\rho(x, y)_l$ . Values of  $r_{u/l} = 1$  correspond to perfectly registered domains and values of  $r_{u/l} = -1$  correspond to perfectly anti-registered domains.

The atoms used in calculating the interleaflet registration are specified by passing selection strings to the `upper_sel` and `lower_sel` arguments of `Registration` (Figure 4). The leaflet membership of all atoms in the two selections must be passed to the `leaflets` argument. As before, this must be a two-dimensional NumPy array of shape  $(N_{\text{lipids}}, N_{\text{frames}})$ . The results are stored in the `registration.registration` attribute as a one-dimensional NumPy array of length  $N_{\text{frames}}$ . The array contains the Pearson correlation coefficient of the two-dimensional leaflet densities at each frame.

The example in Figure 4 demonstrates how to compute the registration of cholesterol across the upper and lower leaflets. However, in simulations of phase-separating mixtures, it is useful to know the degree of registration of  $L_o$  domains rather than the registration of a specific molecular species. If the phase of each lipid at each frame is known, `Registration` can be used to

**A**

```

1 from lipophilic.lib.neighbors import Neighbors
2
3 neighbors = Neighbors(
4     universe=u,
5     lipid_sel="name GL1 GL2 AM1 AM2 ROH",
6     cutoff=12
7 )
8
9 neighbours.run(start=None, stop=None, step=None, verbose=True)

```

**B**

```

1 largest_cluster, lipid_indices = neighbours.largest_cluster(
2     cluster_sel="name GM*",
3     filter_by=leaflets.leaflets == 1,
4     return_indices=True
5 )

```

**C**

```

1 counts, enrichment = neighbours.count_neighbors(return_enrichment=True)

```

**Figure 5.** (A) Creating the neighbor adjacency matrix. (B) Finding the largest cluster of glycolipids at each frame as well as the residue indices of lipids in the largest cluster. (C) Calculating the enrichment/depletion index of each lipid species.

calculate the registration of  $L_o$  or  $L_d$  domains over time. There are various approaches to determining the phase of lipids, from simple metrics such as the deuterium order parameter to more powerful machine learning methods such as hidden Markov models,<sup>29,30,36,56</sup> Smooth Overlap of Atomic Positions,<sup>57</sup> or those employed by MLLPA.<sup>7</sup> If the lipid phase data are stored in a two-dimensional NumPy array of shape  $(N_{\text{lipids}}, N_{\text{frames}})$ , it can be used to create a boolean mask that will tell Registration which lipids to include in the analysis. For example, if our array is named `lipid_phase_data` and its elements are strings of either "Lo" or "Ld", then we can select the  $L_o$  lipids for analysis by passing the boolean mask `lipid_phase_data == "Lo"` to the `filter_by` argument of Registration.

**2.4. Neighbor Matrix.** The plasma membrane comprises hundreds of different lipid species. In this complex mixture, lateral heterogeneities and aggregates of specific lipid species arise spontaneously. Over the past decade, this compositional complexity has begun to feature in MD simulations of membranes.<sup>19,27,34,58–63</sup> In these simulations, the lateral organization of the membrane is typically quantified via a consideration of local lipid environments. Specifically, the lipid enrichment index of species B around species A,  $E_{AB}$ , may be defined as<sup>19</sup>

$$E_{AB} = N_{AB} / \langle N_B \rangle$$

where  $N_{AB}$  is the number of molecules of species B around species A, and  $\langle N_B \rangle$  is the mean number of species B around any species.

The class `lipophilic.lib.neighbors.Neighbors` provides methods for computing the lipid enrichment index and for identifying the largest cluster of a specific species of lipids over time. Both of these analyses first require the construction of an adjacency matrix,  $A$ , that describes whether each pair of lipid molecules are neighboring one another or not. Two lipids are considered neighbors if they have any atoms within a user-defined cutoff distance,  $d_{\text{cutoff}}$ , of one another. The adjacency matrix can be created by passing an atom selection

and a value of  $d_{\text{cutoff}}$  to the `lipid_sel` and `cutoff` arguments, respectively, of `Neighbors` (Figure 5A). The `run` method is called to construct an adjacency matrix for each frame of the trajectory specified using the `start`, `stop`, and `step` arguments. The results are available in the `neighbors.neighbors` attribute as a NumPy array of SciPy sparse matrices. There is one adjacency matrix for each frame, and each matrix is of shape  $(N_{\text{lipids}}, N_{\text{lipids}})$ . These matrices can be used for further analysis either via helper methods of `Neighbors` or via user-written scripts.

**2.4.1. Largest Cluster.** Some glycolipids in the plasma membrane are known to aggregate, forming platforms for cell-signaling.<sup>64–66</sup> The size of the largest cluster of glycolipids in the neuronal plasma membrane<sup>34</sup> can be calculated using the `largest_cluster` method of `Neighbors` (Figure 5B). For this, an atom selection must be provided to the `cluster_sel` argument. In the example in Figure 5B, it is specified that only lipids in the upper leaflet should be included in the calculation by passing a boolean mask to the `filter_by` keyword. The `return_indices` argument is used to specify that the residue indices of the lipid molecules in the largest cluster at each frame are also to be returned. There is no need to call a `run` method or to specify which frames of the analysis to use—the same frames specified in Figure 5A will be used for the cluster analysis.

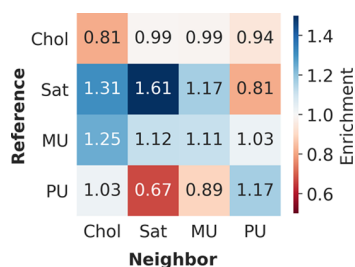
The results are not stored in an attribute in our `neighbors` object. Instead, the largest cluster size and the residue indices of the lipid molecules in the largest cluster are each returned as a NumPy array. The former is a one-dimensional array containing the number of lipids in the largest cluster at each frame. The latter is a list of NumPy arrays. Each array in the list corresponds to a single frame and contains the residue indices of the lipid molecules in the largest cluster at that frame. Knowing the indices of the lipids in the largest cluster allows for further analysis, such as calculating the lateral diffusion coefficient of lipid molecules in the cluster.<sup>37</sup>

To find the largest cluster at a given frame, the `neighbors.neighbors` sparse adjacency matrix is first sliced to give a matrix for the current frame only,  $A_{\text{frame}}$ . The `connected_components` function of SciPy is then used to find all connected components at the current frame. NumPy's `unique` function, with `return_counts` set to `True`, is then used to identify the largest connected component and thus the largest cluster size.

**2.4.2. Enrichment Index.** After constructing the adjacency matrix, the `count_neighbours` method can be used to determine the local environment of each lipid molecule (Figure 5C). For a single lipid, its local lipid environment is defined as the number of neighbors of each species. In the example in Figure 5C, `return_enrichment=True` is set to specify that the lipid enrichment index is also to be returned.

As with `neighbors.largest_cluster`, the results are not stored in an attribute in our `neighbors` object. The neighbor counts and the enrichment index are each returned as a Pandas DataFrame. The DataFrame of neighbor counts contains—for each lipid at each frame—the residue name, lipid residue index, frame number, number of neighbors of each species, and total number of neighbors. The lipid enrichment DataFrame contains the enrichment index of each lipid species at each frame. These data can easily be used to calculate the mean enrichment of each species, or it can be plotted over time to determine whether the lateral mixing of lipids has equilibrated.

**2.4.3. User-Defined Counts.** By default, the `count_neighbours` method will calculate the number of neighboring species around each individual lipid. This is done using the residue name of each lipid. However, it is also possible to use any ordinal or string data for counting lipid neighbors. For example, the enrichment index of lipids in the neuronal plasma membrane can be calculated based on their tail saturation (Figure 6). For



**Figure 6.** Enrichment/depletion index of lipids in the neuronal plasma membrane based on their tail saturation.

this, first a two-dimensional NumPy array of shape  $(N_{\text{lipids}}, N_{\text{frames}})$  that contains the saturation of each lipid needs to be created (Figure S1). Then, this array is passed to the `count_by` argument of `count_neighbours`, and the local lipid environment and enrichment index will be determined based on the information in this array.

**2.5. On-the-Fly Transformations.** MDAnalysis has a powerful set of on-the-fly trajectory transformations. These transformations can do away with the need to create multiple instances of the same trajectory using, for example, the GROMACS `trjconv` tool. Instead, the transformations are applied each time a frame is loaded into memory by MDAnalysis. LiPyphilic extends the set of transformations available in MDAnalysis to include the ability to repair membranes split across periodic boundaries and to perform

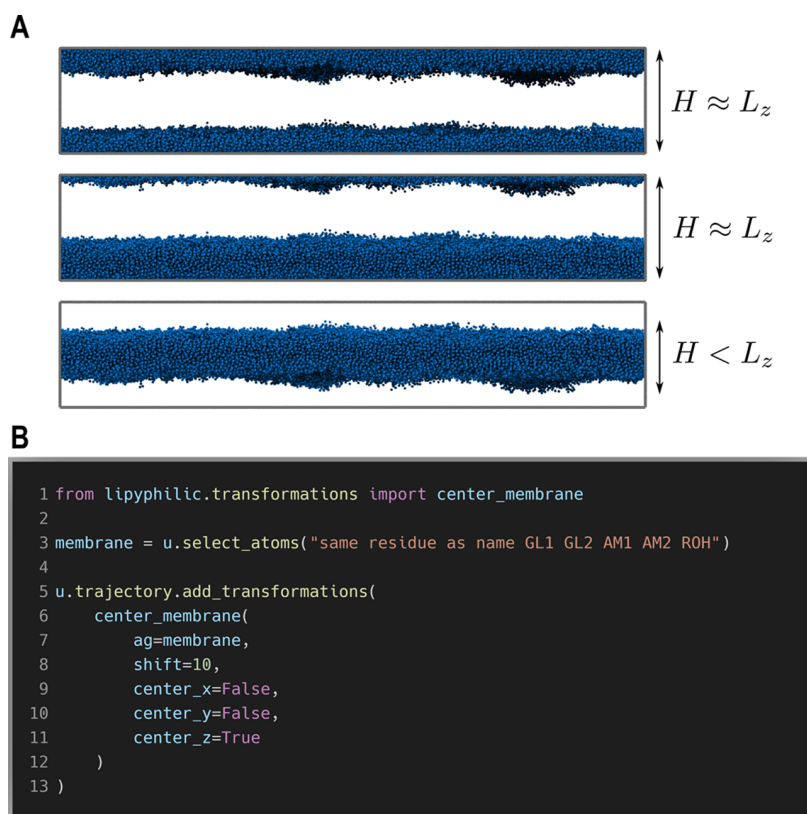
`nojump` trajectory unwrapping. The latter prevents an atom from being wrapped into the primary unit cell when it crosses a periodic boundary.

**2.5.1. Center Membranes.** The callable class `lipypphilic.transformations.center_membrane` can be used to fix a membrane—or any supramolecular structure—that is split across periodic boundaries and then center it in a box, providing it is not self-interacting across the periodic boundaries. For each frame, all atoms in the system are iteratively shifted along a specified set of dimensions until the membrane is no longer split across the periodic boundaries (Figure 7A). After each translation, all atoms are wrapped back into the primary unit cell. For example, to check if a bilayer is split across the periodic boundary in the  $z$ -dimension, its extent in  $z$ ,  $H$ , could be compared to the box length in  $z$ ,  $L_z$ . If  $H$  is within a user-specified cutoff value of  $L_z$ , the bilayer is split across  $z$  and is thus translated in this dimension. Once the bilayer is no longer split across boundaries, it is then moved to the center of the box in  $z$ .

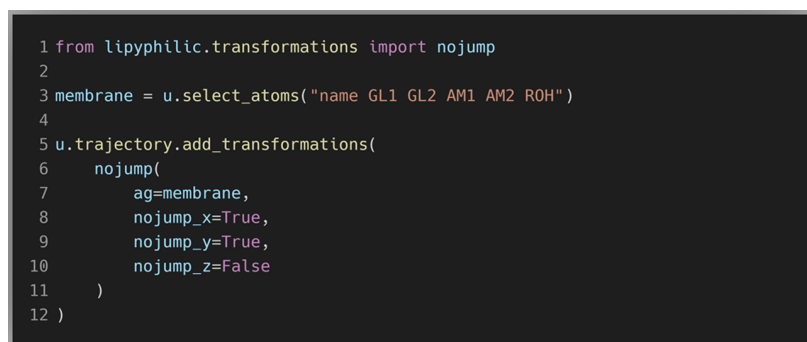
This transformation can be applied to an MDAnalysis universe, `u`, using the `u.trajectory.add_transformation` method (Figure 7A). This method takes as input a transformation or a list of transformations. The `center_membrane` callable class is passed to `add_transformation` along with the arguments for `center_membrane`. The atoms that comprise the membrane are specified using the `ag` argument. The atom selection should include all atoms in the membrane and not a subset of atoms—otherwise the extent of the membrane cannot be calculated accurately. In the example in Figure 7A, the bilayer is centered in only the  $z$ -dimension; however, the membrane can be made whole in each dimension independently. This is controlled using the `center_x`, `center_y`, and `center_z` arguments. The `shift` argument is used to specify the distance in Å that the membrane will be translated at each iteration. Too small a value of `shift` would require many iterations to make a membrane whole. Too large a value, on the other hand, may result in a membrane being translated nearly the length of the unit cell and thus remaining broken. We have found a translation of 10 Å to be suitable for bilayers, but the optimal value will depend on the membrane structure and the size of the system.

**2.5.2. nojump Trajectory Unwrapping.** `lipypphilic.transformations.nojump` can be used to prevent atoms from jumping across periodic boundaries. It is analogous, but not equivalent, to using the GROMACS command `trjconv` with the flag `-pbc nojump`. This transformation can be applied to an MDAnalysis universe in much the same way as `lipypphilic.transformations.center_membrane`. We must pass an atom selection to the `ag` argument of `nojump` and specify the dimensions to which the transformation should be applied (Figure 8).

Upon adding this transformation to your trajectory, `nojump` will perform an initial pass over the trajectory. It will determine the frames at which each atom crosses a boundary, keeping a record of the net movement of each atom at each boundary. This net movement across each boundary is used to determine the distance an atom must be translated in order to be moved from its wrapped position to its unwrapped position. Subsequently, every time a new frame is loaded into memory by MDAnalysis, such as when iterating over the trajectory, the relevant translation is applied to each atom to move it to its unwrapped coordinates.



**Figure 7.** (A) A membrane split across periodic boundaries can be made whole and centered by iteratively translating the system of particles. After each translation, a check is performed to determine whether the membrane is still split across boundaries. If the extent of the membrane in  $z$ ,  $H$ , is approximately equal to the box length in  $z$ ,  $L_z$ , then, the membrane is split across the periodic boundary. (B) Code snippet for applying the transformation to an MDAnalysis universe. The on-the-fly transformation can be applied to each dimension independently.



**Figure 8.** “nojump” on-the-fly transformation can be applied to any AtomGroup in an MDAnalysis Universe. The transformation can be applied to each dimension independently.

Below, we describe the nojump unwrapping algorithm implemented in LiPyphilic. We also explain how this algorithm avoids the artifacts introduced by the standard unwrapping scheme that, to our knowledge, is employed by all MD simulation-related software. Specifically, the standard unwrapping scheme fails to account for fluctuating system sizes caused by barostats in  $NPT$  ensemble simulations.

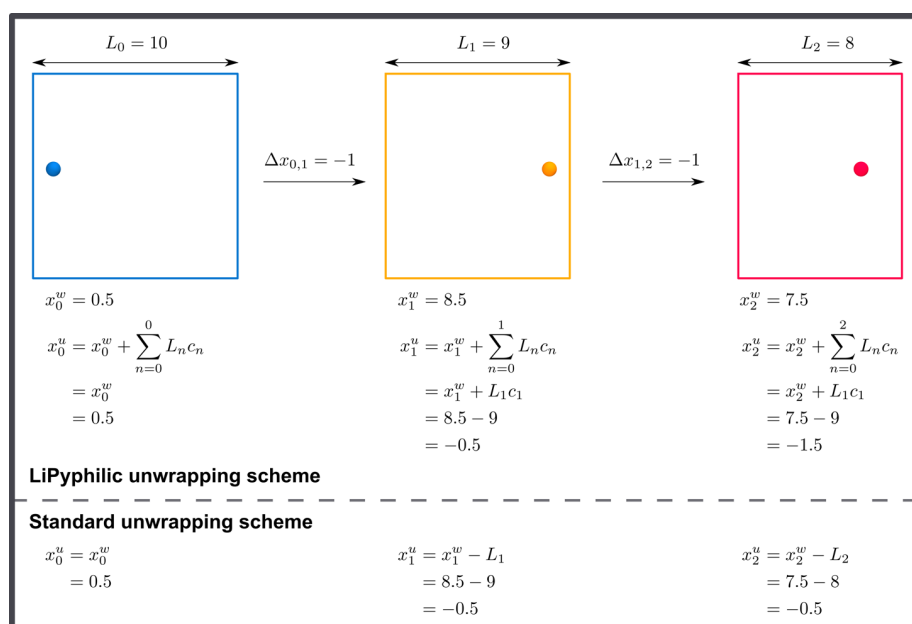
**2.5.3. Unwrapping Scheme.** The unwrapped position of a particle at frame  $N$ , denoted  $x_N^u$  is given by

$$x_N^u = x_N^w + \sum_{n=0}^N L_n c_n$$

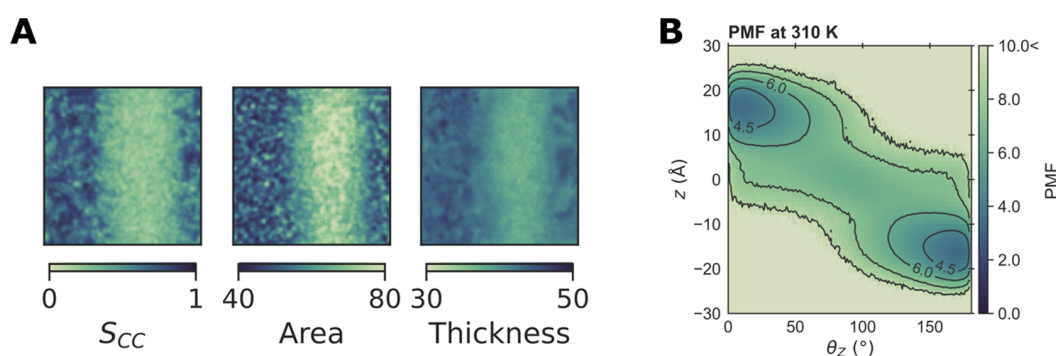
where  $x_N^w$  is the wrapped position of the particle at frame  $N$ , and  $\sum_{n=0}^N L_n c_n$  accounts for the displacement that results from all jumps across periodic boundaries from frame 0 to frame  $N$ .  $L_n$  is the box length at frame  $n$ , with the box centered at  $L/2$ . This box length is multiplied by a factor  $c_n$ , which depends on whether an atom crossed a periodic boundary from frame  $n - 1$  to frame  $n$

$$c_n = \begin{cases} -1 & x_n^w - x_{n-1}^w > L_n/2 \\ 1 & x_n^w - x_{n-1}^w < -L_n/2 \\ 0 & \text{otherwise} \end{cases}$$

where  $x_{n-1}^w$  is the particle's wrapped position at frame  $n - 1$ . At frame  $n = 0$ , for which there is no previous position, we take  $x_{-1}^w$



**Figure 9.** To correctly unwrap atomic coordinates, we must know size of the box at the frame at which a jump across periodic boundaries occurred. The standard unwrapping scheme produces an incorrect unwrapped coordinate at the second frame,  $x_2^u$ . Superscripts w and u denote whether the coordinate is wrapped or unwrapped, and subscripts  $n$  denote the frame number.



**Figure 10.** Plots produced using LiPyphilic. See the interactive tutorials for usage examples. (A) Projection onto the membrane plane of the coarse-grained order parameter ( $S_{CC}$ ), area per lipid ( $\text{\AA}^2$ ), and local membrane thickness ( $\text{\AA}$ ) of the phospholipids in an equimolar mixture of DPPC, DOPC, and cholesterol. (B) Potential of mean force (PMF) of cholesterol orientation ( $\theta_z$ ) and height ( $z$ ).

to be the particle's raw atomic coordinate at frame  $n = 0$ . That is, if the particle is not in the primary unit cell at frame  $n = 0$ , we calculate the displacement required to move from  $x_0^w$  to  $x_0^u$ . Note that this method will correctly unwrap coordinates for orthorhombic systems. A further correction can be applied for triclinic boxes,<sup>67</sup> which we plan to implement in a future release.

This unwrapping scheme is equivalent to that recently described by von Bülow et al.<sup>67</sup> although it was derived independently. Both our unwrapping algorithm and that of von Bülow et al. avoid the problems that the standard unwrapping scheme suffers from. To calculate  $x_N^u$ , the standard unwrapping scheme iteratively adds the box length at frame  $N$  to the wrapped coordinate at frame  $N$  until  $|x_n^w - x_{n-1}^u| < L_n/2$ . However, in the example in Figure 9, this would result in  $x_2^u = x_2^w - L_2 = -0.5$  instead of the correct value  $x_2^u = -1.5$ . von Bülow et al. demonstrated clearly the effect that this inaccurate unwrapping of atomic coordinates has on the calculated diffusion coefficient.<sup>67</sup>

The unwrapping scheme described above, and previously by von Bülow et al.,<sup>67</sup> correctly accounts for the fluctuating box size

in the *NPT* ensemble. However, it is only accurate in the case where coordinates are stored every timestep. In fact, it is impossible to correctly unwrap coordinates unless we store them at every timestep. This logically follows from the same argument made above—to correctly unwrap coordinates, we must know the length of the box at the timestep at which the jump occurred. See the [Supporting Information](#) for further details.

**2.6. Other Analysis Tools.** Although we have described some of the analysis tools that make LiPyphilic unique in the previous sections, there is much more functionality in LiPyphilic that is fully detailed in the documentation (<https://lipyphilic.readthedocs.io/en/latest>). This functionality includes calculating the coarse-grained lipid order parameter,<sup>68</sup> the area per lipid for planar bilayers, the lateral diffusion coefficient, and the membrane thickness of planar bilayers. There are also tools for calculating thickness, orientations, and  $z$ -positions of lipid molecules in a planar bilayer. Regarding the area per lipid tool, we recommend using either FATS<sup>3</sup> or MemSurfer<sup>4</sup> if you have a curved membrane. These tools are designed specifically to deal with undulating bilayers and non-bilayer structures,

**Table 1. Benchmark Times for Analysis Tools in LiPyphilic Using a MARTINI Bilayer of 12,000 Lipids<sup>a</sup>**

analysis/transformation	LiPyphilic class/method	Time per frame (ms)	
		LiPyphilic	comparison
leaflet identification	AssignLeaflets	17.54	168.47 <sup>b</sup>
	AssignCurvedLeaflets	29.85	168.47 <sup>b</sup>
flip-flop <sup>d</sup>	FlipFlop	0.5	
interleaflet registration	Registration	151.81	
construct neighbor matrix	Neighbours	260.97	
largest cluster <sup>e</sup>	Neighbours.largest_cluster	1.85	
enrichment index	Neighbours.count_neighbours	100.93	b
bilayer thickness	MembThickness	16.35	394.11 <sup>b</sup>
lipid thickness	ZThickness	37.62	
lipid height	ZPositions	19.05	
lipid orientation	ZAngles	22.87	
area per lipid	AreaPerLipid	1589.29	206.75 <sup>b</sup>
coarse-grained order parameter <sup>f</sup>	SCC	16.10	119.09 <sup>c</sup>
unwrap membrane	Center_membrane	23.37	
nojump unwrapping	Nojump	23.89	12.04 <sup>c</sup>

<sup>a</sup>Where possible, performance is compared with either FATS�iM 0.2.2 or GROMACS 2020.4. The FATS�iM benchmarks were performed using eight OpenMP threads. All other benchmarks were performed in series. <sup>b</sup>FATS�iM. <sup>c</sup>GROMACS. <sup>d</sup>Time per molecule (ms) over 1600 frames. <sup>e</sup>Time per frame (ms) for a subset of 2000 lipids. <sup>f</sup>Time per frame (ms) for the sn-1 tail of DPPC (4000 lipids).

whereas the area per lipid tool of LiPyphilic will only produce reliable values in the case of planar bilayers.

In general, LiPyphilic does not handle plotting of analysis data. However, it does have plotting utilities for visualizing joint potentials of mean force (PMFs)—such as the PMF of cholesterol orientation and height—and for the projection of membrane properties onto the *xy* plane (Figure 10). There are full descriptions of all LiPyphilic tools in the documentation, and our interactive tutorials (available at <https://lipyphilic.readthedocs.io/en/latest/reference/tutorials.html>) provide examples of how to use the analysis tools and plot the results.

### 3. BENCHMARKING

For benchmarking the LiPyphilic analysis tools, we have performed a simulation of an equimolar mixture of DPPC/DOPC/Cholesterol using the MARTINI 2 force field.<sup>69,70</sup> We created a symmetric bilayer of 12,000 lipids in total using the CHARMM-GUI MARTINI Maker.<sup>71</sup> The production run was performed for 8.0  $\mu$ s and coordinates were stored every 5.0 ns, giving a trajectory of 1600 frames. Where analogous analysis tools are available in either FATS�iM<sup>3</sup> 0.2.2 or GROMACS<sup>72</sup> 2020.4, we compare their performance with that of LiPyphilic. We benchmark against FATS�iM as this is generally the fastest membrane analysis tool available.<sup>3</sup> The FATS�iM benchmarks were performed using eight OpenMP threads. All other benchmarks were performed in serial. All of the benchmarks can be seen in Table 1.

LiPyphilic is generally very fast, with most analysis tools and trajectory transformations taking on the order of 10 ms per frame for the 12,000 lipid membrane. For example, both methods of assigning leaflets are faster than the corresponding implementation in FATS�iM. This is down to the algorithms used for assigning lipids to leaflets—FATS�iM calculates a local membrane normal for each lipid based on the point cloud of neighboring lipids and then generates leaflets based on the distance and relative orientation of groups of lipids. `AssignLeaflets`, on the other hand, uses only the distance in *z* to the midplane. `AssignCurvedLeaflets`, meanwhile, is computationally expensive for the first frame. This is because a graph is constructed from the positions of non-translocating lipid

headgroups, and from this, the leaflets are identified as the two largest connected components. Subsequent frames, however, assign potentially translocating lipids to a leaflet based on their distance to each leaflet. Thus, the longer a trajectory, the more computationally efficient `AssignCurvedLeaflets` becomes. There are two further benefits of assigning leaflets with LiPyphilic: (i) molecules may reside in the midplane and (ii) the results are stored in a single NumPy array, whereas FATS�iM creates a GROMACS index file for each frame of the trajectory.

LiPyphilic is significantly faster in calculating the bilayer thickness although the implementation in FATS�iM is more sophisticated. Part of the reason LiPyphilic is faster is that it uses the leaflet information calculated by `AssignLeaflets` or `AssignCurvedLeaflets`, whereas FATS�iM assigns the lipids to leaflets at each frame before it calculates the bilayer thickness. The algorithm used in LiPyphilic for calculating membrane thickness is also simpler (but less versatile). The FATS�iM thickness tool effectively constructs a smoothed surface for each leaflet and calculates the distance from each lipid to the apposing leaflet. In this way, FATS�iM is able to calculate the thickness of both planar and non-planar bilayers. LiPyphilic's `MembThickness` tool, however, constructs a two-dimensional surface of each leaflet and then calculates the bilayer thickness as the mean separation in *z* between the two leaflet surfaces. This means that `MembThickness` is only appropriate for planar bilayers.

LiPyphilic is also faster than GROMACS when it comes to calculating the coarse-grained order parameter,  $S_{CC}$ . Using LiPyphilic to calculate  $S_{CC}$  is also simpler than using GROMACS, which requires the creation of a separate index group for each unique atom along a tail of each lipid species.

The calculation of interleaflet registration, construction of a neighbor matrix, and calculation of the lipid enrichment index are slower, on the order of 100 ms per frame. This is still relatively fast—although they are different analyses, these times are comparable to the performance of the various analyses available in FATS�iM such as the area per lipid and membrane thickness calculations.



The slowest analysis in LiPyphilic is the calculation of the area per lipid. This takes over 1.5 s per frame and is 7.6 times slower than FATSliM. We therefore recommend using FATSliM for the area per lipid calculation if you have a large membrane (>1000 lipids). It should be remembered, however, that the FATSliM analyses were run in parallel over eight cores, whereas all other benchmarks were performed in serial. Parallelizing the analysis in LiPyphilic could result in a similar performance boost. Although this is out of the scope of the package in its current state, we do have plans to parallelize the slower tools in LiPyphilic in due course. In this respect, we are particularly interested in the development of Parallel MDAnalysis (PMDA).<sup>73</sup> PMDA is based on MDAnalysis and uses Dask to parallelize the analysis modules.<sup>74</sup> The analysis classes inherit a modified abstract base class that is specifically designed to make the parallelization straightforward. We will wait until PMDA is out of the alpha stage of development before assessing which modules would benefit from being parallelized.

Finally, the on-the-fly transformations are fast, with `center_membrane` and `nojump` taking 23.37 and 23.89 ms per frame, respectively. Upon applying the `nojump` transformation, a first pass over the trajectory is performed to calculate the translations that need to be applied at each frame. This first pass takes 13.98 ms per frame for the 12,000 atoms selected in the benchmark. After this first pass, the time to load a frame into memory and apply the translations is 9.91 ms. This performance is put into perspective by considering that iterating over the trajectory with MDAnalysis takes 8.54 ms per frame itself, with no transformations applied. The total time per frame of `nojump` (23.89 ms) is approximately twice that required by the GROMACS `trjconv` tool to do the same transformation. Using `nojump` has two benefits over GROMACS `trjconv` in that it (i) prevents the need to create duplicate trajectories and (ii) accounts for box size fluctuations caused by barostats.

#### 4. SOFTWARE ENGINEERING

LiPyphilic is a free, open-source software licensed under the GNU General Public License v2 or later. In developing LiPyphilic, we have followed the best practices in modern scientific software engineering.<sup>75</sup> We use version control, unit testing, and continuous integration, and we have a fully documented API with examples of how to use each analysis tool.

The full development history and planned improvements of the project are available to view on GitHub, at <https://github.com/p-j-smith/lipyphilic>. LiPyphilic loosely follows the GitHub-flow model of software development<sup>76</sup>—developing directly from the master branch and releasing new versions soon after new functionality or fixes are added. We encourage users to submit feature requests and bug reports via GitHub, and we welcome any question about usage on our discussion page at <https://github.com/p-j-smith/lipyphilic/discussions>.

Unit testing in LiPyphilic is performed using Pytest.<sup>77</sup> We have constructed a set of toy systems for testing each analysis tool. These systems are typically composed of two sets of atoms each arranged on a hexagonal lattice in *xy*, with the two lattices separated vertically in *z*. This setup approximates the topology of the headgroups of a lipid bilayer. Using these toy systems, we know what the results of each analysis tool should be a priori. We can thus test each analysis tool with full confidence in the results if the tests pass, without relying on regression tests that involve highly complex systems. Further, using Pytest-cov,<sup>78</sup> we have ensured that all analysis tools and trajectory transformations have 100% test coverage.

Finally, LiPyphilic is simple to install. We have packaged LiPyphilic to make it available for installation using widely used package managers. The easiest way to install LiPyphilic along with all of its dependencies is through Anaconda.<sup>79</sup> Alternatively, it can be installed via the Python Package Index using Pip.<sup>80</sup>

#### 5. CONCLUSIONS

We have developed a new Python package for the analysis of lipid membrane simulations. We have focused on providing functionality not available in other membrane analysis tools, such as calculating the lipid enrichment/depletion index, the degree of interleaflet registration, and the flip-flop rate of molecules between leaflets. LiPyphilic is a modular object-oriented package that makes extensive use of NumPy,<sup>48</sup> SciPy,<sup>49</sup> and MDAnalysis<sup>50,51</sup> for efficient computation. For analyzing a 12,000 lipid MARTINI membrane, the analysis classes typically take on the order of 10–100 ms per frame. This is comparable to the performance of analysis tools in GROMACS<sup>72</sup> and FATSliM<sup>3</sup>—a very fast package for membrane analysis. All analysis tools in LiPyphilic share the same API as those of MDAnalysis. This shared API makes LiPyphilic simple to learn for current users of MDAnalysis.

The modularity of LiPyphilic, along with its focus on integrating with the wider scientific Python stack, means the output of other analysis tools such as FATSliM<sup>3</sup> or MLLPA<sup>7</sup> can be used as input for further analysis in LiPyphilic. Further, the output of LiPyphilic is in the form of NumPy arrays, SciPy sparse matrices, or Pandas Dataframes. This means the results can readily be plotted or further analyzed using the standard libraries of the scientific Python stack. For examples of how to do so, see our interactive tutorials available at <https://lipyphilic.readthedocs.io/en/latest/reference/tutorials.html>.

LiPyphilic is built upon sound software engineering principles. It uses version control, is fully unit-tested, employs continuous integration, and has extensive documentation. LiPyphilic is also trivial to install—it can be installed using either Anaconda or Pip.<sup>79,80</sup> We encourage users to submit feature requests and bug reports via GitHub and are always open to new contributors to the project.

#### ■ ASSOCIATED CONTENT

##### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00447>.

Comparison of LiPyphilic with other software for lipid membrane analysis; Python script for calculating the lipid enrichment index based on tail saturation; and discussion of practical issues with `nojump` trajectory unwrapping (PDF)

#### ■ AUTHOR INFORMATION

##### Corresponding Authors

Paul Smith — Department of Physics, King's College London, London WC2R 2LS, U.K.; [orcid.org/0000-0002-3676-5318](https://orcid.org/0000-0002-3676-5318); Email: [paul.smith@kcl.ac.uk](mailto:paul.smith@kcl.ac.uk)

Christian D. Lorenz — Department of Physics, King's College London, London WC2R 2LS, U.K.; [orcid.org/0000-0003-1028-4804](https://orcid.org/0000-0003-1028-4804); Email: [chris.lorenz@kcl.ac.uk](mailto:chris.lorenz@kcl.ac.uk)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00447>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202/1, EP/R029431/1, EP/T022213), this work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>) and the UK Materials and Molecular Modelling Hub (MMM Hub) for computational resources, which is partially funded by EPSRC (EP/P020194/1, EP/T022213), to carry out the MD simulations of DPPC/DOPC/Cholesterol reported in this manuscript. P.S. acknowledges the funding provided by the EPSRC DTP Studentship Block Grant (EP/N509498/1). C.D.L. acknowledges the supportive research environment of the EPSRC center for Doctoral Training in Cross-Disciplinary Approaches to Non-Equilibrium Systems (CANES, no. EP/L015854/1).

## REFERENCES

- Grecco, H. E.; Schmick, M.; Bastiaens, P. I. H. Signaling from the living plasma membrane. *Cell* **2011**, *144*, 897–909.
- Sunshine, H.; Iruela-Arispe, M. L. Membrane lipids and cell signaling. *Curr. Opin. Lipidol.* **2017**, *28*, 408–413.
- Buchoux, S. FATSlim: a fast and robust software to analyze MD simulations of membranes. *Bioinformatics* **2016**, *33*, 133–134.
- Bhatia, H.; Ingólfsson, H. I.; Carpenter, T. S.; Lightstone, F. C.; Bremer, P.-T. MemSurfer: A Tool for Robust Computation and Characterization of Curved Membranes. *J. Chem. Theory Comput.* **2019**, *15*, 6411–6421.
- Song, W.; Corey, R. A.; Ansell, B.; Cassidy, K.; Horrell, M.; Duncan, A.; Stansfeld, P. J.; Sansom, M. PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations. **2021**, bioRxiv:2021.07.14.452312.
- Sejdiu, B. I.; Tieleman, D. P. ProLint: a web-based framework for the automated data analysis and visualization of lipid-protein interactions. *Nucleic Acids Res.* **2021**, *49*, W544–W550.
- Walter, V.; Ruscher, C.; Benzerara, O.; Thalmann, F. MLLPA: A Machine Learning-assisted Python module to study phase-specific events in lipid membranes. *J. Comput. Chem.* **2021**, *42*, 930–943.
- Romo, T. D.; Grossfield, A. LOOS: An extensible platform for the structural analysis of simulations. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2009; pp 2332–2335.
- Romo, T. D.; Leioatts, N.; Grossfield, A. Lightweight object oriented structure analysis: Tools for building tools to analyze molecular dynamics simulations. *J. Comput. Chem.* **2014**, *35*, 2305–2318.
- Tieleman, D. P.; Marrink, S.-J. Lipids Out of Equilibrium: Energetics of Desorption and Pore Mediated Flip-Flop. *J. Am. Chem. Soc.* **2006**, *128*, 12462–12467.
- Gurtovenko, A. A.; Vattulainen, I. Molecular Mechanism for Lipid Flip-Flops. *J. Phys. Chem. B* **2007**, *111*, 13554–13559.
- Róg, T.; Stimson, L. M.; Pasenkiewicz-Gierula, M.; Vattulainen, I.; Karttunen, M. Replacing the cholesterol hydroxyl group with the ketone group facilitates sterol flip-flop and promotes membrane fluidity. *J. Phys. Chem. B* **2008**, *112*, 1946–1952.
- Risselada, H. J.; Marrink, S. J. The molecular face of lipid rafts in model membranes. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17367–17372.
- Kucerka, N.; Perlmutter, J. D.; Pan, J.; Tristram-Nagle, S.; Katsaras, J.; Sachs, J. N. The effect of cholesterol on short- and long-chain monounsaturated lipid bilayers as determined by molecular dynamics simulations and X-ray scattering. *Biophys. J.* **2008**, *95*, 2792–2805.
- Bennett, W. F. D.; MacCallum, J. L.; Hinner, M. J.; Marrink, S. J.; Tieleman, D. P. Molecular view of cholesterol flip-flop and chemical potential in different membrane environments. *J. Am. Chem. Soc.* **2009**, *131*, 12714–12720.
- Bennett, W. F. D.; Tieleman, D. P. Molecular simulation of rapid translocation of cholesterol, diacylglycerol, and ceramide in model raft and nonraft membranes. *J. Lipid Res.* **2012**, *53*, 421–429.
- Ogushi, F.; Ishitsuka, R.; Kobayashi, T.; Sugita, Y. Rapid flip-flop motions of diacylglycerol and ceramide in phospholipid bilayers. *Chem. Phys. Lett.* **2012**, *522*, 96–102.
- Choubey, A.; Kalia, R. K.; Malmstadt, N.; Nakano, A.; Vashishta, P. Cholesterol translocation in a phospholipid membrane. *Biophys. J.* **2013**, *104*, 2429–2436.
- Ingólfsson, H. I.; Melo, M. N.; Van Eerden, F. J.; Arnarez, C.; Lopez, C. A.; Wassenaar, T. A.; Periole, X.; De Vries, A. H.; Tieleman, D. P.; Marrink, S. J. Lipid organization of the plasma membrane. *J. Am. Chem. Soc.* **2014**, *136*, 14554–14559.
- Marquardt, D.; Kučerka, N.; Wassall, S. R.; Harroun, T. A.; Katsaras, J. Cholesterol's location in lipid bilayers. *Chem. Phys. Lipids* **2016**, *199*, 17–25.
- Javanainen, M.; Martinez-Seara, H. Rapid diffusion of cholesterol along polyunsaturated membranes: Via deep dives. *Phys. Chem. Chem. Phys.* **2019**, *21*, 11660–11669.
- Miettinen, M. S.; Lipowsky, R. Bilayer Membranes with Frequent Flip-Flops Have Tensionless Leaflets. *Nano Lett.* **2019**, *19*, 5011–5016.
- Oh, Y.; Sung, B. J. Facilitated and Non-Gaussian Diffusion of Cholesterol in Liquid Ordered Phase Bilayers Depends on the Flip-Flop and Spatial Arrangement of Cholesterol. *J. Phys. Chem. Lett.* **2018**, *9*, 6529–6535.
- Aghaaminiha, M.; Farnoud, A. M.; Sharma, S. Quantitative relationship between cholesterol distribution and ordering of lipids in asymmetric lipid bilayers. *Soft Matter* **2021**, *17*, 2742–2752.
- Filipe, H. A. L.; Javanainen, M.; Salvador, A.; Galvã, A. M.; Vattulainen, I.; Luís, L. M. S.; João Moreno, M. Quantitative Assessment of Methods Used To Obtain Rate Constants from Molecular Dynamics Simulations—Translocation of Cholesterol across Lipid Bilayers. *J. Chem. Theory Comput.* **2018**, *14*, 3840.
- Gu, R.-X.; Baoukina, S.; Tieleman, D. P. Cholesterol Flip-Flop in Heterogeneous Membranes. *J. Chem. Theory Comput.* **2019**, *15*, 2064–2070.
- Baral, S.; Levental, I.; Lyman, E. Composition dependence of cholesterol flip-flop rates in physiological mixtures. *Chem. Phys. Lipids* **2020**, *232*, 104967.
- Carter, J. W.; Gonzalez, M. A.; Brooks, N. J.; Seddon, J. M.; Bresme, F. Flip-flop asymmetry of cholesterol in model membranes induced by thermal gradients. *Soft Matter* **2020**, *16*, 5925–5932.
- Sodt, A. J.; Sandar, M. L.; Gawrisch, K.; Pastor, R. W.; Lyman, E. The molecular structure of the liquid-ordered phase of lipid bilayers. *J. Am. Chem. Soc.* **2014**, *136*, 725–732.
- Sodt, A. J.; Pastor, R. W.; Lyman, E. Hexagonal Substructure and Hydrogen Bonding in Liquid-Ordered Phases Containing Palmitoyl Sphingomyelin. *Biophys. J.* **2015**, *109*, 948–955.
- Ackerman, D. G.; Feigenson, G. W. Multiscale modeling of four-component lipid mixtures: Domain composition, size, alignment, and properties of the phase interface. *J. Phys. Chem. B* **2015**, *119*, 4240–4250.
- Lin, X.; Lorent, J. H.; Skinkle, A. D.; Levental, K. R.; Waxham, M. N.; Gorge, A. A.; Levental, I. Domain stability in biomimetic membranes driven by lipid polyunsaturation. *J. Phys. Chem. B* **2016**, *120*, 11930–11941.
- Sodt, A. J.; Venable, R. M.; Lyman, E.; Pastor, R. W. Nonadditive Compositional Curvature Energetics of Lipid Bilayers. *Phys. Rev. Lett.* **2016**, *117*, 138104.
- Ingólfsson, H. I.; Carpenter, T. S.; Bhatia, H.; Bremer, P.-T.; Marrink, S. J.; Lightstone, F. C. Computational Lipidomics of the Neuronal Plasma Membrane. *Biophys. J.* **2017**, *113*, 2271–2280.
- Koshiyama, K.; Taneo, M.; Shigematsu, T.; Wada, S. Bicelle-to-Vesicle Transition of a Binary Phospholipid Mixture Guided by Controlled Local Lipid Compositions: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2019**, *123*, 3118–3123.

- (36) Smith, P.; Quinn, P. J.; Lorenz, C. D. Two coexisting membrane structures are defined by lateral and transbilayer interactions between sphingomyelin and cholesterol. *Langmuir* **2020**, *36*, 9786–9799.
- (37) Gu, R.-X.; Baoukina, S.; Tieleman, D. P. Phase Separation in Atomistic Simulations of Model Membranes. *J. Am. Chem. Soc.* **2020**, *142*, 2844–2856.
- (38) Perlmutter, J. D.; Sachs, J. N. Interleaflet Interaction and Asymmetry in Phase Separated Lipid Bilayers: Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 6563–6577.
- (39) Nickels, J. D.; Smith, J. C.; Cheng, X. Lateral organization, bilayer asymmetry, and inter-leaflet coupling of biological membranes. *Chem. Phys. Lipids* **2015**, *192*, 87–99.
- (40) Fowler, P. W.; Williamson, J. J.; Sansom, M. S. P.; Olmsted, P. D. Roles of Interleaflet Coupling and Hydrophobic Mismatch in Lipid Membrane Phase-Separation Kinetics. *J. Am. Chem. Soc.* **2016**, *138*, 11633–11642.
- (41) Reigada, R. Alteration of interleaflet coupling due to compounds displaying rapid translocation in lipid membranes. *Sci. Rep.* **2016**, *6*, 32934.
- (42) Fujimoto, T.; Parmryd, I. Interleaflet coupling, pinning, and leaflet asymmetry-major players in plasma membrane nanodomain formation. *Front. Cell Dev. Biol.* **2017**, *4*, 155.
- (43) Galimzyanov, T. R.; Kuzmin, P. I.; Pohl, P.; Akimov, S. A. Undulations Drive Domain Registration from the Two Membrane Leaflets. *Biophys. J.* **2017**, *112*, 339–345.
- (44) Thallmair, S.; Ingólfsson, H. I.; Marrink, S. J. Cholesterol Flip-Flop Impacts Domain Registration in Plasma Membrane Models. *J. Phys. Chem. Lett.* **2018**, *9*, 5527–5533.
- (45) Weiner, M. D.; Feigenson, G. W. Molecular Dynamics Simulations Reveal Leaflet Coupling in Compositionally Asymmetric Phase-Separated Lipid Membranes. *J. Phys. Chem. B* **2019**, *123*, 3968–3975.
- (46) Zhang, S.; Lin, X. Lipid Acyl Chain cis Double Bond Position Modulates Membrane Domain Registration/Anti-Registration. *J. Am. Chem. Soc.* **2019**, *141*, 15884–15890.
- (47) Sarmiento, M. J.; Hof, M.; Sächl, R. Interleaflet Coupling of Lipid Nanodomains—Insights From in vitro Systems. *Front. Cell Dev. Biol.* **2020**, *8*, 284.
- (48) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.
- (49) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; van der Walt, S. J.; Brett, M.; Wilson, J.; Millman, K. J.; Mayorov, N.; Nelson, A. R. J.; Jones, E.; Kern, R.; Larson, E.; Carey, C. J.; Polat, I.; Feng, Y.; Moore, E. W.; VanderPlas, J.; Laxalde, D.; Perktold, J.; Cimrman, R.; Henriksen, I.; Quintero, E. A.; Harris, C. R.; Archibald, A. M.; Ribeiro, A. H.; Pedregosa, F.; van Mulbregt, P.; Vijaykumar, A.; Bardelli, A. P.; Rothberg, A.; Hilboll, A.; Kloeckner, A.; Scopatz, A.; Lee, A.; Rokem, A.; Woods, C. N.; Fulton, C.; Masson, C.; Häggström, C.; Fitzgerald, C.; Nicholson, D. A.; Hagen, D. R.; Pasechnik, D. V.; Olivetti, E.; Martin, E.; Wieser, E.; Silva, F.; Lenders, F.; Wilhelm, F.; Young, G.; Price, G. A.; Ingold, G.-L.; Allen, G. E.; Lee, G. R.; Audren, H.; Probst, I.; Dietrich, J. P.; Silterra, J.; Webber, J. T.; Glavič, J.; Nothman, J.; Buchner, J.; Kulick, J.; Schönberger, J. L.; de Miranda Cardoso, J. V.; Reimer, J.; Harrington, J.; Rodríguez, J. L. C.; Nunez-Iglesias, J.; Kuczynski, J.; Tritz, K.; Thoma, M.; Newville, M.; Kümmeler, M.; Bolingbroke, M.; Tartre, M.; Pak, M.; Smith, N. J.; Nowaczyk, N.; Shebanov, N.; Pavlyk, O.; Brodtkorb, P. A.; Lee, P.; McGibbon, R. T.; Feldbauer, R.; Lewis, S.; Tygier, S.; Sievert, S.; Vigna, S.; Peterson, S.; More, S.; Pudlik, T.; Oshima, T.; Pingel, T. J.; Robitaille, T. P.; Spura, T.; Jones, T. R.; Cera, T.; Leslie, T.; Zito, T.; Krauss, T.; Upadhyay, U.; Halchenko, Y. O.; Vázquez-Baeza, Y. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat. Methods* **2020**, *17*, 261–272.
- (50) Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.
- (51) Gowers, R.; Linke, M.; Barnoud, J.; Reddy, T.; Melo, M.; Seyler, S.; Domański, J.; Dotson, D.; Buchoux, S.; Kenney, I.; Beckstein, O. MDAAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. In *Proceedings of the 15th Python in Science Conference*; Benthall, S., Rostrup, S., Eds.; SciPy: Austin, Texas, 2016; July 11–17.
- (52) Weiner, M. D.; Feigenson, G. W. Presence and Role of Midplane Cholesterol in Lipid Bilayers Containing Registered or Antiregistered Phase Domains. *J. Phys. Chem. B* **2018**, *122*, 8193.
- (53) Doktorova, M.; Symons, J. L.; Levental, I. Structural and functional consequences of reversible lipid asymmetry in living membranes. *Nat. Chem. Biol.* **2020**, *16*, 1321–1330.
- (54) Bhatia, H.; Bremer, P.-T. Example usage of MemSurfer—a bilayer membrane consisting of three types of lipid. [https://github.com/LLNL/MemSurfer/blob/develop/examples/ex\\_g3lipid.py](https://github.com/LLNL/MemSurfer/blob/develop/examples/ex_g3lipid.py) (accessed April 9, 2021).
- (55) Heberle, F. A.; Doktorova, M.; Scott, H. L.; Skinkle, A. D.; Waxham, M. N.; Levental, I. Direct label-free imaging of nanodomains in biomimetic and biological membranes by cryogenic electron microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 19943–19952.
- (56) Park, S.; Im, W. Analysis of Lipid Order States and Domains in Lipid Bilayer Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 688–697.
- (57) Capelli, R.; Gardin, A.; Empereur-mot, C.; Doni, G.; Pavan, G. M. A Data-Driven Dimensionality Reduction Approach to Compare and Classify Lipid Force Fields. *J. Phys. Chem. B* **2021**, *125*, 7785–7796.
- (58) Koldso, H.; Shorthouse, D.; Hélie, J.; Sansom, M. S. P. Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers. *PLoS Comput. Biol.* **2014**, *10*, No. e1003911.
- (59) Lorent, J. H.; Levental, K. R.; Ganesan, L.; Rivera-Longworth, G.; Sezgin, E.; Doktorova, M.; Lyman, E.; Levental, I. Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape. *Nat. Chem. Biol.* **2020**, *16*, 644–652.
- (60) Ingólfsson, H. I.; Bhatia, H.; Zeppelin, T.; Bennett, W. F. D.; Carpenter, K. A.; Hsu, P.-C.; Dharuman, G.; Bremer, P.-T.; Schiøtt, B.; Lightstone, F. C.; Carpenter, T. S. Capturing Biologically Complex Tissue-Specific Membranes at Different Levels of Compositional Complexity. *J. Phys. Chem. B* **2020**, *124*, 7819–7829.
- (61) Wilson, K. A.; MacDermott-Opeskin, H. I.; Riley, E.; Lin, Y.; O'Mara, M. L. Understanding the Link between Lipid Diversity and the Biophysical Properties of the Neuronal Plasma Membrane. *Biochemistry* **2020**, *59*, 3010–3018.
- (62) Wilson, K. A.; Fairweather, S. J.; MacDermott-Opeskin, H. I.; Wang, L.; Morris, R. A.; O'Mara, M. L. The role of plasmalogens, Forssman lipids, and sphingolipid hydroxylation in modulating the biophysical properties of the epithelial plasma membrane. *J. Chem. Phys.* **2021**, *154*, 095101.
- (63) Sze, M. Y.; Gillams, R. J.; McLain, S. E.; Lorenz, C. D. Effects of lipid heterogeneity on model human brain lipid membranes. *Soft Matter* **2021**, *17*, 126–135.
- (64) Sonnino, S.; Mauri, L.; Chigorno, V.; Prinetti, A. Gangliosides as components of lipid membrane domains. *Glycobiology* **2007**, *17*, 1R–13R.
- (65) Prinetti, A.; Loberto, N.; Chigorno, V.; Sonnino, S. Glycosphingolipid behaviour in complex membranes. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 184–193.
- (66) Westerlund, B.; Slotte, J. P. How the molecular features of glycosphingolipids affect domain formation in fluid membranes. *Biochim. Biophys. Acta, Biomembr.* **2009**, *1788*, 194–201.
- (67) von Bülow, S.; Bullerjahn, J. T.; Hummer, G. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure. *J. Chem. Phys.* **2020**, *153*, 021101.
- (68) Seo, S.; Murata, M.; Shinoda, W. Pivotal Role of Interdigitation in Interleaflet Interactions: Implications from Molecular Dynamics Simulations. *J. Phys. Chem. Lett.* **2020**, *11*, 5171–5176.

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

(70) Melo, M. N.; Ingólfsson, H. I.; Marrink, S. J. Parameters for Martini sterols and hopanoids based on a virtual-site description. *J. Chem. Phys.* **2015**, *143*, 243152.

(71) Qi, Y.; Ingólfsson, H. I.; Cheng, X.; Lee, J.; Marrink, S. J.; Im, W. CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field. *J. Chem. Theory Comput.* **2015**, *11*, 4486–4494.

(72) 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.

(73) Fan, S.; Linke, M.; Paraskevatos, I.; Gowers, R.; Gecht, M.; Beckstein, O. Rostrup. In *Proceedings of the 18th Python in Science Conference*; Calloway, C., Lippa, D., Niederhut, D., Shupe, D., Eds.; SciPy: Austin, Texas, 2019, July 8–14.

(74) Rocklin, M. Dask: Parallel computation with blocked algorithms and task scheduling. In *Proceedings of the 14th Python in Science Conference*; Huff, K., Bergstra, J., Eds.; SciPy: Austin, Texas, 2015, July 6–12.

(75) Jarrod Millman, K.; Pérez, F. *Implementing Reproducible Research*; CRC Press, 2014; pp 149–183.

(76) Flow, G. The best way to use Git and GitHub. <https://githubflow.github.io/> (accessed April 13, 2021).

(77) Krekel, H.; Oliveira, B.; Pfannschmidt, R.; Bruynooghe, F.; Laughner, B.; Bruhin, F. pytest 6.2.2. 2004, <https://github.com/pytest-dev/pytest> (accessed April 13, 2021).

(78) pytest-cov development team. pytest-cov 2.11.1. 2006, <https://github.com/pytest-dev/pytest-cov> (accessed April 13, 2021).

(79) Anaconda, Anaconda Software Distribution. <https://anaconda.com> (accessed April 13, 2021).

(80) PyPA. pip—The Python Package Installer. <https://pip.pypa.io/en/stable/> (accessed April 13, 2021).



ACS Institute  
Learn. Develop. Excel.

**A comprehensive and authoritative learning platform supporting the broad chemistry community across the spectrum of learning.**

[institute.acs.org](https://institute.acs.org)

American Chemical Society



## Chapter 6

# Conclusions

The aim of this thesis was to provide new insight into the molecular origin of emergent phenomena in lipid membranes. Through all-atom and coarse-grained MD simulations, I have studied the process of phase separation in model membranes and the interaction of cholesterol and sphingomyelin in equimolar binary mixtures. The latter study provided a possible explanation for cholesterol preferentially mixing with sphingolipids over glycerophospholipids. This preferential mixing is important in the formation of biological lipid rafts — structures that are essential for many cellular processes. In addition to the findings reported in this thesis, the novel analysis methods — such as the conformational clustering of lipid pairs — and software developed in this thesis will be made openly available for use by other researchers.

In Chapter 3, I used coarse-grained simulations to study the effect of cholesterol oxidation on domain formation in model membranes. I found that domain formation may be disrupted by three mechanisms: i) a broader distribution of orientations of the sterol, which disrupts the packing of neighbouring lipids, ii) a decrease in the relative affinity of the sterol for saturated phospholipids over unsaturated phospholipids leads to less-complete phase separation, and iii) a decrease in the rate of translocation of the sterol across the membrane may impact domain registration in more complex mixtures. These findings corroborate previous simulation studies<sup>60,175,176</sup> and provide a molecular level description of domain formation disruption observed experimentally.<sup>178</sup>

However, Chapter 3 also provided a reminder that even when findings of MD simulations seem correct, and align with previous experimental and simulation studies, there

may be underlying issues with the force field that render the results unreliable. The results presented in Chapter 3 are unreliable due to a lack of conservation of energy that arises when simulating the MARTINI model of cholesterol with the default LINCS parameters in GROMACS. Thallmair et al. have recently shown that the breaking of conservation of energy is not unique to the MARTINI model of cholesterol.<sup>314</sup> Simulating any model that contains virtual sites using the default LINCS parameters of GROMACS will cause energy to be drained from the system via the virtual sites.

In Chapter 4, I used all-atom MD simulations to study cholesterol-sphingomyelin interactions in an equimolar mixture of cholesterol and PSM. SAXS and WAXS measurements of this binary mixture identified two coexisting bilayer structures, which are differentiated by their thicknesses. From the MD simulations, I provided a detailed description of the two coexisting bilayers. The thicker bilayer is enriched in PSM, depleted in cholesterol, and is characterised by an intermolecular hydrogen bond network formed by neighbouring PSM molecules. Further, the acyl tails of PSM molecules in the thicker bilayer are, on average, more extended, more ordered, and interdigitate less with the opposing leaflet.

Chapter 4 also illustrates that lipid-lipid interactions can be studied via an unsupervised clustering the conformations of neighbouring lipid pairs. Using this approach, I uncovered a potential mechanism by which cholesterol preferentially mixes with sphingolipids over glyceropholipids. This preferential mixing is important in the formation of lipid rafts in the cellular membrane, which themselves are essential for proper cellular functioning.

In Chapter 5, I described LiPyphilic — a Python package I have created for analysing MD simulations of lipid membranes. LiPyphilic provides analysis tools not available in other software, including the identification of sterol flip-flop events, the classification of local lipid environments, and the quantification of interleaflet registration. LiPyphilic also contains two ‘on-the-fly’ transformations to i) repair membranes that are split across periodic boundaries and ii) perform a nojump unwrapping of atomic coordinates. By having these tools in a well-tested and easy to install framework, other researchers no longer need to write in-house scripts for these analyses. Further, the tools can be used

with confidence in the accuracy of the results.

The nojump unwrapping transformation implemented in LiPyphilic will be of particular use to the entire MD community interested in the self-diffusion of any molecular species. Before calculating the diffusion coefficient of a molecule, a nojump transformation must first be applied to the atomic coordinates. Currently, however, all other nojump unwrapping implementations fail to account for the box size fluctuations generated by barostats under the NPT ensemble. At small box sizes, this failure can lead to an overestimation of the calculated diffusion coefficient.<sup>305</sup> As the box size increases, the fluctuations in box size become smaller compared to the size of the box, and thus the calculated diffusion coefficient converges toward the true value. This is why, for many years, it was thought that self-diffusion coefficients calculated from MD simulations depended on the dimensions of the system. However, LiPyphilic now provides a fast, simple means of correctly unwrapping atomic coordinates and thus obtaining a diffusion coefficient that does not depend on box size.

## Chapter 7

# Further Work

*‘Science is an ongoing process. It never ends. There is no single ultimate truth to be achieved, after which all the scientists can retire.’*

— Carl Sagan, ‘Cosmos’

As with many scientific endeavors, this thesis has raised more questions than it has answered. Below I will briefly discuss some potential avenues of further research that are of particular interest to me.

The domain formation process in DPPC:DOPC:Chol mixtures can not be studied using the MARTINI force field. This is because phase separation is not observed when the appropriate, more conservative, LINCS parameters are used. However, the SIRAH coarse-grained force-field was recently parameterised to correctly reproduce the phase behaviour of single- and multi-component lipid mixtures.<sup>35</sup> It has already been used to study the role of interdigitation in driving the registration of ordered domains in bilayers with asymmetric lipid compositions.<sup>7</sup> Reproducing the work in Chapter 3 using the SIRAH force field may provide more reliable insight into the disruption of domain formation caused by cholesterol oxidation. A further benefit of using the SIRAH force field is that a more biologically-relevant membrane composition could be used whilst still observing domain formation.<sup>7</sup> However, SIRAH does not employ the same building-block philosophy as MARTINI; the parameterisation of new molecules is non-trivial. Thus, atomistic simulations of 7-ketocholesterol would need to be performed, from which a



SIRAH model could be parameterised via the iterative Boltzmann inversion procedure.

In Chapter 4, I constructed hidden Markov models (HMM) based on local phosphate-phosphate distances to identify two coexisting bilayer structures in an equimolar mixture of cholesterol and PSM. These coexisting bilayer structures are possibly analogous to the coexisting raft-like and non-raft-like regions in biological membranes. As this approach to constructing the HMM is general, it could be used to study phase separation in more biologically-relevant lipid mixtures. This method of constructing HMMs also has potential benefits compared to previous approaches. Previously, hidden Markov models have been used to identify phase separated regions based on local lipid compositions,<sup>121,234,280</sup> with regions rich in cholesterol, sphingomyelin, and saturated lipids assumed to be raft-like. There are, however, many different types of raft-forming mixtures, and the lipid composition of raft and non-raft regions may not be known a priori. However, it is always the case that raft-like regions are thicker than their non-raft-like counterparts. As such, local phosphate-phosphate distances may prove to be a useful metric for determining the phase of a bilayer patch without biasing the results with assumptions about local lipid compositions.

The conformational clustering of lipid pairs developed in Chapter 4 provided insight into the distinct modes of interaction between sphingomyelin and cholesterol. Recently, Soloviov et al. used X-ray scattering measurements and MD simulations to show that transient lipid pairs form in the liquid-ordered phase.<sup>316</sup> The authors found that the energy of optical phonon modes between pairs of lipids correlates with the degree of phase separation in the membrane. It would be interesting to calculate the phonon mode energies between cholesterol and sphingomyelin for the six different cholesterol-sphingomyelin conformations identified in Chapter 4. The approach taken by Soloviov et al. would mean the energies are averaged over all six conformations. However, by calculating the energy of phonon modes for each conformation separately, we may gain a deeper understanding of how lipid-lipid interactions at the subnanometer level drive domain formation. For instance, we may find that one specific (energetically favourable) conformation is particularly important in driving the interaction between the two lipid species. This conformation might then act as a nucleation site for the formation of larger-

scale liquid-ordered structures.<sup>187</sup>

There are many ways in which LiPyphilic can be made more useful for researchers. I have already added indirect support for triclinic systems by creating an on-the-fly transformation to convert triclinic coordinates into their orthorhombic representation. There are also many more analyses that could be added to the package. For instance, I would like to add a tool to calculate the deuterium order parameter of lipid tails in all-atom simulations. Currently, many widely-used implementations do not properly account for the presence of C-C double bonds in unsaturated lipids.<sup>317</sup> In addition, I would like to add a tool that calculates the bending rigidity of a membrane from the distribution of lipid tilts and splays.<sup>318–320</sup>

There have been numerous Python packages for analysing lipid membrane simulations developed over the past year, including MLLPA,<sup>258</sup> ProLint,<sup>259</sup> PyLipID,<sup>260</sup> and LiPyphilic.<sup>192</sup> Together, these packages provide an excellent suite of tools for analysing MD simulations of lipid membranes. However, the long-term support of these packages is not guaranteed. All four of these packages were created by graduate students or post-docs. Whilst it is possible the respective research groups will take over the maintenance of a package, this is not an easy task — depending on the complexity of the software, it can take many months of full-time work for someone to become familiar with a new code base. The ideal solution is to create a community of users that also contribute to the maintenance and development of the software. However, this is very difficult to achieve, especially for new projects. MDAnalysis has recently launched the concept of MDAKits. These are Python packages, based on MDAnalysis, that provide more specialised analysis tools than those available in MDAnalysis. MDAKits are hosted by the MDAnalysis organisation, meaning they benefit from wide exposure to the large user-base of MDAnalysis. It may therefore be a good idea to consider requesting for LiPyphilic to become an MDAKit for lipid membrane analysis. More generally, however, the sustainability of scientific software is something that needs to be addressed by research funding bodies in order facilitate the long-term support of open-source software.

## **Appendix A**

# Supporting Information: Two coexisting membrane structures are defined by lateral and transbilayer interactions between sphingomyelin and cholesterol

Paul Smith,<sup>†</sup> Peter J. Quinn,<sup>\*,‡</sup> and Christian D. Lorenz<sup>\*,†</sup>

<sup>†</sup>*Department of Physics, King's College London, London, WC2R 2LS, UK*

<sup>‡</sup>*Institute of Pharmaceutical Science, King's College London, London SE1 9NH, UK*

E-mail: p.quinn@kcl.ac.uk; chris.lorenz@kcl.ac.uk

Pages: S1-S21

Figures: S1-S27

Tables: S1-S2

## CHARMM atom names

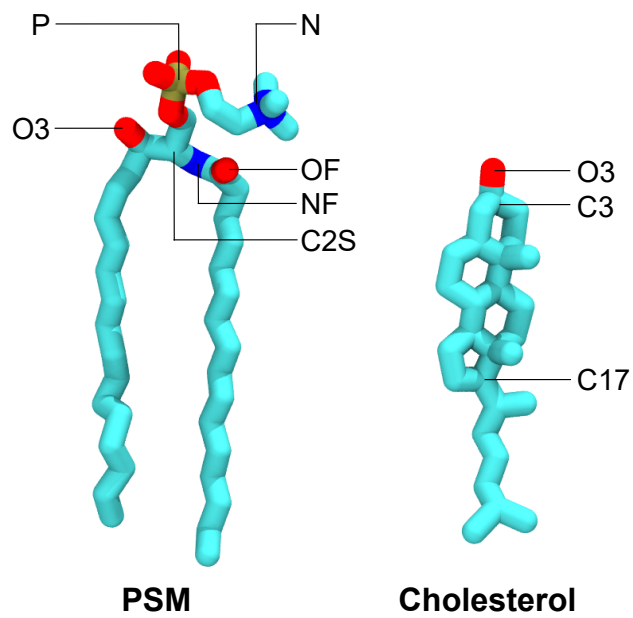


Figure S1: CHARMM atom names of specific atoms used in analysis of the simulations.

## Definition of measured angles

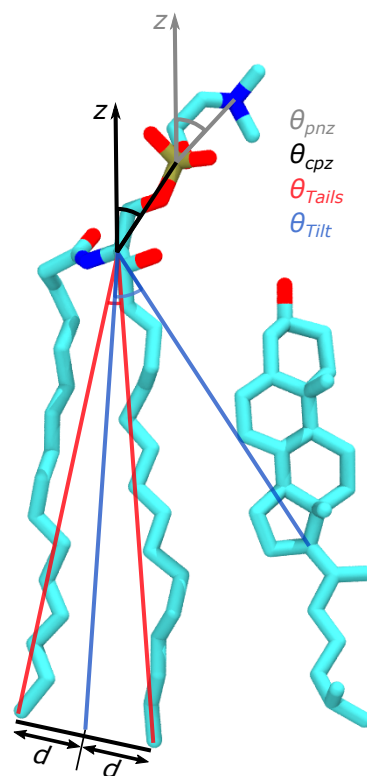


Figure S2: Definition of the angles used to describe the orientation of PSM and cholesterol within our simulated bilayer:  $\theta_{pnz}$ ,  $\theta_{cpz}$ ,  $\theta_{Tails}$  and  $\theta_{Tilt}$ .

## Definition of PSM and cholesterol faces

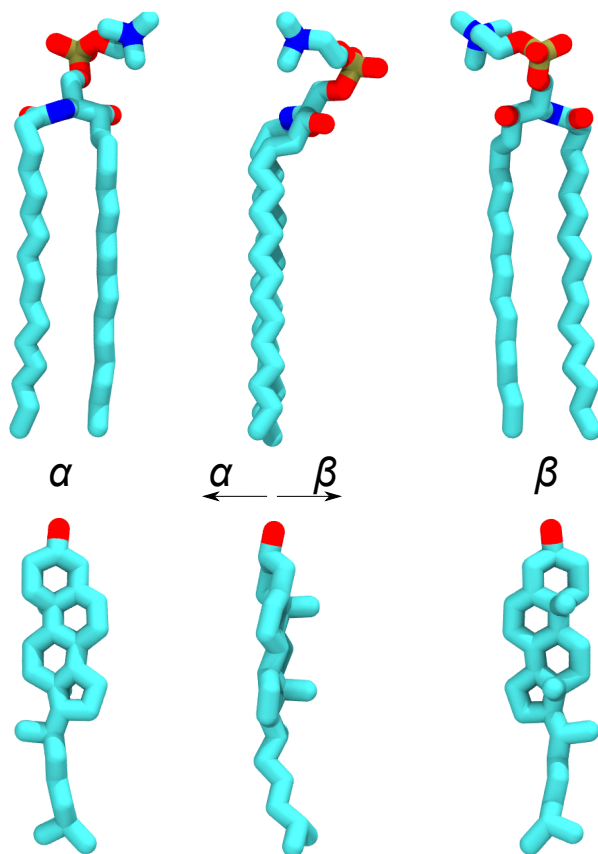


Figure S3: Definition of the  $\alpha$  and  $\beta$  faces of PSM and cholesterol. The carbonyl and hydroxyl groups of PSM protrude into its  $\beta$  face, in an analogous way to the methyl groups of cholesterol protruding into its own  $\beta$  face.

## Coupled Correlations

Property	$\rho$
$N_{HB}$	-0.01
Area	-0.08
Headgroup thickness	-0.01
SPH thickness	-0.00
FA thickness	-0.01
SPH Interdigitation	-0.18
FA Interdigitation	-0.17
SPH SCD	-0.00
FA SCD	0.00
$\theta_{pnz}$	-0.01
$\theta_{cpz}$	-0.01
$\theta_{Tails}$	0.01
Hydration	-0.01

Figure S4: Pearson correlation coefficient between the physical properties of PSM and those of its transbilayer couple.



## Interdigitation

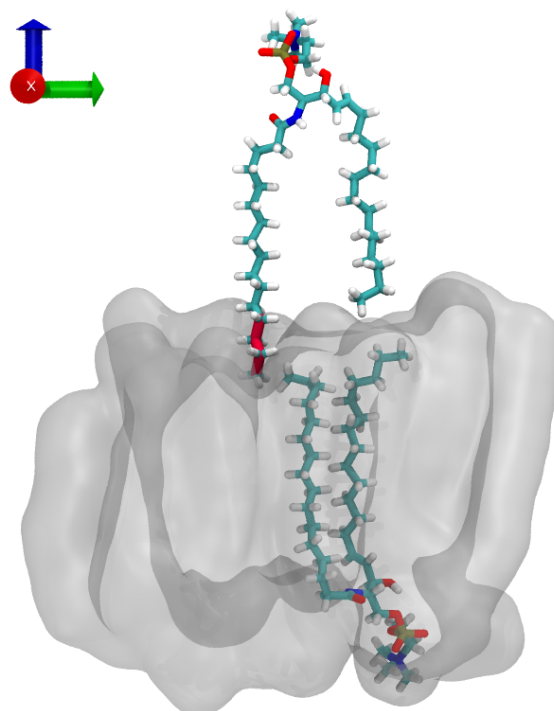


Figure S5: The FA tail of PSM in the upper leaflet penetrates into the surface formed by tails in the opposing leaflet. The tails of the coupled lipid (in the lower leaflet) are disordered, creating a void which the ordered FA tail in the reference lipid (upper leaflet) can fill.

The only property of PSM correlated with that of its couple is interdigitation (see Figure S4). Interdigitation of PSM occurs via disordered tails in one leaflet creating a void for more-ordered tails in the opposing leaflet to fill. To calculate the degree of interdigitation, we construct an intrinsic surface from the PSM tails of a given leaflet using a 2d histogram. Then for each PSM molecule in the opposing leaflet we find the maximum extent of penetration into this surface. We used 9 bins in each dimensions - that is, bin widths of approximately 10 Å.

## Bilayers Identified by Simulation

Table S1: Mean values of membrane thickness ( $\text{\AA}$ ) for  $B_1$  and  $B_2$ , defined as the C2S-C2S, P-P, and N-N distances in  $z$  between transbilayer coupled PSM molecules. The lower and upper bounds of the 95% confidence intervals are also shown. Confidence intervals were calculated via bootstrapping with 1,000 resamples.

	C2S			P			N		
	Lower	Mean	Upper	Lower	Mean	Upper	Lower	Mean	Upper
$B_1$	40.326	40.329	40.332	47.093	47.096	47.099	49.504	49.510	49.516
$B_2$	38.450	38.453	38.457	44.490	44.494	44.498	47.615	47.621	47.627

Table S2: Mean values of  $2r = d/\sin(\theta)$  (in  $\text{\AA}$ ) and 95% confidence intervals, calculated via bootstrapping with 1,000 resamples. The distance  $2r$  corresponds to the mean distance from the center of mass of the PSM hydrocarbon tails to the center of mass of a neighboring cholesterol molecule. We assume that two distinct mean distances (corresponding to  $WAXS_1$  and  $WAXS_2$ ) arise via either: i) PSM-cholesterol interactions in the two bilayers identified by simulations ( $B_1$  and  $B_2$ ); or ii) PSM interacting with the  $\alpha$  or  $\beta$  face of cholesterol ( $Chol \alpha$  or  $Chol \beta$ ). The latter assumption (ii) provides closer agreement with  $WAXS_1$  and  $WAXS_2$ . The values of  $2r$  for  $Chol \alpha$  and  $Chol \beta$  are statistically distinct with  $p < 0.05$ . The difference between  $Chol \alpha$  and  $Chol \beta$  is on the order of  $1 \times 10^{-1} \text{\AA}$ , whilst their 95% confidence intervals are on the order of  $1 \times 10^{-2} \text{\AA}$ .

	Lower	Mean	Upper
$B_1$	7.289	7.296	7.304
$B_2$	7.182	7.190	7.198
$Chol \alpha$	7.344	7.352	7.360
$Chol \beta$	7.134	7.142	7.149

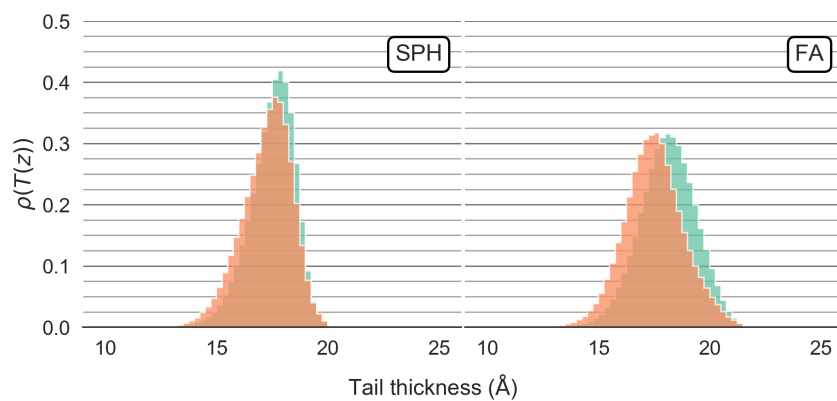


Figure S6: Distribution of the SPH and FA tail thicknesses ( $T(z)$ ) for lipids in  $B_1$  (green) and  $B_2$  (orange).  $T(z)$  is defined as the maximum extent in  $z$  of the heavy (non-Hydrogen) atoms in a given tail.

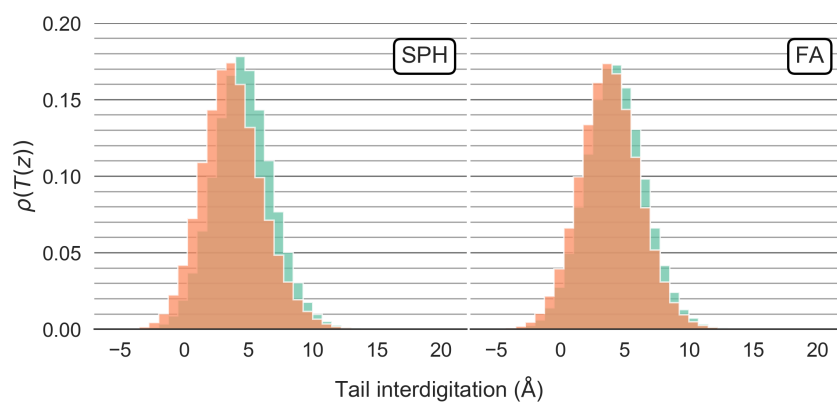


Figure S7: Distribution of the degree of tail interdigitation ( $T(z)$ ) for the SPH and FA tails of PSM in  $B_1$  (green) and  $B_2$  (orange). Negative values indicate penetration of the tail into the opposing leaflet.

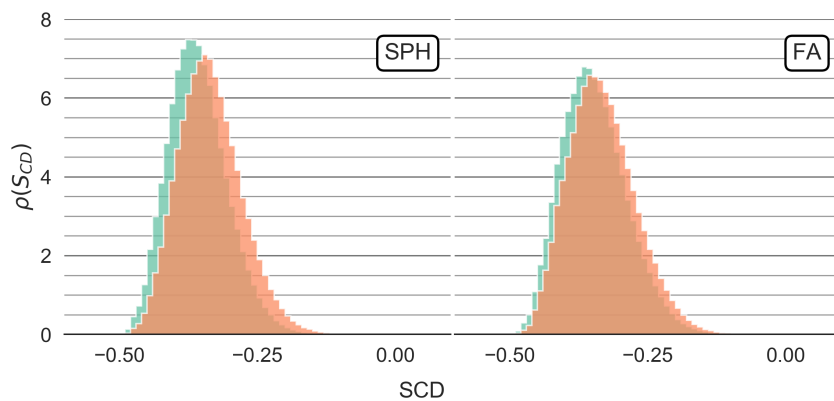


Figure S8: Distributions of the mean deuterium order parameter ( $S_{CD}$ ) for the SPH and FA tails of the PSM in  $B_1$  (green) and  $B_2$  (orange).

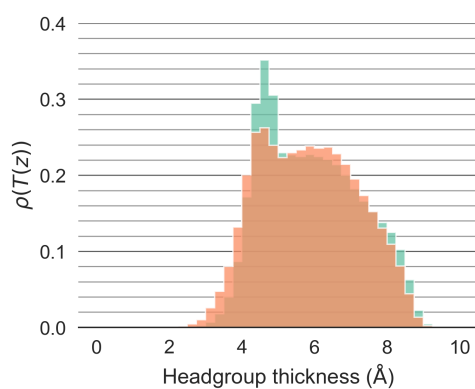


Figure S9: Distribution of the headgroup thickness ( $T(z)$ ) of lipids in  $B_1$  (green) and  $B_2$  (orange).  $T(z)$  is defined as the maximum extent in  $z$  of the heavy (non-Hydrogen) atoms not in the SPH or FA tails.

Property	$\rho$
$N_{HB}$	-0.21
$N_{HB}$ PSM : Chol	-0.22
$N_{HB}$ PSM : PSM	-0.07
Area	0.07
Headgroup thickness	0.03
SPH thickness	0.16
FA thickness	0.28
SPH Interdigitation	0.23
FA Interdigitation	0.13
SPH SCD	-0.25
FA SCD	-0.17
$\theta_{pnz}$	0.16
$\theta_{cpz}$	-0.40
$\theta_{Tails}$	-0.05
Hydration	0.18
Insertion	0.45

Figure S10: Pearson correlation coefficient between the physical properties of PSM and the membrane thickness (defined by the local phosphate-phosphate distance).

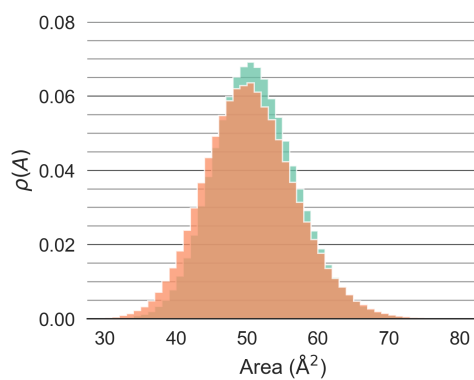


Figure S11: Distribution of the area ( $A$ ) of PSM in  $B_1$  (green) and  $B_2$  (orange) as calculated via a Voronoi tessellation of each leaflet.

## Hydrogen Bonds

In our discussion on the two bilayers in the main text, we reported a large difference between the  $\theta_{cpz}$  and  $\theta_{Tails}$  distributions, as well as the number of cholesterol-PSM phosphate (CP) and cholesterol-PSM hydroxyl (CH) hydrogen bonds. Here we see that these two hydrogen bonds have a large influence on the  $\theta_{cpz}$  angle, whilst the presence of the PSM amide-cholesterol (AC) hydrogen bond distorts the distribution of  $\theta_{Tails}$ . We also see that these two hydrogen bonds, CP and CH, contribute significantly to the difference in P-P distributions for  $B_1$  and  $B_2$ , whilst AC and cholesterol-PSM carbonyl (CE) have very little affect on membrane thickness.

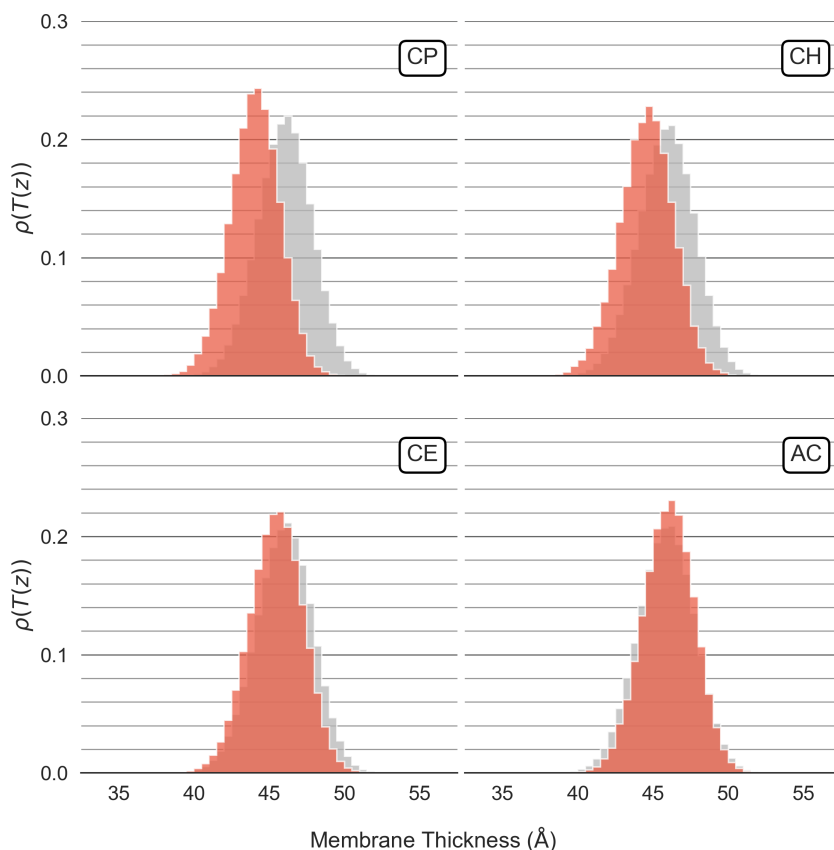


Figure S12: Distribution of P-P distances for lipids that are hydrogen bonded (red) or not (grey), for each type of cholesterol-PSM hydrogen bond.

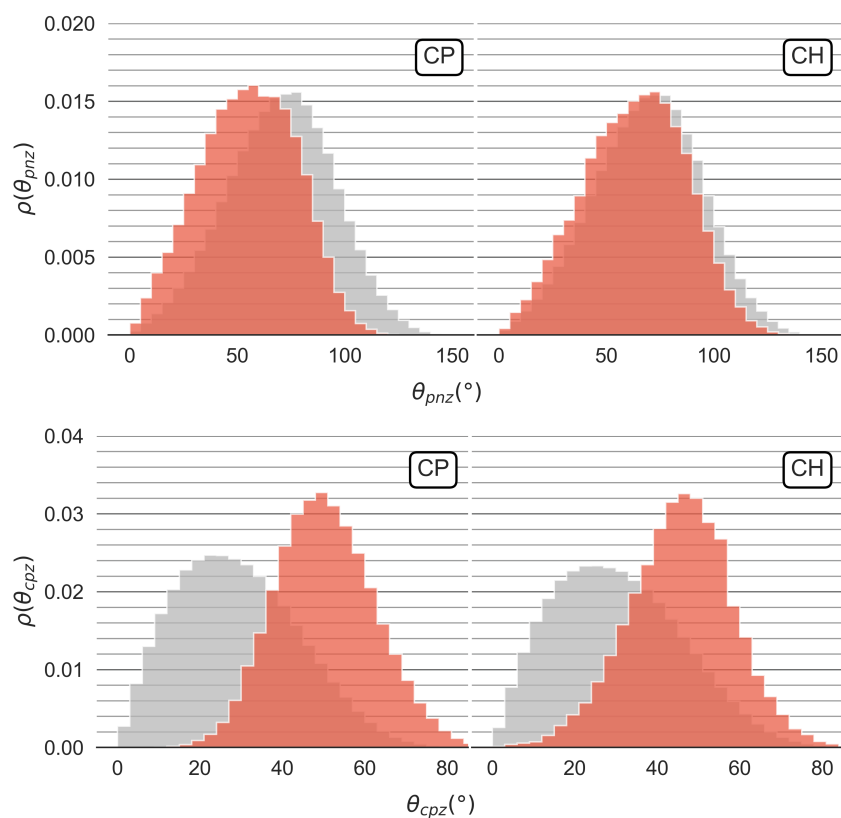


Figure S13: PSM PN headgroup angle ( $\theta_{pnz}$ ) and C2S-P angle ( $\theta_{cpz}$ ) for lipids that are hydrogen bonded (red) or not (grey), for cholesterol-PSM phosphate (CP) and cholesterol-PSM hydroxyl (CH) hydrogen bonds.



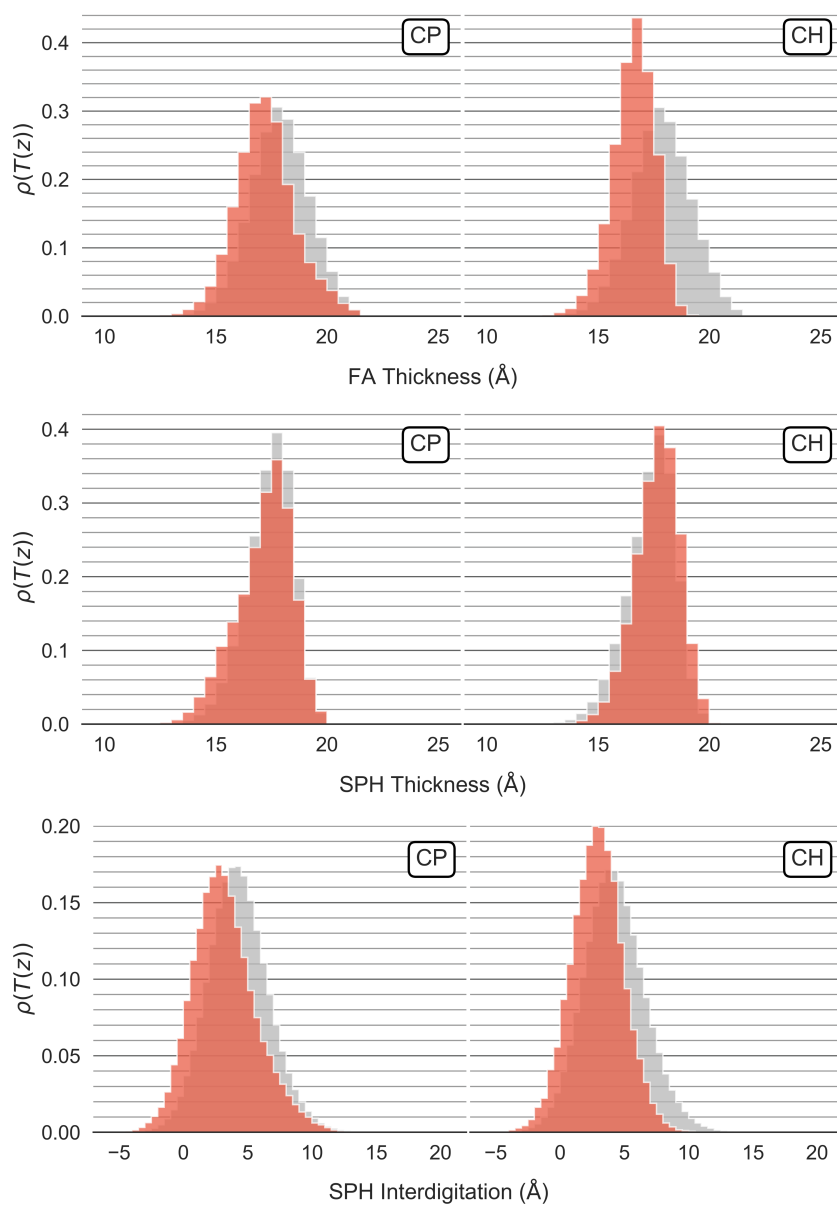


Figure S14: N-linked fatty acid (FA) thickness, sphingosine (SPH) thickness and SPH interdigitation for lipids that are hydrogen bonded (red) or not (grey), for cholesterol-PSM phosphate (CP) and cholesterol-PSM hydroxyl (CH) hydrogen bonds.

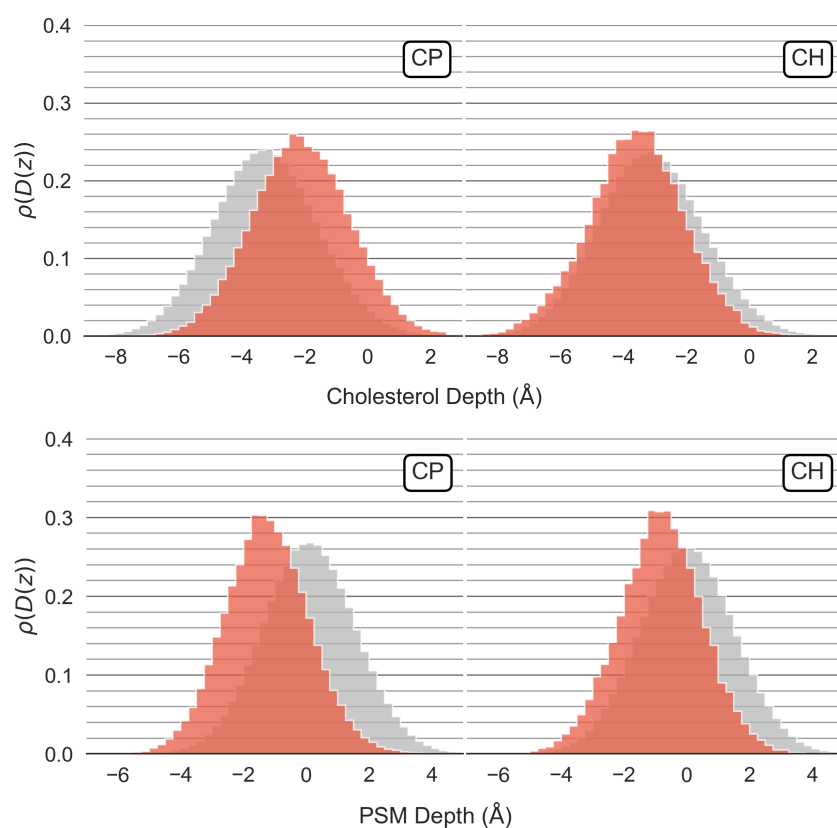


Figure S15: Cholesterol and PSM insertion depth ( $\text{\AA}$ ) for molecules that are hydrogen bonded (red) or not (grey), for cholesterol-PSM phosphate (CP) and cholesterol-PSM hydroxyl (CH) hydrogen bonds.

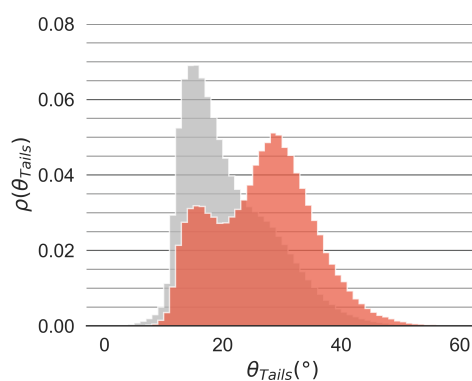


Figure S16: Distribution of  $\theta_{Tails}$  for PSM with (red) and without (grey) PSM amide-cholesterol hydrogen bonds.

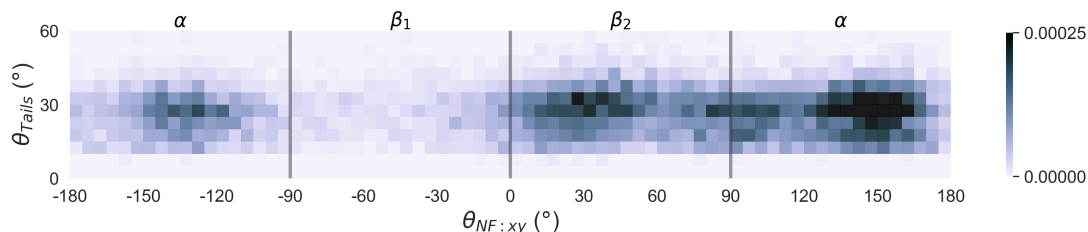


Figure S17: Distribution of  $\theta_{Tails}$  for PSM with amide-cholesterol hydrogen bonds, as a function of the orientation of the PSM around cholesterol. The orientation is defined as the angle between the vector made from the center of mass of cholesterol to the NF atom of PSM, and the  $y$ -axis in the  $xy$ -plane, where the line  $y|_{x=0}$  bisects the  $\beta$  face of cholesterol.

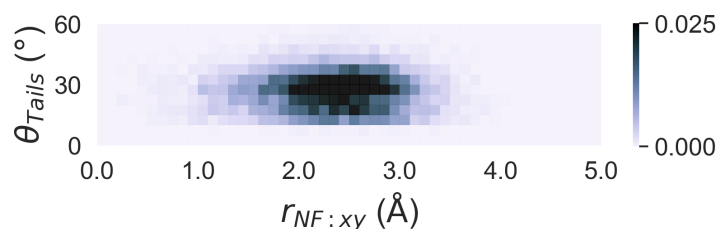


Figure S18: Distribution of  $\theta_{Tails}$  for PSM with amide-cholesterol hydrogen bonds, as a function of the distance in  $xy$  of PSM from cholesterol. The  $xy$  distance is calculated from the center of mass of cholesterol to the NF atom of PSM.

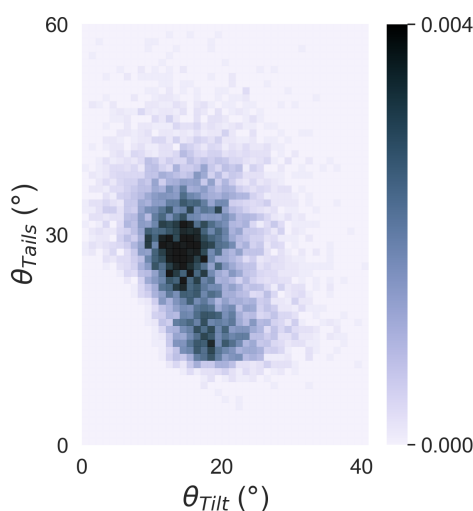


Figure S19: Distribution of  $\theta_{Tails}$  for PSM with amide-cholesterol hydrogen bonds, as a function of the relative tilt angle between PSM and cholesterol. The tilt angle is defined as the angle made by the center of mass of the PSM tails, the PSM C2S atom, and the cholesterol C17 atom (see Figure S2).

### Lateral distribution of PSM around cholesterol in the two bilayers

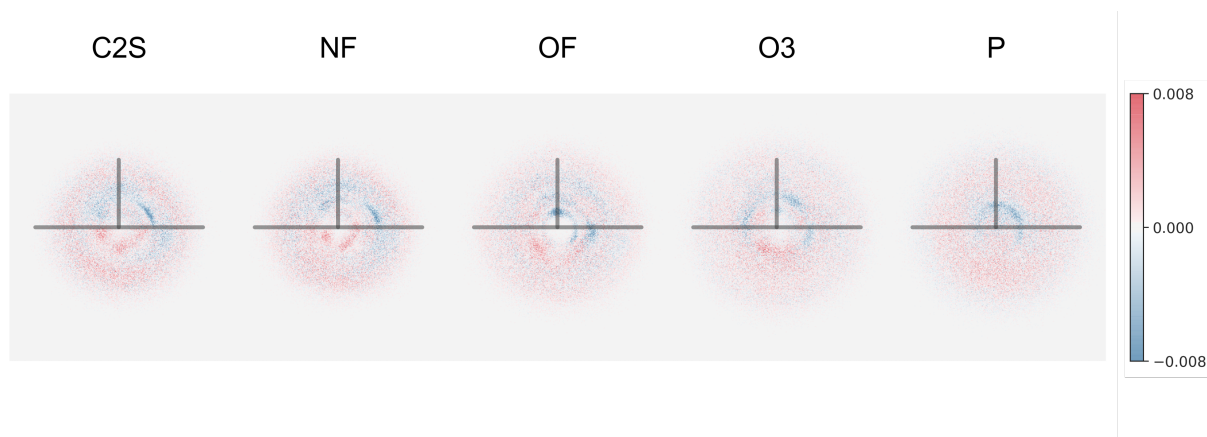


Figure S20: Difference in the distribution of PSM atoms around a neighboring cholesterol molecule between  $B_1$  and  $B_2$ . The horizontal grey bars represent  $25 \text{ \AA}$  and the vertical bars represent  $10 \text{ \AA}$ . See Figure S3 for a definition of the  $\alpha$  and  $\beta$  faces of cholesterol. Positive values (red) indicate increased density in  $B_1$ .

### Bilayers Identified by Simulation: Effect on Cholesterol

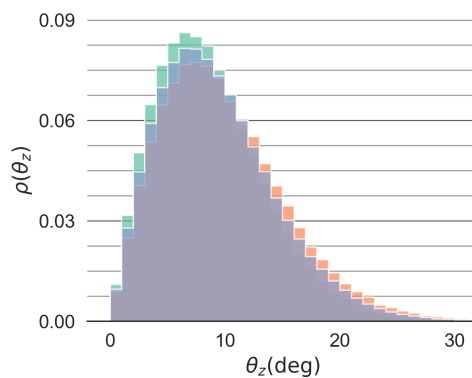


Figure S21: Distribution of the cholesterol tilt angle ( $\theta_z$ ) for the cholesterol that neighbors only  $B_1$  (green), only  $B_2$  (orange) or both  $B_1$  and  $B_2$  (blue).

## Clusters

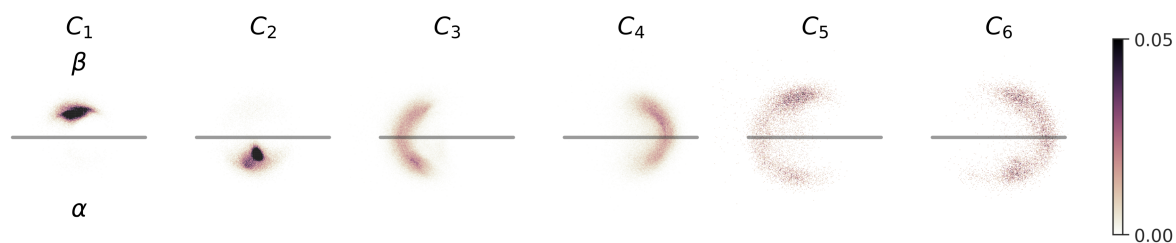


Figure S22: Distribution of the location of cholesterol around a neighboring PSM molecule for clusters  $C_1$  to  $C_6$ . Horizontal grey bars represent 25 Å. See Figure S3 for a definition of the  $\alpha$  and  $\beta$  faces of PSM. The  $\alpha$  and  $\beta$  faces are akin to those of cholesterol, with the roughness of the PSM  $\beta$  face due to the protrusion of the carbonyl and hydroxyl groups.

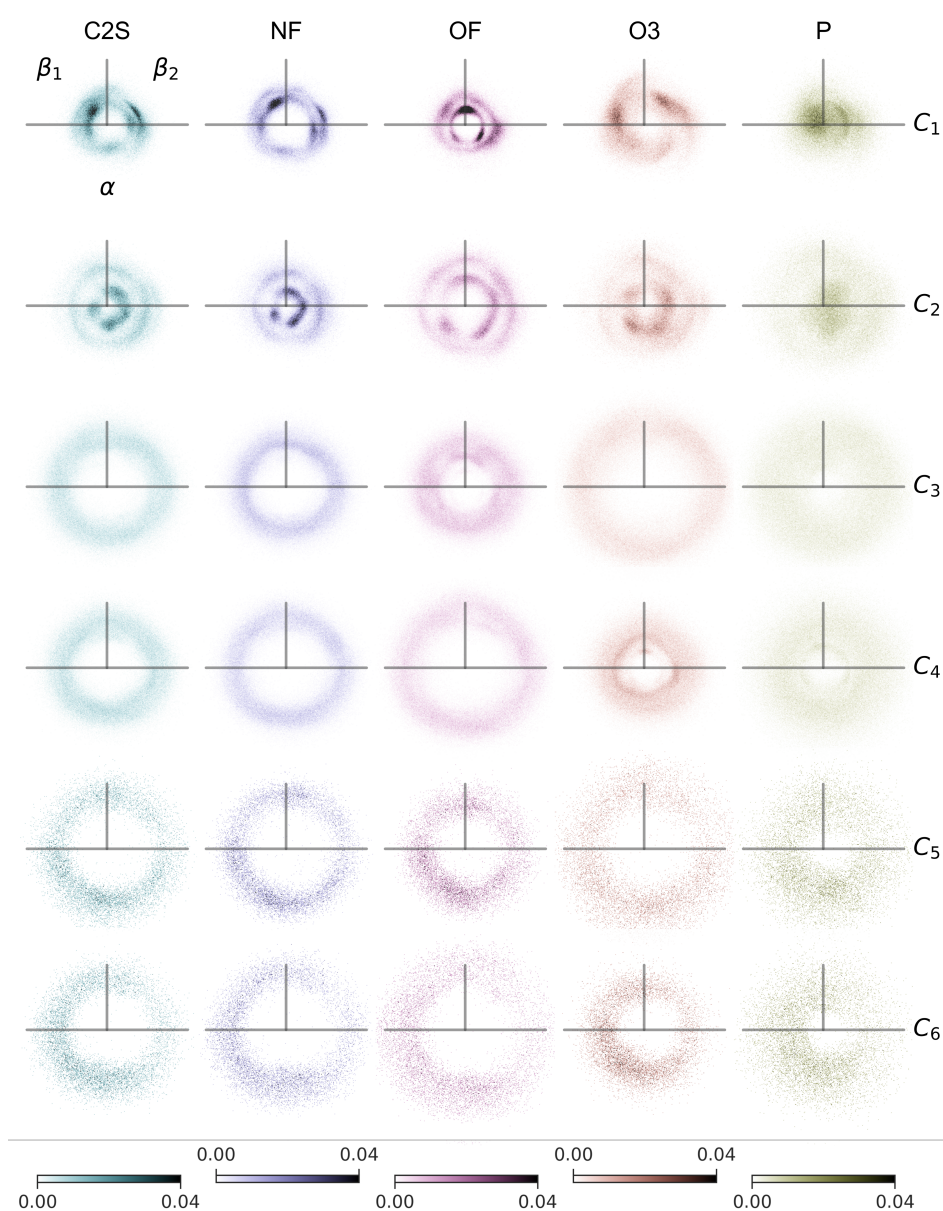


Figure S23: Distribution of PSM atoms around a neighboring cholesterol molecule for clusters  $C_1$  to  $C_6$ . Horizontal grey bars represent 25 Å and vertical represent 10 Å. See Figure S3 for a definition of the  $\alpha$  and  $\beta$  faces of cholesterol.

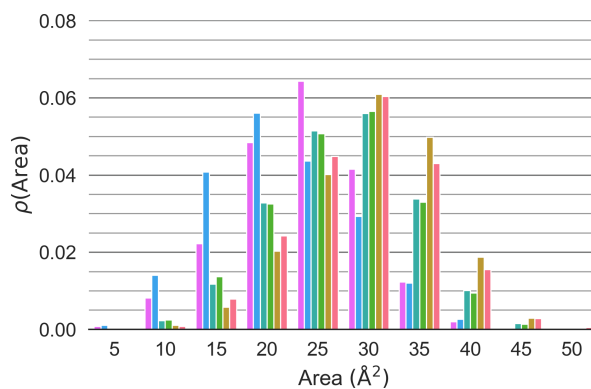


Figure S24: Distribution of the area per lipid for cholesterol in clusters  $C_1$  to  $C_6$ . Colors correspond to those in Figure 6.

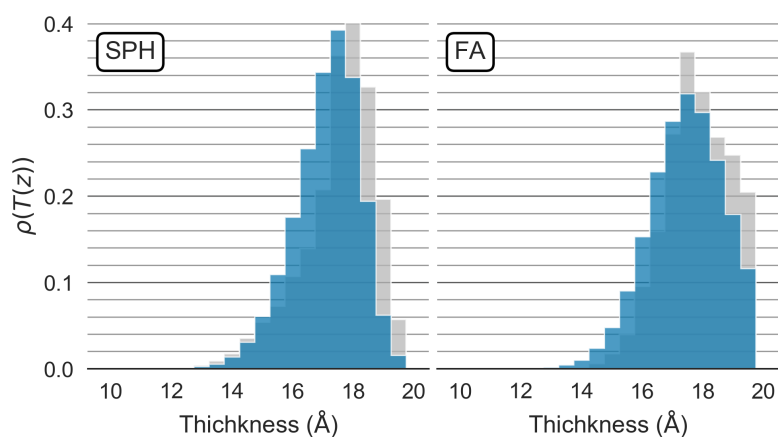


Figure S25: Distribution of the SPH and FA thickness of PSM neighboring cholesterol (blue) and PSM having no cholesterol neighbors (grey).

Hydrogen bonding from the amide group of PSM to cholesterol (AC) only occurs via PSM's  $\alpha$  face, whereas cholesterol-PSM carbonyl (CE) hydrogen bonding occurs only via the  $\beta$  face of PSM.

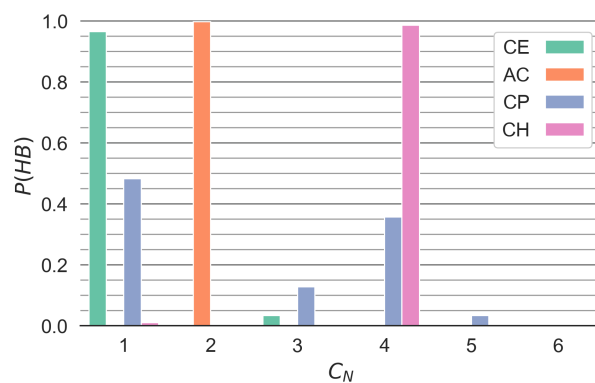


Figure S26: Fraction of a given hydrogen-bonded PSM-cholesterol pair belonging to each cluster,  $C_1$  to  $C_6$ , for each hydrogen bond type (CE: cholesterol-PSM carbonyl; AC: PSM amide-cholesterol; CP: cholesterol-PSM phosphate; and CH: cholesterol-PSM hydroxyl)

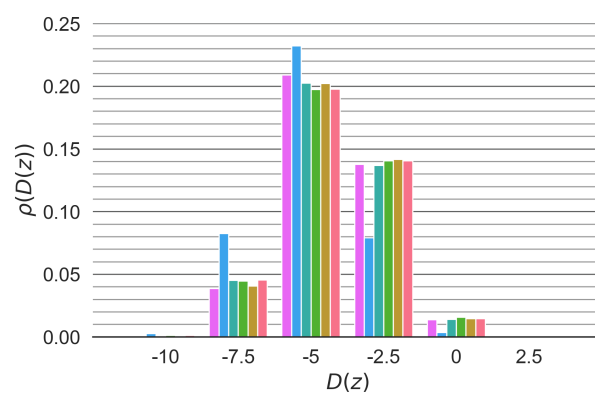


Figure S27: Distribution of the depth ( $D(z)$ ) at which cholesterol C3 atoms sit below the mean height of PSM C2S atoms in the corresponding leaflet, for cholesterol in clusters  $C_1$  to  $C_6$ . Colors correspond to those in Figure 6.



## **Appendix B**

## Supporting Information for ‘LiPyphilic: A Python toolkit for the analysis of lipid membrane simulations’

Table S1: Comparison of membrane analysis tools available in various software packages

Figure S1: Python script for calculating the lipid enrichment index based on tail saturation.

Section S1: Problems with nojump trajectory unwrapping

## Comparison with other software

Table S1: A comparison of the analysis tools available in LiPyphilic, FATSLiM,<sup>1</sup> and MLLPA.<sup>2</sup>

Analysis/transformation	LiPyphilic	FATSLiM	MLLPA
Leaflet identification	✓	✓	✓
Flip-flop	✓	×	×
Interleaflet registration	✓†	×	×
Largest cluster	✓	×	×
Local lipid environment	✓	×	✓
Enrichment index	✓	×	×
Bilayer thickness	✓†	✓	×
Lipid thickness	✓†	×	×
Lipid height	✓†	×	×
Lipid orientation	✓†	×	×
Area per lipid	✓†	✓	✓
Coarse-grained order parameter	✓	×	×
Unwrap membrane	✓	×	×
NoJump unwrapping	✓	×	×
Plotting utilities	✓	×	×
Lipid phase identification	×	×	✓

<sup>†</sup> Planar bilayers only

We have not included a comparison with PyLipID<sup>3</sup> or ProLint<sup>4</sup> — these packages contain a lot of useful tools for the analysis of lipid-protein interactions, but have little overlap with LiPyphilic in terms of functionality.

```

1 # Determine the saturation of each individual lipid
2 saturation = []
3 membrane = u.select_atoms("name GL1 GL2 AM1 AM2 ROH")
4
5 for lipid in np.unique(membrane.resnames):
6
7     lipid_residues = membrane.residues.atoms.select_atoms(f"resname {lipid}").residues
8     lipid_atoms = lipid_residues[0].atoms
9
10    # It is cholesterol if it has an ROH bead
11    if "ROH" in lipid_atoms.names:
12        saturation.append("C")
13
14    else:
15
16        num_doulbe = sum([True for name in lipid_atoms.names if name.startswith("D")])
17
18        # It is saturated if it has no D beads
19        if num_doulbe == 0:
20            saturation.append("S")
21
22        # It is monounsaturated if it has 1 D bead
23        elif num_doulbe == 1:
24            saturation.append("M")
25
26        # It is polyunsaturated if it has 2+ D beads
27        else:
28            saturation.append("P")
29
30 saturation = np.asarray(saturation)
31
32 # Now create a two-dimensional NumPy array where each lipid is labelled by its saturation
33 count_by = np.full((membrane.n_residues, neighbours.n_frames), fill_value="", dtype=str)
34 for lipid, sat in zip(np.unique(membrane.resnames), saturation):
35
36     lipid_indices = membrane.residues.resnames == lipid
37     count_by[lipid_indices, :] = sat
38
39 # Finally, count the neighbours based on tail saturation
40 counts, enrichment = neighbours.count_neighbours(
41     count_by=count_by,
42     return_enrichment=True
43 )

```

Figure S1: Workflow for calculating the enrichment index of lipids in the neuronal plasma membrane<sup>5</sup> based on their degree of tail saturation. This assumes the neighbour adjacency matrix has already been constructed as shown in Figure 5A.

## More trajectory unwrapping problems

The unwrapping scheme described in the main text, and previously by von Bulow *et al.*,<sup>6</sup> correctly accounts for the fluctuating box size in the NPT ensemble. However, it is only accurate in the case where coordinates are stored every timestep. In fact, it is impossible to correctly unwrap coordinates unless we store them at every timestep. This logically follows from the same argument made in the main text - to correctly unwrap coordinates we must know the length of the box at the timestep at which the jump occurred.

This can be seen when considering Figure 9, and assuming that coordinates are written every other frame. In this case, the coordinates are only known at timesteps  $N = \{0, 2\}$  and so:

$$x_N^u = x_N^w + \sum_{n=0}^{N/2} L_{2n} c_{2n}$$

where  $x_N^w$  is the wrapped position of the particle at frame  $N$ , and  $\sum_{n=0}^{N/2} L_{2n} c_{2n}$  accounts for the displacement that results from all jumps across periodic boundaries from frame 0 to frame  $N$ , determined using coordinates stored *every other* frame.  $L_{2n}$  is the box length at frame  $2n$  and:

$$c_{2n} = \begin{cases} -1 & x_{2n}^w - x_{2n-2}^w > L_{2n}/2 \\ 1 & x_{2n}^w - x_{2n-2}^w < -L_{2n}/2 \\ 0 & \text{otherwise} \end{cases}$$

In this example, using the atomic coordinates in Figure [9](#):

$$\begin{aligned}
x_2^u &= x_2^w - \sum_{n=0}^{N/2} L_{2n} c_{2n} \\
&= 7.5 - 8.0 \\
&= -0.5
\end{aligned}$$

which would give a displacement from frame 0 to frame 2 of  $-1$ , rather than  $-2$ . Therefore, the length of the box must be known at the frame at which the jump actually occurred. If this information is not known, then artifacts may be introduced into the unwrapping of atomic coordinates, even using the unwrapping scheme described here and by von Bulow *et al.*<sup>6</sup>

## References

- (1) Buchoux, S. FATSLiM: a fast and robust software to analyze MD simulations of membranes. *Bioinformatics* **2016**, *33*, 133-134.
- (2) Walter, V.; Ruscher, C.; Benzerara, O.; Thalmann, F. A Machine Learning-assisted Python module to study phase-specific events in lipid membranes. *J. Comput. Chem.* **2021**, *42*, 930-943.
- (3) Song, W.; Corey, R. A.; Ansell, B.; Cassidy, K.; Horrell, M.; Duncan, A.; Stansfeld, P. J.; Sansom, M. PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations. *bioRxiv* **2021**, 2021.07.14.452312..
- (4) Sejdiu, B. I.; Tieleman, D. P. ProLint: a web-based framework for the automated data analysis and visualization of lipid-protein interactions. *Nucleic Acids Res.* **2021**, *49*, W544–W550.
- (5) Ingólfsson, H. I.; Carpenter, T. S.; Bhatia, H.; Bremer, P. T.; Marrink, S. J.; Lightstone, F. C. Computational Lipidomics of the Neuronal Plasma Membrane. *Biophys. J.* **2017**, *113*, 2271–2280.
- (6) Von Bülow, S.; Bullerjahn, J. T.; Hummer, G. Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure. *J. Chem. Phys.* **2020**, *153*, 021101

# Bibliography

- [1] Siewert J Marrink, Valentina Corradi, Paulo CT Souza, Helgi I Ingoíffson, D Peter Tieleman, and Mark SP Sansom. “Computational Modeling of Realistic Cell Membranes”. *Chem. Rev.* 119.9 (2019), pp. 6184–6226. DOI: 10.1021/acs.chemrev.8b00460.
- [2] Bruno Pontes, Yareni Ayala, Anna Carolina C. Fonseca, Luciana F. Romão, Raccele F. Amaral, Leonardo T. Salgado, Flavia R. Lima, Marcos Farina, Nathan B. Viana, Vivaldo Moura-Neto, and H. Moysés Nussenzveig. “Membrane Elastic Properties and Cell Function”. *PLOS One* 8.7 (July 2013), pp. 1–13. DOI: 10.1371/journal.pone.0067708.
- [3] Zheng Shi, Zachary T. Graber, Tobias Baumgart, Howard A. Stone, and Adam E. Cohen. “Cell Membranes Resist Flow”. *Cell* 175.7 (2018), 1769–1779.e13. DOI: <https://doi.org/10.1016/j.cell.2018.09.054>.
- [4] J.H. Lorent, K.R. Levental, L. Ganesan, G. Rivera-Longsworth, E. Sezgin, M. Doktorova, E. Lyman, and I. Levental. “Plasma membranes are asymmetric in lipid unsaturation, packing and protein shape”. *Nat. Chem. Biol.* 16.6 (June 2020), pp. 644–652. DOI: 10.1038/s41589-020-0529-6.
- [5] Henri F. Avela and Heli Sirén. “Advances in lipidomics”. *Clin. Chim. Acta* 510 (2020), pp. 123–141. DOI: <https://doi.org/10.1016/j.cca.2020.06.049>.
- [6] Gerhard Liebisch, Juan Antonio Vizcaíno, Harald Köfeler, Martin Trötz Müller, William J. Griffiths, Gerd Schmitz, Friedrich Spener, and Michael J.O. Wakelam. “Shorthand notation for lipid structures derived from mass spectrometry”. *J. Lipid Res.* 54.6 (June 2013), pp. 1523–1530. DOI: 10.1194/jlr.M033506.



- [7] Sangjae Seo, Michio Murata, and Wataru Shinoda. “Pivotal Role of Interdigitation in Interleaflet Interactions: Implications from Molecular Dynamics Simulations”. *J. Phys. Chem. Lett* 11.13 (July 2020), pp. 5171–5176. DOI: 10.1021/acs.jpcclett.0c01317.
- [8] Edgar Eduard Kooijman and Koert N.J. Burger. “Biophysics and function of phosphatidic acid: A molecular perspective”. *Biochim Biophys Acta Mol Cell Biol Lipids* 1791.9 (2009). Phospholipase D, pp. 881–888. DOI: <https://doi.org/10.1016/j.bbalip.2009.04.001>.
- [9] Tomohiro Kimura, William Jennings, and Richard M. Epand. “Roles of specific lipid species in the cell and their molecular mechanism”. *Prog. Lipid Res.* 62 (2016), pp. 75–92. DOI: <https://doi.org/10.1016/j.plipres.2016.02.001>.
- [10] Marianna Agassandian and Rama K. Mallampalli. “Surfactant phospholipid metabolism”. *Biochim Biophys Acta Mol Cell Biol Lipids* 1831.3 (2013). Phospholipids and phospholipid metabolism, pp. 612–625. DOI: <https://doi.org/10.1016/j.bbalip.2012.09.010>.
- [11] Edwin J.A Veldhuizen and Henk P Haagsman. “Role of pulmonary surfactant components in surface film formation and dynamics”. *Biochim Biophys Acta Biomembr* 1467.2 (2000), pp. 255–270. DOI: [https://doi.org/10.1016/S0005-2736\(00\)00256-X](https://doi.org/10.1016/S0005-2736(00)00256-X).
- [12] Jason G. Kay and Gregory D. Fairn. “Distribution, dynamics and functional roles of phosphatidylserine within the cell”. *Cell Commun. Signal.* 17.1 (Oct. 2019), pp. 1–8. DOI: 10.1186/s12964-019-0438-z.
- [13] Jelske N. van der Veen, John P. Kennelly, Sereana Wan, Jean E. Vance, Dennis E. Vance, and René L. Jacobs. “The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease”. *Biochim Biophys Acta Biomembr* 1859.9, Part B (2017). Membrane Lipid Therapy: Drugs Targeting Biomembranes, pp. 1558–1572. DOI: <https://doi.org/10.1016/j.bbamem.2017.04.006>.

- [14] James P. Zewe, April M. Miller, Sahana Sangappa, Rachel C. Wills, Brady D. Goulden, and Gerald R.V. Hammond. “Probing the subcellular distribution of phosphatidylinositol reveals a surprising lack at the plasma membrane”. *J. Cell Biol.* 219.3 (Mar. 2020). DOI: 10.1083/JCB.201906127.
- [15] Zsafia Szentpetery, Peter Várnai, and Tamas Balla. “Acute manipulation of Golgi phosphoinositides to assess their importance in cellular trafficking and signaling”. *Proc. Natl. Acad. Sci. U.S.A* 107.18 (2010), pp. 8225–8230. DOI: 10.1073/pnas.1000157107.
- [16] Amanda Krajnik, Joseph A. Brazzo, Kalyanaraman Vaidyanathan, Tuhin Das, Javier Redondo-Muñoz, and Yongho Bae. “Phosphoinositide Signaling and Mechanotransduction in Cardiovascular Biology and Disease”. *Front. Cell Dev. Biol.* 8 (Dec. 2020), p. 1588. DOI: 10.3389/fcell.2020.595849.
- [17] Björn H. Falkenburger, Jill B. Jensen, Eamonn J. Dickson, Byung Chang Suh, and Bertil Hille. “Phosphoinositides: Lipid regulators of membrane proteins”. *J. Physiol. (Lond.)* 588.17 (Sept. 2010), pp. 3179–3185. DOI: 10.1113/jphysiol.2010.192153.
- [18] B S Robinson, D W Johnson, and A Poulos. “Novel molecular species of sphingomyelin containing 2-hydroxylated polyenoic very-long-chain fatty acids in mammalian testes and spermatozoa.” *J. Biol. Chem.* 267.3 (1992), pp. 1746–1751. DOI: [https://doi.org/10.1016/S0021-9258\(18\)46009-7](https://doi.org/10.1016/S0021-9258(18)46009-7).
- [19] Paul Smith, Dylan M. Owen, Christian D. Lorenz, and Maria Makarova. “Asymmetric glycerophospholipids impart distinctive biophysical properties to lipid bilayers”. *Biophys. J.* 120.9 (2021), pp. 1746–1754. DOI: <https://doi.org/10.1016/j.bpj.2021.02.046>.
- [20] Ren Sheng, Yong Chen, Heon Yung Gee, Ewa Stec, Heather R. Melowic, Nicholas R. Blatner, Moe P. Tun, Yonjung Kim, Morten Källberg, Takahiro K. Fujiwara, Ji Hye Hong, Kwang Pyo Kim, Hui Lu, Akihiro Kusumi, Min Goo Lee, and Wonhwa Cho. “Cholesterol modulates cell signaling and protein network-

- ing by specifically interacting with PDZ domain-containing scaffold proteins”. *Nat. Commun.* 3.1 (Dec. 2012), pp. 1–9. DOI: 10.1038/ncomms2221.
- [21] Frédéric de Meyer and Berend Smit. “Effect of cholesterol on the structure of a phospholipid bilayer”. *Proc. Natl. Acad. Sci. U.S.A* 106.10 (2009), pp. 3654–3658. DOI: 10.1073/pnas.0809959106.
- [22] Jie Hu, Zhonghua Zhang, Wen Jun Shen, and Salman Azhar. “Cellular cholesterol delivery, intracellular processing and utilization for biosynthesis of steroid hormones”. *Nutr. Metab.* 7.1 (June 2010), p. 47. DOI: 10.1186/1743-7075-7-47.
- [23] A.E. Co Wen and C.B. Campbell. “Bile Salt Metabolism:I. The Physiology of Bile Salts”. *Aust. Nz. J. Med.* 7.6 (1977), pp. 579–586. DOI: <https://doi.org/10.1111/j.1445-5994.1977.tb02312.x>.
- [24] Pablo V. Escribá, Andrés Ozaita, Catalina Ribas, Antonio Miralles, Elfrieda Fodor, Tibor Farkas, and Jesús A. García-Sevilla. “Role of lipid polymorphism in G protein-membrane interactions: Nonlamellar-prone phospholipids and peripheral protein binding to membranes”. *Proc. Natl. Acad. Sci. U.S.A* 94.21 (1997), pp. 11375–11380. DOI: 10.1073/pnas.94.21.11375.
- [25] Alex H. de Vries, Serge Yefimov, Alan E. Mark, and Siewert J. Marrink. “Molecular structure of the lecithin ripple phase”. *Proc. Natl. Acad. Sci. U.S.A* 102.15 (2005), pp. 5392–5396. DOI: 10.1073/pnas.0408249102.
- [26] J. Hjort Ipsen, G. Karlström, O.G. Mourtsen, H. Wennerström, and M.J. Zuckermann. “Phase equilibria in the phosphatidylcholine-cholesterol system”. *Biochim Biophys Acta Biomembr* 905.1 (Nov. 1987), pp. 162–172. DOI: 10.1016/0005-2736(87)90020-4.
- [27] Daniel Lingwood and Kai Simons. “Lipid rafts as a membrane-organizing principle”. *Science* 327.5961 (Jan. 2010), pp. 46–50. DOI: 10.1126/science.1174621.
- [28] Mary L. Kraft. “Plasma membrane organization and function: Moving past lipid rafts”. *Mol. Biol. Cell* 24.18 (Sept. 2013), pp. 2765–2768. DOI: 10.1091/mbc.E13-03-0165.

- [29] Chang Wang, Martin R. Krause, and Steven L. Regen. “Push and Pull Forces in Lipid Raft Formation: The Push Can Be as Important as the Pull”. *J. Am. Chem. Soc.* 137.2 (Jan. 2015), pp. 664–666. DOI: 10.1021/ja5115437.
- [30] Jonathan D. Nickels, Xiaolin Cheng, Barmak Mostofian, Christopher Stanley, Benjamin Lindner, Frederick A. Heberle, Stefania Perticaroli, Mikhail Feygen-son, Takeshi Egami, Robert F. Standaert, Jeremy C. Smith, Dean A.A. Myles, Michael Ohl, and John Katsaras. “Mechanical Properties of Nanoscopic Lipid Domains”. *J. Am. Chem. Soc.* 137.50 (Dec. 2015), pp. 15772–15780. DOI: 10.1021/jacs.5b08894.
- [31] Ilya Levental and Sarah L. Veatch. “The Continuing Mystery of Lipid Rafts”. *J. Mol. Biol.* 428.24 (Dec. 2016), pp. 4749–4764. DOI: 10.1016/J.JMB.2016.08.022.
- [32] Mélanie Carquin, Ludovic D’Auria, Hélène Pollet, Ernesto R Bongarzone, and Donatienne Tyteca. “Recent progress on lipid lateral heterogeneity in plasma membranes: From rafts to submicrometric domains”. *Prog. Lipid Res.* 62 (Apr. 2016), pp. 1–24. DOI: 10.1016/j.plipres.2015.12.004.
- [33] Mary L. Kraft. “Sphingolipid organization in the plasma membrane and the mechanisms that influence it”. *Front. Cell Dev. Biol.* 4 (Jan. 2017), p. 154. DOI: 10.3389/fcell.2016.00154.
- [34] Friederike Schmid. “Physical mechanisms of micro- and nanodomain formation in multicomponent lipid membranes”. *Biochim Biophys Acta Biomembr* 1859.4 (Apr. 2017), pp. 509–528. DOI: 10.1016/j.bbamem.2016.10.021.
- [35] Erdinc Sezgin, Ilya Levental, Satyajit Mayor, and Christian Eggeling. “The mystery of membrane organization: composition, regulation and roles of lipid rafts”. *Nat. Rev. Mol. Cell Bio.* 18 (2017), pp. 361–374. DOI: 10.1038/nrm.2017.16.
- [36] Erhard Bieberich. “Sphingolipids and lipid rafts: Novel concepts and methods of analysis”. *Chem. Phys. Lipids* 216 (Nov. 2018), pp. 114–131. DOI: 10.1016/j.chemphyslip.2018.08.003.
- [37] Diego Krapf. “Compartmentalization of the plasma membrane”. *Curr. Opin. Cell Biol.* 53 (Aug. 2018), pp. 15–21. DOI: 10.1016/J.CEB.2018.04.002.

- [38] Krishnan Raghunathan and Anne K. Kenworthy. “Dynamic pattern generation in cell membranes: Current insights into membrane organization”. *Biochim Biophys Acta Biomembr* 1860.10 (Oct. 2018), pp. 2018–2031. DOI: 10.1016/J.BBAMEM.2018.05.002.
- [39] Giray Enkavi, Matti Javanainen, Waldemar Kulig, Tomasz Róg, and Ilpo Vattulainen. “Multiscale Simulations of Biological Membranes: The Challenge To Understand Biological Phenomena in a Living Substance”. *Chem. Rev.* 119.9 (Mar. 2019), pp. 5607–5774. DOI: 10.1021/acs.chemrev.8b00538.
- [40] Ilya Levental, Hong-Yin Wang, and Deepti Bharti. “Membrane Heterogeneity Beyond the Plasma Membrane”. *Front. Cell Dev. Biol.* 8 (2020), p. 1186. DOI: 10.3389/FCELL.2020.580814.
- [41] Arnd Pralle. “Modulation and dynamics of cell membrane heterogeneities”. *Chem. Phys. Lipids* 233 (Nov. 2020), p. 105006. DOI: 10.1016/j.chemphyslip.2020.105006.
- [42] Alena Khmelinskaia, Joaquim M.T. Marquês, André E.P. Bastos, Catarina A.C. Antunes, Andreia Bento-Oliveira, Silvia Scolari, Gerson M. da S. Lobo, Rui Malhó, Andreas Herrmann, H. Susana Marinho, and Rodrigo F.M. de Almeida. “Liquid-Ordered Phase Formation by Mammalian and Yeast Sterols: A Common Feature With Organizational Differences”. *Front. Cell Dev. Biol.* 8 (June 2020), p. 337. DOI: 10.3389/fcell.2020.00337.
- [43] Kai Simons and Elina Ikonen. “Functional rafts in cell membranes”. *Nature* 387.6633 (June 1997), pp. 569–572. DOI: 10.1038/42408.
- [44] Abdul A. Waheed and Eric O. Freed. “Lipids and membrane microdomains in HIV-1 replication”. *Virus Res.* 143.2 (2009), pp. 162–176. DOI: <https://doi.org/10.1016/j.virusres.2009.04.007>.
- [45] Taras Sych, Cenk Onur Gurdap, Linda Wedemann, and Erdinc Sezgin. “How Does Liquid-Liquid Phase Separation in Model Membranes Reflect Cell Membrane Heterogeneity?” *Membranes* 11.5 (2021). DOI: 10.3390/membranes11050323.

- [46] Jun Fan, Maria Sammalkorpi, and Mikko Haataja. “Formation and regulation of lipid microdomains in cell membranes: Theory, modeling, and speculation”. *FEBS Lett* 584.9 (2010). *Frontiers in Membrane Biochemistry*, pp. 1678–1684. DOI: <https://doi.org/10.1016/j.febslet.2009.10.051>.
- [47] Sarah L Veatch and Sarah L Keller. “Separation of Liquid Phases in Giant Vesicles of Ternary Mixtures of Phospholipids and Cholesterol”. *Biophys. J.* 85 (2003), pp. 3074–3083.
- [48] Frederick A. Heberle, Jing Wu, Shih Lin Goh, Robin S. Petruzielo, and Gerald W. Feigenson. “Comparison of three ternary lipid bilayer mixtures: FRET and ESR reveal nanodomains”. *Biophys. J.* 99.10 (Nov. 2010), pp. 3309–3318. DOI: [10.1016/j.bpj.2010.09.064](https://doi.org/10.1016/j.bpj.2010.09.064).
- [49] Yu Chia Kao, Pei Chuan Ho, Yuan Kun Tu, I. Ming Jou, and Kuen Jer Tsai. “Lipids and alzheimer’s disease”. *Int. J. Mol. Sci.* 21.4 (Feb. 2020), p. 1505. DOI: [10.3390/ijms21041505](https://doi.org/10.3390/ijms21041505).
- [50] Saranna Fanning, Dennis Selkoe, and Ulf Dettmer. “Parkinson’s disease: proteinopathy or lipidopathy?” *npj Parkinsons Dis.* 6.1 (Dec. 2020), pp. 1–9. DOI: [10.1038/s41531-019-0103-7](https://doi.org/10.1038/s41531-019-0103-7).
- [51] Alexandrine During. “Osteoporosis: A role for lipids”. *Biochimie* 178 (Nov. 2020), pp. 49–55. DOI: [10.1016/j.biochi.2020.08.004](https://doi.org/10.1016/j.biochi.2020.08.004).
- [52] Valerio Chiurchiù, Alessandro Leuti, and Mauro Maccarrone. “Bioactive lipids and chronic inflammation: Managing the fire within”. *Front. Immunol.* 9.1 (Jan. 2018), p. 1. DOI: [10.3389/fimmu.2018.00038](https://doi.org/10.3389/fimmu.2018.00038).
- [53] Lisa M. Butler, Ylenia Perone, Jonas Dehairs, Leslie E. Lupien, Vincent de Laat, Ali Talebi, Massimo Loda, William B. Kinlaw, and Johannes V. Swinnen. “Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention”. *Adv. Drug Deliver. Rev.* 159 (Jan. 2020), pp. 245–293. DOI: [10.1016/j.addr.2020.07.013](https://doi.org/10.1016/j.addr.2020.07.013).

- [54] Ooiean Teng, Candice Ke En Ang, and Xue Li Guan. "Macrophage-bacteria interactions-A lipid-centric relationship". *Front. Immunol.* 8.12 (Dec. 2017), p. 1836. DOI: 10.3389/fimmu.2017.01836.
- [55] Mohamed Abu-Farha, Thangavel Alphonse Thanaraj, Mohammad G. Qad-doumi, Anwar Hashem, Jehad Abubaker, and Fahd Al-Mulla. "The role of lipid metabolism in COVID-19 virus infection and as a drug target". *Int. J. Mol. Sci.* 21.10 (May 2020), p. 3544. DOI: 10.3390/ijms21103544.
- [56] Antonella Rella, Amir M. Farnoud, and Maurizio Del Poeta. "Plasma membrane lipids and their role in fungal virulence". *Prog. Lipid Res.* 61 (Jan. 2016), pp. 63–72. DOI: 10.1016/j.plipres.2015.11.003.
- [57] Michael I. Bukrinsky, Nigora Mukhamedova, and Dmitri Sviridov. "Lipid rafts and pathogens: the art of deception and exploitation: Thematic Review Series: Biology of Lipid Rafts". *J. Lipid Res.* 61.5 (2020), pp. 601–610. DOI: <https://doi.org/10.1194/jlr.TR119000391>.
- [58] Glenn F.W. Walpole, Sergio Grinstein, and Johannes Westman. "The role of lipids in host–pathogen interactions". *IUBMB Life* 70.5 (2018), pp. 384–392. DOI: <https://doi.org/10.1002/iub.1737>.
- [59] Valery N. Bochkov, Olga V. Oskolkova, Konstantin G. Birukov, Anna Liisa Levenonen, Christoph J. Binder, and Johannes Stöckl. "Generation and biological activities of oxidized phospholipids". *Antioxid. Redox Signal* 12.8 (Apr. 2010), pp. 1009–1059. DOI: 10.1089/ars.2009.2597.
- [60] Waldemar Kulig, Agnieszka Olżyńska, Piotr Jurkiewicz, Anu M Kantola, Sanna Komulainen, Moutusi Manna, Mohsen Pourmousa, Mario Vazdar, Lukasz Cwiklik, Tomasz Rog, George Khelashvili, Daniel Harries, Ville-Veikko Telkki, Martin Hof, Ilpo Vattulainen, and Pavel Jungwirth. "Cholesterol under oxidative stress - How lipid membranes sense oxidation as cholesterol is being replaced by oxysterols". *Free Radical Bio. Med.* 84 (2015), pp. 30–41. DOI: 10.1016/j.freeradbiomed.2015.03.006.

- [61] Waldemar Kulig, Lukasz Cwiklik, Piotr Jurkiewicz, Tomasz Rog, and Ilpo Vattulainen. "Cholesterol oxidation products and their biological importance". *Chem. Phys. Lipids* 199 (2016), pp. 144–160. DOI: 10.1016/j.chemphyslip.2016.03.001.
- [62] Monika Bokori-Brown, Peter G Petrov, Mawya A Khafaji, Muhammad K Mughal, Claire E Naylor, Angela C Shore, Kim M Gooding, Francesco Casanova, Tim J Mitchell, Richard W Titball, and C Peter Winlove. "Red blood cell susceptibility to pneumolysin: Correlation with membrane biochemical and physical properties". *J. Biol. Chem.* 291.19 (2016), pp. 10210–10227. DOI: 10.1074/jbc.M115.691899.
- [63] M Aoun, P A Corsetto, G Nugue, G Montorfano, E Ciusani, D Crouzier, P Hogarth, A Gregory, S Hayflick, G Zorzi, A M Rizzo, and V Tiranti. "Changes in Red Blood Cell membrane lipid composition: A new perspective into the pathogenesis of PKAN". *Mol. Genet. Metab.* 121.2 (2017), pp. 180–189.
- [64] Rimsha Munir, Jan Lisec, Johannes V. Swinnen, and Nousheen Zaidi. "Lipid metabolism in cancer cells under metabolic stress". *Brit. J. Cancer* 120.12 (June 2019), pp. 1090–1098. DOI: 10.1038/s41416-019-0451-4.
- [65] Wojciech Szlasa, Iga Zendran, Aleksandra Zalesińska, Mounir Tarek, and Julita Kulbacka. "Lipid composition of the cancer cell membrane". *J. Bioenerg. Biomembr.* 52.5 (Oct. 2020), pp. 321–342. DOI: 10.1007/s10863-020-09846-4.
- [66] Olivier Belzile, Xianming Huang, Jian Gong, Jay Carlson, Alan Schroit, Rolf Brekken, and Bruce Freimark. "Antibody targeting of phosphatidylserine for the detection and immunotherapy of cancer". *ImmunoTargets Ther* Volume 7 (Jan. 2018), pp. 1–14. DOI: 10.2147/itt.s134834.
- [67] Sadnu Budha, Rachel Giese, Aditi Gupta, Olivier De Henau, Roberta Zappasodi, Luis Felipe Campesato, Christopher Barker, Joseph Shan, Jedd D Wolchok, and Taha Merghoub. "Phosphatidylserine targeting antibody in combination with tumor radiation and immune checkpoint blockade promotes anti-



- tumor activity in mouse B16 melanoma". *J. Immunol.* 200.1 Supplement (2018), pp. 122.10–122.10.
- [68] Daan Frenkel. "Introduction to Monte Carlo Methods". In: *Computational Soft Matter: From Synthetic Polymers to Proteins - Lecture Notes*. Ed. by Attig Norbert, Binder Kurt, Grubmüller Helmut, and Kurt Kremer. Bonn: John von Neumann Institute for Computing, 2004.
- [69] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. "Equation of state calculations by fast computing machines". *J. Chem. Phys.* 21.6 (Dec. 1953), pp. 1087–1092. DOI: 10.1063/1.1699114.
- [70] Marshall N. Rosenbluth. "Genesis of the Monte Carlo Algorithm for Statistical Mechanics". In: *AIP Conference Proceedings*. Vol. 690. 1. AIP Publishing, Jan. 2003, pp. 22–30. DOI: 10.1063/1.1632112.
- [71] J.E. Gubernatis. "Marshall Rosenbluth and the Metropolis algorithm". In: *Physics of Plasmas*. Vol. 12. 5. American Institute of PhysicsAIP, May 2005, pp. 1–5. DOI: 10.1063/1.1887186.
- [72] Michel Mareschal. *Almost famous, a woman behind the codes - Part 1: It's only numerics*. 2017.
- [73] Brad Lee Holian. "Exploring the boundary between atoms and the continuum by computers: a personal history". *Eur. Phys. J. H* 46.1 (Dec. 2021), pp. 1–12. DOI: 10.1140/epjh/s13129-021-00010-z.
- [74] Marshall N. Rosenbluth and Arianna W. Rosenbluth. "Further results on Monte Carlo equations of state". *J. Chem. Phys.* 22.5 (Dec. 1954), pp. 881–884. DOI: 10.1063/1.1740207.
- [75] George Michael and Berni Alder. *An Interview with Bernie Alder*. 1997.
- [76] William W Wood. *A Brief History of the Use of the Metropolis Method at LANL in the 1950s*. Tech. rep. 2003.

- [77] Giovanni Battimelli and Giovanni Ciccotti. “Berni Alder and the pioneering times of molecular simulation”. *Eur. Phys. J. H* 43 (2018), pp. 303–335. DOI: 10.1140/epjh/e2018-90027-5.
- [78] B.J. Alder and T.E. Wainwright. “Phase transition for a hard sphere system”. *J. Chem. Phys.* 27.5 (Aug. 1957), pp. 1208–1209. DOI: 10.1063/1.1743957.
- [79] W.W. Wood and J.D. Jacobson. “Preliminary results from a recalculation of the Monte Carlo equation of state of hard spheres”. *J. Chem. Phys.* 27.5 (Aug. 1957), pp. 1207–1208. DOI: 10.1063/1.1743956.
- [80] B.J. Alder and T.E. Wainwright. “Studies in molecular dynamics. I. General method”. *J. Chem. Phys.* 31.2 (Aug. 1959), pp. 459–466. DOI: 10.1063/1.1730376.
- [81] Mary Ann Mansigh and Daan Frenkel. *Almost famous, a woman behind the codes - Part 2: A conversation with Mary Ann Mansigh*. 2017.
- [82] Stewart A. Adcock and J. Andrew McCammon. “Molecular dynamics: Survey of methods for simulating the activity of proteins”. *Chem. Rev.* 106.5 (May 2006), pp. 1589–1615. DOI: 10.1021/cr040426m.
- [83] Elisa Fadda and Robert J. Woods. “Molecular simulations of carbohydrates and protein-carbohydrate interactions: Motivation, issues and prospects”. *Drug Discov. Today* 15.15-16 (Aug. 2010), pp. 596–609. DOI: 10.1016/j.drudis.2010.06.001.
- [84] Jiří Šponer, Giovanni Bussi, Miroslav Krepl, Pavel Banáš, Sandro Bottaro, Richard A. Cunha, Alejandro Gil-Ley, Giovanni Pinamonti, Simón Poblete, Petr Jurečka, Nils G. Walter, and Michal Otyepka. “RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview”. *Chem. Rev.* 118.8 (2018). PMID: 29297679, pp. 4177–4338. DOI: 10.1021/acs.chemrev.7b00427.
- [85] Manas Mondal, Lijiang Yang, Zhicheng Cai, Piya Patra, and Yi Qin Gao. “A perspective on the molecular simulation of DNA from structural and functional aspects”. *Chem. Sci.* 12.15 (Apr. 2021), pp. 5390–5409. DOI: 10.1039/d0sc05329e.

- [86] Melanie P. Muller, Tao Jiang, Chang Sun, Muyun Lihan, Shashank Pant, Paween Mahinthichaichan, Anda Trifan, and Emad Tajkhorshid. "Characterization of Lipid-Protein Interactions and Lipid-Mediated Modulation of Membrane Protein Function through Molecular Simulation". *Chem. Rev.* 119.9 (May 2019), pp. 6086–6161. DOI: 10.1021/acs.chemrev.8b00608.
- [87] Xiaolin Cheng and Jeremy C Smith. "Biological Membrane Organization and Cellular Signaling". *Chem. Rev.* 119 (2019), pp. 5849–5880. DOI: 10.1021/acs.chemrev.8b00439.
- [88] Valentina Corradi, Besian I. Sejdiu, Haydee Mesa-Galloso, Haleh Abdizadeh, Sergei Yu. Noskov, Siewert J. Marrink, and D. Peter Tieleman. "Emerging Diversity in Lipid-Protein Interactions". *Chem. Rev.* 119.9 (2019), pp. 5775–5848. DOI: 10.1021/acs.chemrev.8b00451.
- [89] J.B. Gibson, A.N. Goland, M. Milgram, and G.H. Vineyard. "Dynamics of Radiation Damage". *Phys. Rev.* 120 (4 Nov. 1960), pp. 1229–1253. DOI: 10.1103/PhysRev.120.1229.
- [90] A. Rahman. "Correlations in the Motion of Atoms in Liquid Argon". *Phys. Rev.* 136 (2A Oct. 1964), A405–A411. DOI: 10.1103/PhysRev.136.A405.
- [91] A.J. Kox, J.P.J. Michels, and F.W. Wiegel. "Simulation of a lipid monolayer using molecular dynamics". *Nature* 287.5780 (Sept. 1980), pp. 317–319. DOI: 10.1038/287317a0.
- [92] P. van der Ploeg and H.J.C. Berendsen. "Molecular dynamics simulation of a bilayer membrane". *J. Chem. Phys.* 76.6 (June 1982), pp. 3271–3276. DOI: 10.1063/1.443321.
- [93] E. Egberts and H.J.C. Berendsen. "Molecular dynamics simulation of a smectic liquid crystal with atomic detail". *J. Chem. Phys.* 89.6 (1988), pp. 3718–3732. DOI: 10.1063/1.454893.
- [94] Max L. Berkowitz and K. Raghavan. "Computer Simulation of a Water/Membrane Interface". *Langmuir* 7.6 (June 1991), pp. 1042–1044. DOI: 10.1021/la00054a002.

- [95] S.E. Feller, R.W. Pastor, A. Rojnuckarin, S. Bogusz, and B.R. Brooks. "Effect of electrostatic force truncation on interfacial and transport properties of water". *J. Phys. Chem.* 100.42 (1996), pp. 17011–17020. DOI: 10.1021/jp9614658.
- [96] William L. Jorgensen, David S. Maxwell, and Julian Tirado-Rives. "Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids". *J. Am. Chem. Soc.* 118.45 (Nov. 1996), pp. 11225–11236. DOI: 10.1021/ja962176o.
- [97] Jiwei Wang, Megha, and Erwin London. "Relationship between Sterol/Steroid Structure and Participation in Ordered Lipid Domains (Lipid Rafts): Implications for Lipid Raft Structure and Function". *Biochemistry* 43.4 (Feb. 2004), pp. 1010–1018. DOI: 10.1021/bi035696y.
- [98] Siewert J. Marrink, H. Jelger Risselada, Serge Yefimov, D. Peter Tieleman, and Alex H. De Vries. "The MARTINI force field: Coarse grained model for biomolecular simulations". *J. Phys. Chem. B.* 111.27 (July 2007), pp. 7812–7824. DOI: 10.1021/jp071097f.
- [99] Karl N. Kirschner, Austin B. Yongye, Sarah M. Tschampel, Jorge González-Outeiriño, Charlisa R. Daniels, B. Lachele Foley, and Robert J. Woods. "GLYCAM06: A generalizable biomolecular force field. carbohydrates". *J. Comput. Chem.* 29.4 (Mar. 2008), pp. 622–655. DOI: 10.1002/jcc.20820.
- [100] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A.D. Mackerell. "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields". *J. Comput. Chem.* 31.4 (Mar. 2010), pp. 671–690. DOI: 10.1002/jcc.21367.
- [101] Olgun Guvench, Sairam S. Mallajosyula, E. Prabhu Raman, Elizabeth Hatcher, Kenno Vanommeslaeghe, Theresa J. Foster, Francis W. Jamison, and Alexander D. MacKerell. "CHARMM additive all-atom force field for carbohydrate derivatives and its utility in polysaccharide and carbohydrate-protein modeling". *J. Chem. Theory Comput.* 7.10 (Oct. 2011), pp. 3162–3180. DOI: 10.1021/ct200328p.

- [102] Jeffery B. Klauda, Richard M. Venable, J. Alfredo Freites, Joseph W. O'Connor, Douglas J. Tobias, Carlos Mondragon-Ramirez, Igor Vorobyov, Alexander D. MacKerell, and Richard W. Pastor. "Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types". *J. Phys. Chem. B* 114.23 (June 2010), pp. 7830–7843. DOI: 10.1021/jp101759q.
- [103] Richard M. Venable, Alexander J. Sodt, Brent Rogaski, Huan Rui, Elizabeth Hatcher, Alexander D. MacKerell, Richard W. Pastor, and Jeffery B. Klauda. "CHARMM all-atom additive force field for sphingomyelin: Elucidation of hydrogen bonding and of positive curvature". *Biophys. J.* 107.1 (July 2014), pp. 134–145. DOI: 10.1016/j.bpj.2014.05.034.
- [104] Callum J. Dickson, Benjamin D. Madej, Åge A. Skjevik, Robin M. Betz, Knut Teigen, Ian R. Gould, and Ross C. Walker. "Lipid14: The amber lipid force field". *J. Chem. Theory Comput.* 10.2 (Feb. 2014), pp. 865–879. DOI: 10.1021/ct4010307.
- [105] Jing Huang, Sarah Rauscher, Grzegorz Nawrocki, Ting Ran, Michael Feig, Bert L. De Groot, Helmut Grubmüller, and Alexander D. MacKerell. "CHARMM36m: An improved force field for folded and intrinsically disordered proteins". *Nat. Methods* 14.1 (Dec. 2017), pp. 71–73. DOI: 10.1038/nmeth.4067.
- [106] Helgi I. Ingólfsson, Manuel N. Melo, Floris J. Van Eerden, Clément Arnarez, Cesar A. Lopez, Tsjerk A. Wassenaar, Xavier Periole, Alex H. De Vries, D. Peter Tieleman, and Siewert J. Marrink. "Lipid organization of the plasma membrane". *J. Am. Chem. Soc.* 136.41 (Jan. 2014), pp. 14554–14559. DOI: 10.1021/ja507832e.
- [107] Helgi I. Ingólfsson, Timothy S. Carpenter, Harsh Bhatia, Peer Timo Bremer, Siewert J. Marrink, and Felice C. Lightstone. "Computational Lipidomics of the Neuronal Plasma Membrane". *Biophys. J.* 113.10 (2017), pp. 2271–2280. DOI: 10.1016/j.bpj.2017.10.017.

- [108] Swapnil Baral, Ilya Levental, and Edward Lyman. “Composition dependence of cholesterol flip-flop rates in physiological mixtures”. *Chem. Phys. Lipids* 232 (Oct. 2020), p. 104967. DOI: 10.1016/j.chemphyslip.2020.104967.
- [109] Valentina Corradi, Eduardo Mendez-Villuendas, Helgi I. Ingólfsson, Ruo-Xu Gu, Iwona Siuda, Manuel N. Melo, Anastassia Moussatova, Lucien J. DeGagné, Besian I. Sejdiu, Gurpreet Singh, Tsjerk A. Wassenaar, Karelia Delgado Magner, Siewert J. Marrink, and D. Peter Tieleman. “Lipid–Protein Interactions Are Unique Fingerprints for Membrane Proteins”. *ACS Cent. Sci.* 4.6 (2018). PMID: 29974066, pp. 709–717. DOI: 10.1021/acscentsci.8b00143.
- [110] Alvin Yu, Alexander J. Pak, Peng He, Viviana Monje-Galvan, Lorenzo Casalino, Zied Gaieb, Abigail C. Dommer, Rommie E. Amaro, and Gregory A. Voth. “A multiscale coarse-grained model of the SARS-CoV-2 virion”. *Biophys. J.* 120.6 (2021), pp. 1097–1104. DOI: <https://doi.org/10.1016/j.bpj.2020.10.048>.
- [111] Lena Harker-Kirschneck, Buzz Baum, and Anđela Šarić. “Changes in ESCRT-III filament geometry drive membrane remodelling and fission in silico”. *BMC Biol.* 17.1 (Oct. 2019), pp. 1–8. DOI: 10.1186/s12915-019-0700-2.
- [112] Gabriel Tarrason Risa, Fredrik Hurtig, Sian Bray, Anne E. Hafner, Lena Harker-Kirschneck, Peter Faull, Colin Davis, Dimitra Papatziomou, Delyan R. Mutavchiev, Catherine Fan, Leticia Meneguello, Andre Arashiro Pulschen, Gautam Dey, Siân Culley, Mairi Kilkenny, Diorge P. Souza, Luca Pellegrini, Robertus A.M. de Bruin, Ricardo Henriques, Ambrosius P. Snijders, Anđela Šarić, Ann Christin Lindås, Nicholas P. Robinson, and Buzz Baum. “The proteasome controls ESCRT-III-mediated cell division in an archaeon”. *Science* 369.6504 (Aug. 2020). DOI: 10.1126/science.aaz2532.
- [113] William L Jorgensen, Jayaraman Chandrasekhar, Jeffrey D Madura, Roger W Impey, and Michael L Klein. “Comparison of Simple Potential Functions for Simulating Liquid Water”. *J. Chem. Phys.* 79.2 (1983), pp. 926–935. DOI: 10.1063/1.445869.

- [114] Sereina Riniker. “Fixed-Charge Atomistic Force Fields for Molecular Dynamics Simulations in the Condensed Phase: An Overview”. *J. Chem. Inf. Model.* 58.3 (2018). PMID: 29510041, pp. 565–578. DOI: 10.1021/acs.jcim.8b00042.
- [115] Michael F. Russo and Adri C.T. van Duin. “Atomistic-scale simulations of chemical reactions: Bridging from quantum chemistry to engineering”. *Nucl. Instrum. Methods Phys. Res., Sect. B* 269.14 (2011). Computer Simulations of Radiation Effects in Solids, pp. 1549–1554. DOI: <https://doi.org/10.1016/j.nimb.2010.12.053>.
- [116] Michael P. Allen and Dominic J. Tildesley. *Computer simulation of liquids: Second edition*. Oxford University Press, Nov. 2017, pp. 1–626. DOI: 10.1093/oso/9780198803195.001.0001.
- [117] Flaviu S. Cipcigan, Vlad P. Sokhan, Andrew P. Jones, Jason Crain, and Glenn J. Martyna. “Hydrogen bonding and molecular orientation at the liquid–vapour interface of water”. *Phys. Chem. Chem. Phys.* 17 (14 2015), pp. 8660–8669. DOI: 10.1039/C4CP05506C.
- [118] David J. Smith, Jeffrey B. Klauda, and Alexander J. Sodt. “Simulation Best Practices for Lipid Membranes [Article VI.0]”. *LiveCoMS* 1.1 (2019). DOI: 10.33011/livecoms.1.1.5966.
- [119] Fatemeh Namdari and Gholamabbas Parsafar. “Calculation of effective pair potential via the volumetric data”. *Phase Transit.* 87.1 (2014), pp. 38–58. DOI: 10.1080/01411594.2013.804184.
- [120] Pouyan Khakbaz and Jeffery B. Klauda. “Investigation of phase transitions of saturated phosphocholine lipid bilayers via molecular dynamics simulations”. *Biochim Biophys Acta Biomembr* 1860.8 (Aug. 2018), pp. 1489–1501. DOI: 10.1016/j.bbamem.2018.04.014.
- [121] Ruo-Xu Gu, Svetlana Baoukina, and D. Peter Tieleman. “Phase Separation in Atomistic Simulations of Model Membranes”. *J. Am. Chem. Soc.* 142.6 (Jan. 2020), pp. 2844–2856. DOI: 10.1021/jacs.9b11057.

- [122] L. Martínez, R. Andrade, E.G. Birgin, and J.M. Martínez. “PACKMOL: A package for building initial configurations for molecular dynamics simulations”. *J. Comput. Chem.* 30.13 (2009), pp. 2157–2164. DOI: <https://doi.org/10.1002/jcc.21224>.
- [123] Fabian Grünewald, Riccardo Alessandri, Peter C Kroon, Luca Monticelli, Paulo CT Souza, and Siewert J Marrink. “Polyply: a python suite for facilitating simulations of (bio-) macromolecules and nanomaterials”. *arXiv preprint arXiv:2105.05890* (2021).
- [124] Sunhwan Jo, Taehoon Kim, Vidyashankara G. Iyer, and Wonpil Im. “CHARMM-GUI: A web-based graphical user interface for CHARMM”. *J. Comput. Chem.* 29.11 (2008). DOI: [10.1002/jcc.20945](https://doi.org/10.1002/jcc.20945).
- [125] Yifei Qi, Helgi I. Ingólfsson, Xi Cheng, Jumin Lee, Siewert J. Marrink, and Wonpil Im. “CHARMM-GUI Martini Maker for Coarse-Grained Simulations with the Martini Force Field”. *J. Chem. Theory Comput.* 11.9 (Sept. 2015), pp. 4486–4494. DOI: [10.1021/acs.jctc.5b00513](https://doi.org/10.1021/acs.jctc.5b00513).
- [126] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindah. “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers”. *SoftwareX* 1-2 (2015). DOI: [10.1016/j.softx.2015.06.001](https://doi.org/10.1016/j.softx.2015.06.001).
- [127] William C. Swope, Hans C. Andersen, Peter H. Berens, and Kent R. Wilson. “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters”. *J. Chem. Phys.* 76.1 (1982), pp. 637–649. DOI: [10.1063/1.442716](https://doi.org/10.1063/1.442716).
- [128] D.C. Rapaport. *The Art of Molecular Dynamics Simulation*. Cambridge University Press, Apr. 2004. DOI: [10.1017/cbo9780511816581](https://doi.org/10.1017/cbo9780511816581).
- [129] Berk Hess, Henk Bekker, Herman J.C. Berendsen, and Johannes G.E.M. Fraaije. “LINCS: A Linear Constraint Solver for molecular simulations”. *J. Comput. Chem.* 18.12 (Sept. 1997), pp. 1463–1472. DOI: [10.1002/\(SICI\)1096-987X\(199709\)18:12<1463::AID-JCC4>3.0.CO;2-H](https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H).



- [130] Jean Paul Ryckaert, Giovanni Ciccotti, and Herman J.C. Berendsen. “Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes”. *J. Comput. Phys.* 23.3 (Mar. 1977), pp. 327–341. DOI: 10.1016/0021-9991(77)90098-5.
- [131] Mark E Tuckerman. *Statistical Mechanics: Theory and Molecular Simulation*. 1st. Oxford: Oxford University Press, 2010.
- [132] Giovanni Bussi, Davide Donadio, and Michele Parrinello. “Canonical sampling through velocity rescaling”. *J. Chem. Phys.* 126.1 (Jan. 2007), p. 014101. DOI: 10.1063/1.2408420.
- [133] Shuichi Nosé. “A unified formulation of the constant temperature molecular dynamics methods”. *J. Chem. Phys.* 81.1 (July 1984), pp. 511–519. DOI: 10.1063/1.447334.
- [134] William G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. *Phys. Rev. A* 31.3 (Mar. 1985), pp. 1695–1697. DOI: 10.1103/PhysRevA.31.1695.
- [135] H. Sun. “COMPASS: An ab Initio Force-Field Optimized for Condensed-Phase Applications - Overview with Details on Alkane and Benzene Compounds”. *J. Phys. Chem. B* 102.38 (1998), pp. 7338–7364. DOI: 10.1021/jp980939v.
- [136] Djurre H. de Jong, Svetlana Baoukina, Helgi I. Ingólfsson, and Siewert J. Marrink. “Martini straight: Boosting performance using a shorter cutoff and GPUs”. *Comput. Phys. Commun.* 199 (2016), pp. 1–7. DOI: <https://doi.org/10.1016/j.cpc.2015.09.014>.
- [137] Tom Darden, Darrin York, and Lee Pedersen. “Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems”. *J. Chem. Phys.* 98.12 (1993), pp. 10089–10092. DOI: 10.1063/1.464397.
- [138] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. “A smooth particle mesh Ewald method”. *J. Chem. Phys.* 103.19 (1995), pp. 8577–8593. DOI: 10.1063/1.470117.

- [139] “Parameters for Martini sterols and hopanoids based on a virtual-site description”. *J. Chem. Phys.* 143.24 (Dec. 2015), p. 243152. DOI: 10.1063/1.4937783.
- [140] Ilario G. Tironi, René Sperb, Paul E. Smith, and Wilfred F. van Gunsteren. “A generalized reaction field method for molecular dynamics simulations”. *J. Chem. Phys.* 102.13 (1995), pp. 5451–5459. DOI: 10.1063/1.469273.
- [141] Siewert J. Marrink and D. Peter Tieleman. “Perspective on the martini model”. *Chem. Soc. Rev.* 42.16 (July 2013), pp. 6801–6822. DOI: 10.1039/c3cs60093a.
- [142] Francesco Trozzi, Xinlei Wang, and Peng Tao. “UMAP as a Dimensionality Reduction Tool for Molecular Dynamics Simulations of Biomacromolecules: A Comparison Study”. *J. Phys. Chem. B* 125.19 (2021). PMID: 33973773, pp. 5022–5034. DOI: 10.1021/acs.jpcc.1c02081.
- [143] Aldo Glielmo, Brooke E. Husic, Alex Rodriguez, Cecilia Clementi, Frank Noé, and Alessandro Laio. “Unsupervised Learning Methods for Molecular Simulation Data”. *Chem. Rev.* 121.16 (2021). PMID: 33945269, pp. 9722–9758. DOI: 10.1021/acs.chemrev.0c01195.
- [144] Leland McInnes, John Healy, and James Melville. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction”. *arXiv* (Feb. 2018).
- [145] Leland McInnes, John Healy, and Steve Astels. “hdbscan: Hierarchical density based clustering”. *J. Open Source Softw.* 2.11 (2017), p. 205. DOI: 10.21105/joss.00205.
- [146] Claudia Malzer and Marcus Baum. “A Hybrid Approach To Hierarchical Density-based Cluster Selection”. *arXiv* (Nov. 2019).
- [147] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. “A density-based algorithm for discovering clusters in large spatial databases with noise”. In: AAAI Press, 1996, pp. 226–231.
- [148] Paul Smith, Peter G. Petrov, and Christian D. Lorenz. “Cholesterol Oxidation Modulates the Formation of Liquid-Ordered Domains in Model Membranes”. *bioRxiv* (2021). DOI: 10.1101/2021.05.24.445501.

- [149] Matti Javanainen, Balazs Fabian, and Hector Martinez-Seara. *Comment on "Capturing Phase Behavior of Ternary Lipid Mixtures with a Refined Martini Coarse-Grained Force Field"*. 2020.
- [150] Alexandre Toulmay and William A. Prinz. "Direct imaging reveals stable, micrometer-scale lipid domains that segregate proteins in live cells". *J. Cell Biol.* 202.1 (July 2013), pp. 35–44. DOI: 10.1083/jcb.201301039.
- [151] Rupali Prasad, Andrzej Sliwa-Gonzalez, and Yves Barral. "Mapping bilayer thickness in the ER membrane". *Sci. Adv.* 6.46 (Nov. 2020), eaba5130. DOI: 10.1126/sciadv.aba5130.
- [152] J.F. McKenna, D.J. Rolfe, S.E.D. Webb, A.F. Tolmie, S.W. Botchway, M.L. Martin-Fernandez, C. Hawes, and J. Runions. "The cell wall regulates dynamics and size of plasma-membrane nanodomains in Arabidopsis". *Proc. Natl. Acad. Sci. U.S.A* 116.26 (June 2019), pp. 12857–12862. DOI: 10.1073/pnas.1819077116.
- [153] Jonathan D. Nickels, Sneha Chatterjee, Christopher B. Stanley, Shuo Qian, Xiaolin Cheng, Dean A.A. Myles, Robert F. Standaert, James G. Elkins, and John Katsaras. "The in vivo structure of biological membranes and evidence for lipid domains". *PLOS Biol.* 15.5 (May 2017). Ed. by Daniel Lopez, e2002214. DOI: 10.1371/journal.pbio.2002214.
- [154] Daniel Lopez and Gudrun Koch. *Exploring functional membrane microdomains in bacteria: an overview*. Apr. 2017. DOI: 10.1016/j.mib.2017.02.001.
- [155] Frederick A. Heberle, Milka Doktorova, Haden L. Scott, Allison D. Skinkle, M. Neal Waxham, and Ilya Levental. "Direct label-free imaging of nanodomains in biomimetic and biological membranes by cryogenic electron microscopy". *Proc. Natl. Acad. Sci. U.S.A* 117.33 (Aug. 2020), pp. 19943–19952. DOI: 10.1073/PNAS.2002200117.
- [156] Leonard J. Foster, Carmen L. De Hoog, and Matthias Mann. "Unbiased quantitative proteomics of lipid rafts reveals high specificity for signaling factors". *Proc. Natl. Acad. Sci. U.S.A* 100.10 (May 2003), pp. 5813–5818. DOI: 10.1073/pnas.0631608100.

- [157] Blanca B. Diaz-Rohrer, Kandice R. Levental, Kai Simons, and Ilya Levental. “Membrane raft association is a determinant of plasma membrane localization”. *Proc. Natl. Acad. Sci. U.S.A* 111.23 (June 2014), pp. 8500–8505. DOI: 10.1073/pnas.1404582111.
- [158] Tomasz J. Proszynski, Robin Klemm, Michel Bagnat, Katharina Gaus, and Kai Simons. “Plasma membrane polarization during mating in yeast cells”. *J. Cell Biol.* 173.6 (June 2006), pp. 861–866. DOI: 10.1083/jcb.200602007.
- [159] Dina Vind-Kezunovic, Claus Hélix Nielsen, Urszula Wojewodzka, and Robert Gniadecki. “Line tension at lipid phase boundaries regulates formation of membrane vesicles in living cells”. *Biochim Biophys Acta Biomembr* 1778.11 (Nov. 2008), pp. 2480–2486. DOI: 10.1016/j.bbamem.2008.05.015.
- [160] Sara Y. Cheng, Yiyi Cao, Marzieh Rouzbehani, and Kwan H. Cheng. “Coarse-grained MD simulations reveal beta-amyloid fibrils of various sizes bind to interfacial liquid-ordered and liquid-disordered regions in phase separated lipid rafts with diverse membrane-bound conformational states”. *Biophys. Chem.* 260 (May 2020), p. 106355. DOI: 10.1016/j.bpc.2020.106355.
- [161] “Pathogens: Raft hijackers”. *Nat. Rev. Immunol.* 3.7 (July 2003), pp. 557–568. DOI: 10.1038/nri1129.
- [162] Huiyong Yin, Libin Xu, and Ned A. Porter. “Free radical lipid peroxidation: Mechanisms and analysis”. *Chem. Rev.* 111.10 (Oct. 2011), pp. 5944–5972. DOI: 10.1021/cr200084z.
- [163] Valérie Deckert, Laurence Perségol, Laurence Viens, Gérard Lizard, Anne Athias, Christian Lallemand, Philippe Gambert, and Laurent Lagrost. “Inhibitors of arterial relaxation among components of human oxidized low-density lipoproteins: Cholesterol derivatives oxidized in position 7 are potent inhibitors of endothelium-dependent relaxation”. *Circulation* 95.3 (1997), pp. 723–731. DOI: 10.1161/01.CIR.95.3.723.

- [164] Gerard Lizard, Maryvonne Moisant, Catherine Cordelet, Serge Monier, Philippe Gambert, and Laurent Lagrost. "Induction of similar features of apoptosis in human and bovine vascular endothelial cells treated by 7-ketocholesterol". *J. Pathol.* 183.3 (Nov. 1997), pp. 330–338. DOI: 10.1002/(SICI)1096-9896(199711)183:3<330::AID-PATH933>3.0.CO;2-7.
- [165] Marie Charlotte Royer, Stéphanie Lemaine-Ewing, Catherine Desrumaux, Serge Monier, Jean Paul Pais de Barros, Anne Athias, Dominique Néel, and Laurent Lagrost. "7-ketocholesterol incorporation into sphingolipid/cholesterol-enriched (Lipid Raft) domains is impaired by vitamin E. A specific role for  $\alpha$ -tocopherol with consequences on cell death". *J. Biol. Chem.* 284.23 (June 2009), pp. 15826–15834. DOI: 10.1074/jbc.M808641200.
- [166] Amira Zarrouk, Anne Vejux, John Mackrill, Yvonne O'Callaghan, Mohamed Hammami, Nora O'Brien, and Gérard Lizard. *Involvement of oxysterols in age-related diseases and ageing processes*. Nov. 2014. DOI: 10.1016/j.arr.2014.09.006.
- [167] Rodolphe Filomenko, Cynthia Fourgeux, Lionel Bretillon, and Ségolène Gambert-Nicot. "Oxysterols: Influence on plasma membrane rafts microdomains and development of ocular diseases". *Steroids* 99.PB (July 2015), pp. 259–265. DOI: 10.1016/j.steroids.2015.02.004.
- [168] Paola Gamba, Gabriella Testa, Simona Gargiulo, Erica Staurenghi, Giuseppe Poli, and Gabriella Leonarduzzi. *Oxidized cholesterol as the driving force behind the development of Alzheimer's disease*. June 2015. DOI: 10.3389/fnagi.2015.00119.
- [169] Winnie Luu, Laura J. Sharpe, Isabelle Capell-Hattam, Ingrid C. Gelissen, and Andrew J. Brown. "Oxysterols: Old Tale, New Twists". *Annu. Rev. Pharmacol.* 56.1 (Jan. 2016), pp. 447–467. DOI: 10.1146/annurev-pharmtox-010715-103233.
- [170] William J. Griffiths and Yuqin Wang. *Oxysterol research: A brief review*. 2019. DOI: 10.1042/BST20180135.
- [171] "Oxysterols and gastrointestinal cancers around the clock". *Front. Endocrinol* 10.JULY (July 2019), p. 483. DOI: 10.3389/fendo.2019.00483.

- [172] Amelia Anderson, Angielyn Campo, Elena Fulton, Anne Corwin, W. Gray Jerome, and Matthew S. O'Connor. *7-Ketocholesterol in disease and aging*. Jan. 2020. DOI: 10.1016/j.redox.2019.101380.
- [173] Andrew J. Brown, Laura J. Sharpe, and Michael J. Rogers. "Oxysterols: From physiological tuners to pharmacological opportunities". *Brit. J. Pharmacol.* (2020), pp. 1–15. DOI: 10.1111/bph.15073.
- [174] Binlu Huang, Bao liang Song, and Chenqi Xu. "Cholesterol metabolism in cancer: mechanisms and therapeutic opportunities". *Nat. Metab.* 2.2 (Feb. 2020), pp. 132–141. DOI: 10.1038/s42255-020-0174-0.
- [175] Dana-Marie Telesford, Dominique Verreault, Victoria Reick-Mitrisin, and Heather C Allen. "Reduced Condensing and Ordering Effects by 7-Ketocholesterol and  $5\beta,6\beta$ -Epoxycholesterol on DPPC Monolayers". *Langmuir* 31.36 (2015), pp. 9859–9869. DOI: 10.1021/acs.langmuir.5b02539.
- [176] Antenor J P Neto and Rodrigo M Cordeiro. "Molecular simulations of the effects of phospholipid and cholesterol peroxidation on lipid membrane properties". *Biochim Biophys Acta Biomembr* 1858 (2016), pp. 2191–2198. DOI: 10.1016/j.bbamem.2016.06.018.
- [177] Arnaud Berthier, Stéphanie Lemaire-Ewing, Céline Prunet, Thomas Montange, Anne Vejux, Jean Paul Pais de Barros, Serge Monier, Philippe Gambert, Gérard Lizard, and Dominique Néel. "7-Ketocholesterol-induced apoptosis". *FEBS J.* 272.12 (June 2005), pp. 3093–3104. DOI: 10.1111/j.1742-4658.2005.04723.x.
- [178] Anne Vejux, Stéphane Guyot, Thomas Montange, Jean Marc Riedinger, Edmond Kahn, and Gérard Lizard. "Phospholipidosis and down-regulation of the PI3-K/PDK-1/Akt signalling pathway are vitamin E inhibitable events associated with 7-ketocholesterol-induced apoptosis". *J. Nutr. Biochem.* 20.1 (Jan. 2009), pp. 45–61. DOI: 10.1016/j.jnutbio.2007.12.001.
- [179] Derek Marsh. "Cholesterol-induced fluid membrane domains: A compendium of lipid-raft ternary phase diagrams". *Biochim Biophys Acta Biomembr* 1788.10 (Oct. 2009), pp. 2114–2123. DOI: 10.1016/j.bbamem.2009.08.004.

- [180] Celsa Díaz-Tejada, Igor Ariz-Extreme, Neha Awasthi, and Jochen S. Hub. “Quantifying Lateral Inhomogeneity of Cholesterol-Containing Membranes”. *J. Phys. Chem. Lett.* 6.23 (Dec. 2015), pp. 4799–4803. DOI: 10.1021/acs.jpcllett.5b02414.
- [181] Priyadarshini Pathak and Erwin London. “The Effect of Membrane Lipid Composition on the Formation of Lipid Ultrananodomains”. *Biophys. J.* 109.8 (Oct. 2015), pp. 1630–1638. DOI: 10.1016/j.bpj.2015.08.029.
- [182] Marek Cebecauer, Mariana Amaro, Piotr Jurkiewicz, Maria Joao Sarmiento, Radek Šachl, Lukasz Cwiklik, and Martin Hof. “Membrane Lipid Nanodomains”. *Chem. Rev.* 118.23 (Dec. 2018), pp. 11259–11297. DOI: 10.1021/acs.chemrev.8b00322.
- [183] Sebastian Thallmair, Helgi I. Ingólfsson, Siewert J. Marrink, Helgi I Ingólfsson, and Siewert J. Marrink. “Cholesterol Flip-Flop Impacts Domain Registration in Plasma Membrane Models”. *J. Phys. Chem. Lett.* 9 (Sept. 2018), pp. 5527–5533. DOI: 10.1021/acs.jpcllett.8b01877.
- [184] Arturo Galván-Hernández, Naritaka Kobayashi, Jorge Hernández-Cobos, Armando Antillón, Seiichiro Nakabayashi, and Iván Ortega-Blake. “Morphology and dynamics of domains in ergosterol or cholesterol containing membranes”. *Biochim Biophys Acta Biomembr* 1862.2 (Feb. 2020), p. 183101. DOI: 10.1016/j.bbamem.2019.183101.
- [185] Guangtao Li, Qing Wang, Shinako Kakuda, and Erwin London. “Nanodomains can persist at physiologic temperature in plasma membrane vesicles and be modulated by altering cell lipids”. *J. Lipid Res.* 61.5 (May 2020), pp. 758–766. DOI: 10.1194/jlr.RA119000565.
- [186] Toyoshi Fujimoto and Ingela Parmryd. “Interleaflet coupling, pinning, and leaflet asymmetry-major players in plasma membrane nanodomain formation”. *Front. Cell Dev. Biol.* 4.1 (Jan. 2017), p. 155. DOI: 10.3389/fcell.2016.00155.

- [187] Dima Bolmatov, Jacob J. Kinnun, John Katsaras, and Maxim O. Lavrentovich. “Phonon-mediated lipid raft formation in biological membranes”. *Chem. Phys. Lipids* 232 (2020), p. 104979. DOI: <https://doi.org/10.1016/j.chemphyslip.2020.104979>.
- [188] Svetlana Baoukina, Eduardo Mendez-Villuendas, and D. Peter Tieleman. “Molecular view of phase coexistence in lipid monolayers”. *J. Am. Chem. Soc.* 134.42 (Oct. 2012), pp. 17543–17553. DOI: 10.1021/ja304792p.
- [189] Timothy S. Carpenter, Cesar A. López, Christopher Neale, Cameron Montour, Helgi I. Ingólfsson, Francesco Di Natale, Felice C. Lightstone, Sandrasegaram Gnanakaran, Cesar A Lo, Christopher Neale, Cameron Montour, Helgi I Ingólfsson, Francesco Di Natale, Felice C. Lightstone, Sandrasegaram Gnanakaran, Cesar A. López, Christopher Neale, Cameron Montour, Helgi I. Ingólfsson, Francesco Di Natale, Felice C. Lightstone, and Sandrasegaram Gnanakaran. “Capturing Phase Behavior of Ternary Lipid Mixtures with a Refined Martini Coarse-Grained Force Field”. *J. Chem. Theor. Comput.* 14 (2018), p. 3. DOI: 10.1021/acs.jctc.8b00496.
- [190] Naveen Michaud-Agrawal, Elizabeth J. Denning, Thomas B. Woolf, and Oliver Beckstein. “MDAnalysis: A toolkit for the analysis of molecular dynamics simulations”. *J. Comput. Chem.* 32.10 (July 2011), pp. 2319–2327. DOI: 10.1002/jcc.21787.
- [191] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations”. In: *In Proceedings of the 15th Python in Science Conference, Austin, Texas, July 11-17, 2016; Benthall, S., Rostrup, S., Eds; SciPy, Austin, 2016*.
- [192] Paul Smith and Christian D. Lorenz. “LiPyphilic: A Python Toolkit for the Analysis of Lipid Membrane Simulations”. *Journal of Chemical Theory and Computation* 17.9 (2021), pp. 5907–5919. DOI: 10.1021/acs.jctc.1c00447.



- [193] Sébastien Buchoux. “FATSLiM: a fast and robust software to analyze MD simulations of membranes”. *Method. Biochem. Anal.* 33.1 (Aug. 2016), pp. 133–134. DOI: 10.1093/bioinformatics/btw563.
- [194] “SciPy 1.0: fundamental algorithms for scientific computing in Python”. *Nat. Methods* 17.3 (Mar. 2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2.
- [195] *HMMLearn GitHub repository*.
- [196] Paul Smith, Peter J. Quinn, and Christian D. Lorenz. “Two coexisting membrane structures are defined by lateral and transbilayer interactions between sphingomyelin and cholesterol”. *Langmuir* 36.33 (Aug. 2020), pp. 9786–9799. DOI: 10.1021/acs.langmuir.0c01237.
- [197] Vyas Ramasubramani, Bradley D. Dice, Eric S. Harper, Matthew P. Spellings, Joshua A. Anderson, and Sharon C. Glotzer. “freud: A Software Suite for High Throughput Analysis of Particle Simulation Data”. *Comput. Phys. Commun.* 254 (2020), p. 107275. DOI: <https://doi.org/10.1016/j.cpc.2020.107275>.
- [198] Pierre de Buyl. “tidynamics: A tiny package to compute the dynamics of stochastic and molecular simulations”. *J. Open Source Softw.* 3.28 (2018), p. 877. DOI: 10.21105/joss.00877.
- [199] Ruo-Xu Gu, Svetlana Baoukina, and D. Peter Tieleman. “Cholesterol Flip-Flop in Heterogeneous Membranes”. *J. Chem. Theory Comput.* 15.3 (2019), pp. 2064–2070. DOI: 10.1021/acs.jctc.8b00933.
- [200] H. Jelger Risselada and Siewert J. Marrink. “The molecular face of lipid rafts in model membranes”. *Proc. Natl. Acad. Sci. U.S.A* 105.45 (Nov. 2008), pp. 17367–17372. DOI: 10.1073/pnas.0807527105.
- [201] “The Affinity of Sterols for Different Phospholipid Classes and Its Impact on Lateral Segregation”. *Biophys. J* 116.2 (Jan. 2019), pp. 296–307. DOI: 10.1016/j.bpj.2018.11.3135.

- [202] Thomas K.M. Nyholm, Oskar Engberg, Victor Hautala, Hiroshi Tsuchikawa, Kai Lan Lin, Michio Murata, and J. Peter Slotte. “Impact of Acyl Chain Mismatch on the Formation and Properties of Sphingomyelin-Cholesterol Domains”. *Biophys. J.* 117.9 (Nov. 2019), pp. 1577–1588. DOI: 10.1016/j.bpj.2019.09.025.
- [203] Oskar Engberg, Kai Lan Lin, Victor Hautala, J. Peter Slotte, and Thomas K.M. Nyholm. “Sphingomyelin Acyl Chains Influence the Formation of Sphingomyelin- and Cholesterol-Enriched Domains”. *Biophys. J.* 119.5 (Sept. 2020), pp. 913–923. DOI: 10.1016/j.bpj.2020.07.014.
- [204] Naveen Mohideen, Michael D. Weiner, and Gerald W. Feigenson. “Bilayer compositional asymmetry influences the nanoscopic to macroscopic phase domain size transition”. *Chem. Phys. Lipids* 232 (Oct. 2020), p. 104972. DOI: 10.1016/j.chemphyslip.2020.104972.
- [205] Matti Javanainen, Hector Martinez-Seara, and Ilpo Vattulainen. “Nanoscale Membrane Domain Formation Driven by Cholesterol”. *Sci. Rep.* 7.1 (Dec. 2017), pp. 1–10. DOI: 10.1038/s41598-017-01247-9.
- [206] Soohyung Park and Wonpil Im. “Analysis of Lipid Order States and Domains in Lipid Bilayer Simulations”. *J. Chem. Theory Comput.* 15 (2018), pp. 688–697. DOI: 10.1021/acs.jctc.8b00828.
- [207] Erdinc Sezgin, Ilya Levental, Michal Grzybek, Günter Schwarzmann, Veronika Mueller, Alf Honigmann, Vladimir N. Belov, Christian Eggeling, Ünal Coskun, Kai Simons, and Petra Schwille. “Partitioning, diffusion, and ligand binding of raft lipid analogs in model and cellular plasma membranes”. *Biochim. Biophys. Acta, Biomembr.* 1818.7 (2012), pp. 1777–1784. DOI: <https://doi.org/10.1016/j.bbamem.2012.03.007>.
- [208] Akash Das, Joseph L. Goldstein, Donald D. Anderson, Michael S. Brown, and Arun Radhakrishnan. “Use of mutant 125I-Perfringolysin O to probe transport and organization of cholesterol in membranes of animal cells”. *Proc. Natl. Acad. Sci. U.S.A* 110.26 (June 2013), pp. 10580–10585. DOI: 10.1073/pnas.1309273110.

- [209] Kouichi Funato, Howard Riezman, and Manuel Muñiz. “Vesicular and non-vesicular lipid export from the ER to the secretory pathway”. *Biochim Biophys Acta Mol Cell Biol Lipids* 1865.1 (Jan. 2020), p. 158453. DOI: 10.1016/j.bbalip.2019.04.013.
- [210] Rodney Elwood Infante and Arun Radhakrishnan. “Continuous transport of a small fraction of plasma membrane cholesterol to endoplasmic reticulum regulates total cellular cholesterol”. *eLife* 6 (Apr. 2017), e25466. DOI: 10.7554/eLife.25466.
- [211] J. Peter Slotte. “The importance of hydrogen bonding in sphingomyelin’s membrane interactions with co-lipids”. *Biochim Biophys Acta Biomembr* 1858.2 (Feb. 2016), pp. 304–310. DOI: 10.1016/j.bbamem.2015.12.008.
- [212] Motohide Murate and Toshihide Kobayashi. “Revisiting transbilayer distribution of lipids in the plasma membrane”. *Chem. Phys. Lipids* 194 (Jan. 2016), pp. 58–71. DOI: 10.1016/j.chemphyslip.2015.08.009.
- [213] Ilya Levental, Kandice R. Levental, and Frederick A. Heberle. *Lipid Rafts: Controversies Resolved, Mysteries Remain*. May 2020. DOI: 10.1016/j.tcb.2020.01.009.
- [214] Benjamin L. Stottrup, Juan TigreLazo, Vision B. Bagonza, Joan C. Kunz, and Joseph A. Zasadzinski. “Comparison of Line Tension Measurement Methods for Lipid Monolayers at Liquid–Liquid Coexistence”. *Langmuir* 35.48 (Dec. 2019), pp. 16053–16061. DOI: 10.1021/acs.langmuir.9b01696.
- [215] Jana Gallová, Daniela Uhríková, Norbert Kučerka, Miroslava Svorková, Sergio S. Funari, Tatiana N. Murugova, László Almásy, Milan Mazúr, and Pavol Balgavý. “Influence of cholesterol and  $\beta$ -sitosterol on the structure of EYPC bilayers”. *J. Membrane Biol.* 243.1-3 (Oct. 2011), pp. 1–13. DOI: 10.1007/s00232-011-9387-1.
- [216] Frederick A. Heberle, Robin S. Petruzielo, Jianjun Pan, Paul Drazba, Norbert Kučerka, Robert F. Standaert, Gerald W. Feigenson, and John Katsaras. “Bilayer Thickness Mismatch Controls Domain Size in Model Membranes”. *J. Am. Chem. Soc* 135.18 (May 2013), pp. 6853–6859. DOI: 10.1021/ja3113615.

- [217] Michal Belička, Anna Weitzer, and Georg Pabst. “High-resolution structure of coexisting nanoscopic and microscopic lipid domains”. *Soft Matter* 13.9 (2017), pp. 1823–1833. DOI: 10.1039/c6sm02727j.
- [218] D.W. Allender, A.J. Sodt, and M. Schick. “Cholesterol-Dependent Bending Energy Is Important in Cholesterol Distribution of the Plasma Membrane”. *Biophys. J.* 116.12 (June 2019), pp. 2356–2366. DOI: 10.1016/j.bpj.2019.03.028.
- [219] Sagar A. Pandit, S. Vasudevan, S.W. Chiu, R. Jay Mashl, Eric Jakobsson, and H.L. Scott. “Sphingomyelin-cholesterol domains in phospholipid membranes: Atomistic simulation”. *Biophys. J.* 87.2 (2004), pp. 1092–1100. DOI: 10.1529/biophysj.104.041939.
- [220] David Regan, Joseph Williams, Paola Borri, and Wolfgang Langbein. “Lipid Bilayer Thickness Measured by Quantitative DIC Reveals Phase Transitions and Effects of Substrate Hydrophilicity”. *Langmuir* 35.43 (Oct. 2019), pp. 13805–13814. DOI: 10.1021/acs.langmuir.9b02538.
- [221] S. Munro. “An investigation of the role of transmembrane domains in Golgi protein retention.” *EMBOJ* 14.19 (Oct. 1995), pp. 4695–4704. DOI: 10.1002/j.1460-2075.1995.tb00151.x.
- [222] George A. Khelashvili and H.L. Scott. “Combined Monte Carlo and molecular dynamics simulation of hydrated 18:0 sphingomyelin-cholesterol lipid bilayers”. *J. Chem. Phys.* 120.20 (May 2004), pp. 9841–9847. DOI: 10.1063/1.1724814.
- [223] Peter J. Quinn and Claude Wolf. “Egg-Sphingomyelin and Cholesterol Form a Stoichiometric Molecular Complex in Bilayers of Egg-Phosphatidylcholine”. *J. Phys. Chem. B* 114.47 (Dec. 2010), pp. 15536–15545. DOI: 10.1021/jp107490a.
- [224] Peter J. Quinn. “Structure of Sphingomyelin Bilayers and Complexes with Cholesterol Forming Membrane Rafts”. *Langmuir* 29.30 (July 2013), pp. 9447–9456. DOI: 10.1021/la4018129.
- [225] Juyang Huang and Gerald W. Feigenson. “A microscopic interaction model of maximum solubility of cholesterol in lipid bilayers”. *Biophys. J.* 76.4 (1999), pp. 2142–2157. DOI: 10.1016/S0006-3495(99)77369-8.

- [226] E. Towns-Andrews, A. Berry, J. Bordas, G.R. Mant, P.K. Murray, K. Roberts, I. Sumner, J.S. Worgan, R. Lewis, and A. Gabriel. “Time-resolved x-ray diffraction station: X-ray optics, detectors, and data acquisition”. *Rev. Sci. Instrum.* 60.7 (July 1989), pp. 2346–2349. DOI: 10.1063/1.1140765.
- [227] Peter J. Quinn. “Sphingolipid symmetry governs membrane lipid raft structure”. *Biochim Biophys Acta Biomembr* 1838.7 (2014), pp. 1922–1930. DOI: 10.1016/j.bbamem.2014.02.021.
- [228] Emilia L. Wu, Xi Cheng, Sunhwan Jo, Huan Rui, Kevin C. Song, Eder M. Dávila-Contreras, Yifei Qi, Jumin Lee, Viviana Monje-Galvan, Richard M. Venable, Jeffery B. Klauda, and Wonpil Im. “CHARMM-GUI membrane builder toward realistic biological membrane simulations”. *J. Comput. Chem.* 35.27 (2014). DOI: 10.1002/jcc.23702.
- [229] Jumin Lee, Xi Cheng, Jason M. Swails, Min Sun Yeom, Peter K. Eastman, Justin A. Lemkul, Shuai Wei, Joshua Buckner, Jong Cheol Jeong, Yifei Qi, Sunhwan Jo, Vijay S. Pande, David A. Case, Charles L. Brooks, Alexander D. MacKerell, Jeffery B. Klauda, and Wonpil Im. “CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field”. *J. Chem. Theory Comput.* 12.1 (2016). DOI: 10.1021/acs.jctc.5b00935.
- [230] David Van Der Spoel, Erik Lindahl, Berk Hess, Gerrit Groenhof, Alan E. Mark, and Herman J.C. Berendsen. “GROMACS: Fast, flexible, and free”. *J. Comput. Chem.* 26.16 (2005). DOI: 10.1002/jcc.20291.
- [231] Richard Gowers, Max Linke, Jonathan Barnoud, Tyler Reddy, Manuel Melo, Sean Seyler, Jan Domański, David Dotson, Sébastien Buchoux, Ian Kenney, and Oliver Beckstein. “MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations”. In: *Proceedings of the 15th Python in Science Conference*. SciPy, 2016, pp. 98–105. DOI: 10.25080/majora-629e541a-00e.
- [232] Fabian Pedregosa, Gael Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron

- Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Edouard Duchesnay. “Scikit-learn: Machine learning in Python”. *J. Mach. Learn. Res.* 12 (Oct. 2011), pp. 2825–2830.
- [233] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J van der Walt, Matthew Brett, Joshua Wilson, K Jarrod Millman, Nikolay Mayorov, Andrew R J Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E A Quintero, Charles R Harris, Anne M Archibald, Antônio H Ribeiro, Fabian Pedregosa, and Paul van Mulbregt. “SciPy 1.0: fundamental algorithms for scientific computing in Python”. *Nat. Methods* 17 (2020), p. 261272. DOI: 10.1038/s41592-019-0686-2.
- [234] Alexander J. Sodt, Richard W. Pastor, and Edward Lyman. “Hexagonal Substructure and Hydrogen Bonding in Liquid-Ordered Phases Containing Palmitoyl Sphingomyelin”. *Biophys. J.* 109.5 (Sept. 2015), pp. 948–955. DOI: 10.1016/j.bpj.2015.07.036.
- [235] Soohyung Park and Wonpil Im. “Analysis of Lipid Order States and Domains in Lipid Bilayer Simulations”. *J. Chem. Theory Comput.* 15 (2019), pp. 688–697. DOI: 10.1021/acs.jctc.8b00828.
- [236] Alan Grossfield, Paul N. Patrone, Daniel R. Roe, Andrew J. Schultz, Daniel Siderius, and Daniel M. Zuckerman. “Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]”. *LiveCoMS* 1.1 (Oct. 2019), p. 5067. DOI: 10.33011/livecoms.1.1.5067.
- [237] C H Rycroft. “VORO++: a three-dimensional voronoi cell library in c++”. *Chaos* 19.4 (2009), p. 4111. DOI: 10.1063/1.3215722.
- [238] Paul Smith, Robert M. Ziolk, Elena Gazzarrini, Dylan M. Owen, and Christian D. Lorenz. “On the interaction of hyaluronic acid with synovial fluid lipid

- membranes”. *Phys. Chem. Chem. Phys.* 21.19 (Sept. 2019), pp. 9845–9857. DOI: 10.1039/c9cp01532a.
- [239] Zhancheng Zhang, Shreyas Y. Bhide, and Max L. Berkowitz. “Molecular dynamics simulations of bilayers containing mixtures of sphingomyelin with cholesterol and phosphatidylcholine with cholesterol”. *J. Phys. Chem. B* 11.44 (Nov. 2007), pp. 12888–12897. DOI: 10.1021/jp074037i.
- [240] Hector Martinez-Seara, Tomasz Róg, Mikko Karttunen, Ilpo Vattulainen, and Ramon Reigada. “Cholesterol Induces Specific Spatial and Orientational Order in Cholesterol/Phospholipid Membranes”. *PLoS ONE* 5.6 (June 2010). Ed. by Neeraj Vij, e11162. DOI: 10.1371/journal.pone.0011162.
- [241] Felix Leeb and Lutz Maibaum. “Spatially Resolving the Condensing Effect of Cholesterol in Lipid Bilayers”. *Biophys. J.* 115.11 (Dec. 2018), pp. 2179–2188. DOI: 10.1016/j.bpj.2018.10.024.
- [242] Leland McInnes, John Healy, and James Melville. *Using UMAP for Clustering*.
- [243] Pavel Buslaev, Valentin Gordeliy, Sergei Grudinin, and Ivan Gushchin. “Principal Component Analysis of Lipid Molecule Conformational Changes in Molecular Dynamics Simulations”. *J. Comput. Chem.* 12 (2016), pp. 1019–1028. DOI: 10.1021/acs.jctc.5b01106.
- [244] John F. Nagle, Pierre Cognet, Fernando G. Dupuy, and Stephanie Tristram-Nagle. “Structure of gel phase DPPC determined by X-ray diffraction”. *Chem. Phys. Lipids* 218 (Jan. 2019), pp. 168–177. DOI: 10.1016/j.chemphyslip.2018.12.011.
- [245] Eric Wang and Jeffery B. Klauda. “Examination of Mixtures Containing Sphingomyelin and Cholesterol by Molecular Dynamics Simulations”. *J. Phys. Chem. B* 121.18 (May 2017), pp. 4833–4844. DOI: 10.1021/acs.jpcc.7b01832.
- [246] Tomokazu Yasuda, Nobuaki Matsumori, Hiroshi Tsuchikawa, Max Lönnfors, Thomas K.M. Nyholm, J. Peter Slotte, and Michio Murata. “Formation of Gel-like Nanodomains in Cholesterol-Containing Sphingomyelin or Phosphatidylcholine Binary Membrane As Examined by Fluorescence Lifetimes and 2H

- NMR Spectra”. *Langmuir* 31.51 (Dec. 2015), pp. 13783–13792. DOI: 10.1021/acs.langmuir.5b03566.
- [247] Yo Yano, Shinya Hanashima, Tomokazu Yasuda, Hiroshi Tsuchikawa, Nobuaki Matsumori, Masanao Kinoshita, Md Abdullah Al Sazzad, J. Peter Slotte, and Michio Murata. “Sphingomyelin Stereoisomers Reveal That Homophilic Interactions Cause Nanodomain Formation”. *Biophys. J.* 115.8 (Oct. 2018), pp. 1530–1540. DOI: 10.1016/j.bpj.2018.08.042.
- [248] Peter J. Quinn and Claude Wolf. “An X-ray diffraction study of model membrane raft structures”. *FEBS J.* 277.22 (Nov. 2010), pp. 4685–4698. DOI: 10.1111/j.1742-4658.2010.07875.x.
- [249] Max Lönnfors, Jacques P.F. Doux, J. Antoinette Killian, Thomas K.M. Nyholm, and J. Peter Slotte. “Sterols have higher affinity for sphingomyelin than for phosphatidylcholine bilayers even at equal Acyl-chain order”. *Biophys. J.* 100.11 (June 2011), pp. 2633–2641. DOI: 10.1016/j.bpj.2011.03.066.
- [250] Thomas K.M. Nyholm, Shishir Jaikishan, Oskar Engberg, Victor Hautala, and J. Peter Slotte. “The Affinity of Sterols for Different Phospholipid Classes and Its Impact on Lateral Segregation”. *Biophys. J.* 116.2 (Jan. 2019), pp. 296–307. DOI: 10.1016/j.bpj.2018.11.3135.
- [251] Tomasz Róg and Marta Pasenkiewicz-Gierula. “Cholesterol-sphingomyelin interactions: A molecular dynamics simulation study”. *Biophys. J.* 91.10 (2006), pp. 3756–3767. DOI: 10.1529/biophysj.106.080887.
- [252] Shinya Hanashima, Kazuhiro Murakami, Michihiro Yura, Yo Yano, Yuichi Umegawa, Hiroshi Tsuchikawa, Nobuaki Matsumori, Sangjae Seo, Wataru Shinoda, and Michio Murata. “Cholesterol-Induced Conformational Change in the Sphingomyelin Headgroup”. *Biophys. J.* 117.2 (July 2019), pp. 307–318. DOI: 10.1016/j.bpj.2019.06.019.
- [253] Drew Marquardt, Barbara Geier, and Georg Pabst. “Asymmetric Lipid Membranes: Towards More Realistic Model Systems”. *Membranes* 5 (2015), pp. 180–196. DOI: 10.3390/membranes5020180.



- [254] Timothée Rivel, Christophe Ramseyer, and Semen Yesylevskyy. “The asymmetry of plasma membranes and their cholesterol content influence the uptake of cisplatin”. *Sci. Rep.* 9.1 (Dec. 2019), pp. 1–14. DOI: 10.1038/s41598-019-41903-w.
- [255] Jonathan D Nickels, Jeremy C Smith, and Xiaolin Cheng. “Lateral organization, bilayer asymmetry, and inter-leaflet coupling of biological membranes”. *Chem. Phys. Lipids* 192 (2015), pp. 87–99. DOI: 10.1016/j.chemphyslip.2015.07.012.
- [256] T.D. Romo and A. Grossfield. “LOOS: An extensible platform for the structural analysis of simulations”. In: *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2009, pp. 2332–2335. DOI: 10.1109/IEMBS.2009.5335065.
- [257] Harsh Bhatia, Helgi I. Ingólfsson, Timothy S. Carpenter, Felice C. Lightstone, and Peer-Timo Bremer. “MemSurfer: A Tool for Robust Computation and Characterization of Curved Membranes”. *J. Chem. Theory Comput.* 15.11 (2019). PMID: 31564100, pp. 6411–6421. DOI: 10.1021/acs.jctc.9b00453.
- [258] Vivien Walter, Céline Ruscher, Olivier Benzerara, and Fabrice Thalmann. “MLLPA: A Machine Learning-assisted Python module to study phase-specific events in lipid membranes”. *J. Comput. Chem.* 42.13 (2021), pp. 930–943. DOI: <https://doi.org/10.1002/jcc.26508>.
- [259] Besian I Sejdiu and D Peter Tieleman. “ProLint: a web-based framework for the automated data analysis and visualization of lipid–protein interactions”. *Nucleic Acids Res.* 49.W1 (May 2021), W544–W550. DOI: 10.1093/nar/gkab409.
- [260] Wanling Song, Robin A. Corey, Bertie Ansell, Keith Cassidy, Michael Horrell, Anna Duncan, Phillip James Stansfeld, and Mark Sansom. “PyLipID: A Python package for analysis of protein-lipid interactions from MD simulations”. *bioRxiv* (2021). 2021.07.14.452312.
- [261] Hernán E. Grecco, Malte Schmick, and Philippe I.H. Bastiaens. “Signaling from the living plasma membrane”. *Cell* 144.6 (Mar. 2011), pp. 897–909. DOI: 10.1016/j.cell.2011.01.029.

- [262] Hannah Sunshine and Maria Luisa Iruela-Arispe. “Membrane lipids and cell signaling”. *Curr. Opin. Lipidol.* 28.5 (Oct. 2017), pp. 408–413. DOI: 10.1097/MOL.000000000000443.
- [263] Tod D. Romo, Nicholas Leioatts, and Alan Grossfield. “Lightweight object oriented structure analysis: Tools for building tools to analyze molecular dynamics simulations”. *J. Comput. Chem.* 35.32 (2014), pp. 2305–2318. DOI: <https://doi.org/10.1002/jcc.23753>.
- [264] D. Peter Tieleman and Siewert-Jan Marrink. “Lipids Out of Equilibrium: Energetics of Desorption and Pore Mediated Flip-Flop”. *J. Am. Chem. Soc.* 128.38 (2006), pp. 12462–12467. DOI: 10.1021/ja062432i.
- [265] Andrey A. Gurtovenko and Ilpo Vattulainen. “Molecular Mechanism for Lipid Flip-Flops”. *J. Phys. Chem. B.* 111.48 (2007), pp. 13554–13559. DOI: 10.1021/jp077094k.
- [266] Tomasz Róg, Lorna M. Stimson, Marta Pasenkiewicz-Gierula, Ilpo Vattulainen, and Mikko Karttunen. “Replacing the cholesterol hydroxyl group with the ketone group facilitates sterol flip-flop and promotes membrane fluidity”. *J. Phys. Chem. B.* 112.7 (Feb. 2008), pp. 1946–1952. DOI: 10.1021/jp075078h.
- [267] H. Jelger Risselad and Siewert J. Marrink. “The molecular face of lipid rafts in model membranes”. *Proc. Natl. Acad. Sci. U.S.A* 105.45 (Nov. 2008), pp. 17367–17372. DOI: 10.1073/pnas.0807527105.
- [268] Norbert Kučerka, Jason D. Perlmutter, Jianjun Pan, Stephanie Tristram-Nagle, John Katsaras, and Jonathan N. Sachs. “The effect of cholesterol on short- and long-chain monounsaturated lipid bilayers as determined by molecular dynamics simulations and X-ray scattering”. *Biophys. J.* 95.6 (Sept. 2008), pp. 2792–2805. DOI: 10.1529/biophysj.107.122465.
- [269] W. F. Drew Bennett, Justin L. MacCallum, Marlon J. Hinner, Siewert J. Marrink, and D. Peter Tieleman. “Molecular view of cholesterol flip-flop and chemical potential in different membrane environments”. *J. Am. Chem. Soc.* 131.35 (Sept. 2009), pp. 12714–12720. DOI: 10.1021/ja903529f.

- [270] W. F. Drew Bennett and D. Peter Tieleman. “Molecular simulation of rapid translocation of cholesterol, diacylglycerol, and ceramide in model raft and non-raft membranes”. *J. Lipid Res.* 53.3 (Mar. 2012), pp. 421–429. DOI: 10.1194/jlr.M022491.
- [271] Fumiko Ogushi, Reiko Ishitsuka, Toshihide Kobayashi, and Yuji Sugita. “Rapid flip-flop motions of diacylglycerol and ceramide in phospholipid bilayers”. *Chem. Phys. Lett.* 522 (Jan. 2012), pp. 96–102. DOI: 10.1016/j.cplett.2011.11.057.
- [272] Amit Choubey, Rajiv K. Kalia, Noah Malmstadt, Aiichiro Nakano, and Priya Vashishta. “Cholesterol translocation in a phospholipid membrane”. *Biophys. J.* 104.11 (June 2013), pp. 2429–2436. DOI: 10.1016/j.bpj.2013.04.036.
- [273] Drew Marquardt, Norbert Kučerka, Stephen R. Wassall, Thad A. Harroun, and John Katsaras. “Cholesterol’s location in lipid bilayers”. *Chem. Phys. Lipids* 199 (Sept. 2016), pp. 17–25. DOI: 10.1016/j.chemphyslip.2016.04.001.
- [274] Matti Javanainen and Hector Martinez-Seara. “Rapid diffusion of cholesterol along polyunsaturated membranes: Via deep dives”. *Phys. Chem. Chem. Phys.* 21.22 (June 2019), pp. 11660–11669. DOI: 10.1039/c9cp02022e.
- [275] Markus S. Miettinen and Reinhard Lipowsky. “Bilayer Membranes with Frequent Flip-Flops Have Tensionless Leaflets”. *Nano Lett.* 19.8 (Aug. 2019), pp. 5011–5016. DOI: 10.1021/acs.nanolett.9b01239.
- [276] Younghoon Oh and Bong June Sung. “Facilitated and Non-Gaussian Diffusion of Cholesterol in Liquid Ordered Phase Bilayers Depends on the Flip-Flop and Spatial Arrangement of Cholesterol”. *J. Phys. Chem. Lett.* 9.22 (Nov. 2018), pp. 6529–6535. DOI: 10.1021/acs.jpcllett.8b02982.
- [277] Mohammadreza Aghaaminiha, Amir M. Farnoud, and Sumit Sharma. “Quantitative relationship between cholesterol distribution and ordering of lipids in asymmetric lipid bilayers”. *Soft Matter* 17.10 (Mar. 2021), pp. 2742–2752. DOI: 10.1039/d0sm01709d.

- [278] Hugo A L Filipe, Matti Javanainen, Armindo Salvador, Adelino M Galva, Ilpo Vattulainen, Luıs, M S Loura, and Maria Joa O Moreno. “Quantitative Assessment of Methods Used To Obtain Rate Constants from Molecular Dynamics Simulations—Translocation of Cholesterol across Lipid Bilayers”. *J. Chem. Theory Comput* 14 (2018), p. 20. DOI: 10.1021/acs.jctc.8b00150.
- [279] James W. Carter, Miguel A. Gonzalez, Nicholas J. Brooks, John M. Seddon, and Fernando Bresme. “Flip-flop asymmetry of cholesterol in model membranes induced by thermal gradients”. *Soft Matter* 16 (25 2020), pp. 5925–5932. DOI: 10.1039/D0SM00546K.
- [280] Alexander J. Sodt, Michael Logan Sandar, Klaus Gawrisch, Richard W. Pastor, and Edward Lyman. “The molecular structure of the liquid-ordered phase of lipid bilayers”. *J. Am. Chem. Soc.* 136.2 (Jan. 2014), pp. 725–732. DOI: 10.1021/ja4105667.
- [281] David G. Ackerman and Gerald W. Feigenson. “Multiscale modeling of four-component lipid mixtures: Domain composition, size, alignment, and properties of the phase interface”. *J. Phys. Chem. B.* 119.11 (Mar. 2015), pp. 4240–4250. DOI: 10.1021/jp511083z.
- [282] Xubo Lin, Joseph H. Lorent, Allison D. Skinkle, Kandice R. Levental, M. Neal Waxham, Alemayehu A. Gorfe, and Ilya Levental. “Domain stability in biomimetic membranes driven by lipid polyunsaturation”. *J. Phys. Chem. B.* 120.46 (Nov. 2016), pp. 11930–11941. DOI: 10.1021/acs.jpccb.6b06815.
- [283] “Nonadditive Compositional Curvature Energetics of Lipid Bilayers”. *Phys. Rev. Lett.* 117.13 (Sept. 2016), p. 138104. DOI: 10.1103/PhysRevLett.117.138104.
- [284] Kenichiro Koshiyama, Masaki Taneo, Taiki Shigematsu, and Shigeo Wada. “Bicelle-to-Vesicle Transition of a Binary Phospholipid Mixture Guided by Controlled Local Lipid Compositions: A Molecular Dynamics Simulation Study”. *J. Phys. Chem. B.* 123.14 (Apr. 2019), pp. 3118–3123. DOI: 10.1021/acs.jpccb.8b10682.

- [285] Jason D. Perlmutter and Jonathan N. Sachs. “Interleaflet interaction and asymmetry in phase separated lipid bilayers: Molecular dynamics simulations”. *J. Am. Chem. Soc.* 133.17 (May 2011), pp. 6563–6577. DOI: 10.1021/ja106626r.
- [286] Philip W. Fowler, John J. Williamson, Mark S.P. Sansom, and Peter D. Olmsted. “Roles of Interleaflet Coupling and Hydrophobic Mismatch in Lipid Membrane Phase-Separation Kinetics”. *J. Am. Chem. Soc.* 138.36 (Sept. 2016), pp. 11633–11642. DOI: 10.1021/jacs.6b04880.
- [287] Ramon Reigada. “Alteration of interleaflet coupling due to compounds displaying rapid translocation in lipid membranes”. *Sci. Rep.* 6.1 (Sept. 2016), pp. 1–15. DOI: 10.1038/srep32934.
- [288] Timur R. Galimzyanov, Peter I. Kuzmin, Peter Pohl, and Sergey A. Akimov. “Undulations Drive Domain Registration from the Two Membrane Leaflets”. *Biophys. J.* 112.2 (Jan. 2017), pp. 339–345. DOI: 10.1016/j.bpj.2016.12.023.
- [289] Michael D. Weiner and Gerald W. Feigenson. “Molecular Dynamics Simulations Reveal Leaflet Coupling in Compositionally Asymmetric Phase-Separated Lipid Membranes”. *J. Phys. Chem. B.* 123.18 (May 2019), pp. 3968–3975. DOI: 10.1021/acs.jpcc.9b03488.
- [290] Siya Zhang and Xubo Lin. “Lipid Acyl Chain cis Double Bond Position Modulates Membrane Domain Registration/Anti-Registration”. *J. Am. Chem. Soc.* 141.40 (2019). PMID: 31532653, pp. 15884–15890. DOI: 10.1021/jacs.9b06977.
- [291] Maria J. Sarmiento, Martin Hof, and Radek Šachl. “Interleaflet Coupling of Lipid Nanodomains – Insights From in vitro Systems”. *Front. Cell Dev. Biol.* 8 (Apr. 2020), p. 284. DOI: 10.3389/fcell.2020.00284.
- [292] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. “Array

- programming with NumPy”. *Nature* 585.7825 (Sept. 2020), pp. 357–362. DOI: 10.1038/s41586-020-2649-2.
- [293] Michael D. Weiner and Gerald W. Feigenson. “Presence and Role of Midplane Cholesterol in Lipid Bilayers Containing Registered or Antiregistered Phase Domains”. *J. Phys. Chem. B* 122 (2018), p. 21. DOI: 10.1021/acs.jpcc.8b03949.
- [294] Milka Doktorova, Jessica L. Symons, and Ilya Levental. “Structural and functional consequences of reversible lipid asymmetry in living membranes”. *Nat. Chem. Biol.* 16.12 (Dec. 2020), pp. 1321–1330. DOI: 10.1038/s41589-020-00688-0.
- [295] Harsh Bhatia and Peer-Timo Bremer. *Example usage of Memsurfer - a bilayer membrane consisting of three types of lipid*. (accessed April 9, 2021).
- [296] Riccardo Capelli, Andrea Gardin, Charly Empereur-mot, Giovanni Doni, and Giovanni M. Pavan. “A Data-Driven Dimensionality Reduction Approach to Compare and Classify Lipid Force Fields”. *J. Phys. Chem. B* 125.28 (2021), pp. 7785–7796. DOI: 10.1021/acs.jpcc.1c02503.
- [297] Heidi Koldsø, David Shorthouse, Jean Hélie, and Mark S.P. Sansom. “Lipid Clustering Correlates with Membrane Curvature as Revealed by Molecular Simulations of Complex Lipid Bilayers”. *PLoS Comput. Biol.* 10.10 (Oct. 2014). Ed. by Cécile Fradin, e1003911. DOI: 10.1371/journal.pcbi.1003911.
- [298] Helgi I. Ingólfsson, Harsh Bhatia, Talia Zeppelin, W. F.Drew Bennett, Kristy A. Carpenter, Pin Chia Hsu, Gautham Dharuman, Peer Timo Bremer, Birgit Schiøtt, Felice C. Lightstone, and Timothy S. Carpenter. “Capturing Biologically Complex Tissue-Specific Membranes at Different Levels of Compositional Complexity”. *J. Phys. Chem. B* 124.36 (Sept. 2020), pp. 7819–7829. DOI: 10.1021/acs.jpcc.0c03368.
- [299] Katie A. Wilson, Hugo I. MacDermott-Opeskin, Eden Riley, Yiechang Lin, and Megan L. O’Mara. “Understanding the Link between Lipid Diversity and the Biophysical Properties of the Neuronal Plasma Membrane”. *Biochemistry* 59.33 (Aug. 2020), pp. 3010–3018. DOI: 10.1021/acs.biochem.0c00524.

- [300] Katie A. Wilson, Stephen J. Fairweather, Hugo I. Macdermott-Opeskin, Lily Wang, Richard A. Morris, and Megan L. O'mara. "The role of plasmalogens, Forssman lipids, and sphingolipid hydroxylation in modulating the biophysical properties of the epithelial plasma membrane". *J. Chem. Phys.* 154.9 (Mar. 2021), p. 095101. DOI: 10.1063/5.0040887.
- [301] May Yee Sze, Richard J. Gillams, Sylvia E. McLain, and Christian D. Lorenz. "Effects of lipid heterogeneity on model human brain lipid membranes". *Soft Matter* 17.1 (Jan. 2021), pp. 126–135. DOI: 10.1039/dosmo1766c.
- [302] Sandro Sonnino, Laura Mauri, Vanna Chigorno, and Alessandro Prinetti. "Gangliosides as components of lipid membrane domains". *Method. Enzymol.* 17.1 (Jan. 2007), 1R–13R. DOI: 10.1093/glycob/cw1052.
- [303] Alessandro Prinetti, Nicoletta Loberto, Vanna Chigorno, and Sandro Sonnino. "Glycosphingolipid behaviour in complex membranes". *Biochim. Biophys. Acta Biomembr.* 1788.1 (Jan. 2009), pp. 184–193. DOI: 10.1016/j.bbamem.2008.09.001.
- [304] Bodil Westerlund and J. Peter Slotte. "How the molecular features of glycosphingolipids affect domain formation in fluid membranes". *Biochim. Biophys. Acta Biomembr.* 1788.1 (Jan. 2009), pp. 194–201. DOI: 10.1016/j.bbamem.2008.11.010.
- [305] Sören Von Bülow, Jakob Tómas Bullerjahn, and Gerhard Hummer. "Systematic errors in diffusion coefficients from long-time molecular dynamics simulations at constant pressure". *J. Chem. Phys.* 153.2 (July 2020), p. 021101. DOI: 10.1063/5.0008316.
- [306] Shujie Fan, Max Linke, Ioannis Paraskevakos, Richard Gowers, Michael Gecht, and Oliver Beckstein. "Rostrup". In: *In Proceedings of the 18th Python in Science Conference, Austin, Texas, July 8-14, 2019; Calloway, C., Lippa, D., Niederbut, D., Shupe, D., Eds; SciPy, Austin, 2019.*
- [307] Matthew Rocklin. "Dask: Parallel computation with blocked algorithms and task scheduling". In: *In Proceedings of the 14th Python in Science Conference, Austin, Texas, July 6-12, 2015; Huff, K., Bergstra, J, Eds; SciPy, Austin, 2015.*

- [308] K Jarrod Millman and Fernando Pérez. “Developing open source scientific practice”. In: *Implementing Reproducible Research*. CRC Press, 2014, pp. 149–183.
- [309] GitHub Flow. *The best way to use Git and GitHub*. (accessed April 13, 2021).
- [310] Holger Krekel, Bruno Oliveira, Ronny Pfannschmidt, Floris Bruynooghe, Brianna Laughler, and Florian Bruhin. *pytest 6.2.2*. (accessed April 13, 2021). 2004.
- [311] pytest-cov development team. *pytest-cov 2.11.1*. (accessed April 13, 2021). 2006.
- [312] Anaconda. *Anaconda Software Distribution*. (accessed April 13, 2021).
- [313] PyPA. *pip - The Python Package Installer*. (accessed April 13, 2021).
- [314] Sebastian Thallmair, Matti Javanainen, Balázs Fábián, Hector Martinez-Seara, and Siewert J. Marrink. “Nonconverged Constraints Cause Artificial Temperature Gradients in Lipid Bilayer Simulations”. *J. Phys. Chem. B* 0.0 (0). PMID: 34398598, null. DOI: 10.1021/acs.jpcc.1c03665.
- [315] Matías R. Machado, Exequiel E. Barrera, Florencia Klein, Martín Sónora, Stefano Silva, and Sergio Pantano. “The SIRAH 2.0 Force Field: Altius, Fortius, Citius”. *J. Chem. Theory Comput.* 15.4 (2019), pp. 2719–2733. DOI: 10.1021/acs.jctc.9b00006.
- [316] Dmytro Soloviov, Yong Q. Cai, Dima Bolmatov, Alexey Suvorov, Kirill Zhernenkov, Dmitry Zav’yalov, Alexey Bosak, Hiroshi Uchiyama, and Mikhail Zhernenkov. “Functional lipid pairs as building blocks of phase-separated membranes”. *Proc. Natl. Acad. Sci. U.S.A* 117.9 (2020), pp. 4749–4757. DOI: 10.1073/pnas.1919264117.
- [317] Thomas J. Piggot, Jane R. Allison, Richard B. Sessions, and Jonathan W. Essex. “On the Calculation of Acyl Chain Order Parameters from Lipid Simulations”. *J. Chem. Theory Comput.* 13.11 (2017), pp. 5683–5696. DOI: 10.1021/acs.jctc.7b00643.



- [318] George Khelashvili, Benjamin Kollmitzer, Peter Heftberger, Georg Pabst, and Daniel Harries. “Calculating the bending modulus for multicomponent lipid membranes in different thermodynamic phases”. *J. Chem. Theory Comput.* 9.9 (Sept. 2013), pp. 3866–3871. DOI: 10.1021/ct400492e.
- [319] Niklaus Johner, Daniel Harries, and George Khelashvili. “Curvature and lipid packing modulate the elastic properties of lipid assemblies: Comparing HII and lamellar phases”. *J. Phys. Chem. Lett.* 5.23 (Dec. 2014), pp. 4201–4206. DOI: 10.1021/jz5022284.
- [320] M. Doktorova, D. Harries, and G. Khelashvili. “Determination of bending rigidity and tilt modulus of lipid membranes from real-space fluctuation analysis of molecular dynamics simulations”. *Phys. Chem. Chem. Phys.* 19.25 (June 2017), pp. 16806–16818. DOI: 10.1039/c7cp01921a.