Molecular Simulations of Supercooled Water at the Interface with Lipid Bilayers

by

Christopher Malcolm Miles

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Analytical Science

Submitted to The University of Warwick

Department of Chemistry

December, 2022

Contents

List of Tables

List of Figures

List of Abbreviations

A complete list of each abbreviation used is given below, along with the page number in which it is first used.

Abbreviation	**Meaning**	**Page**
I_h	hexagonal ice	1
HC	hexagonal cage	2
I_c	cubic ice	3
DDC	double diamond cage	3
CNT	classical nucleation theory	3
INA	ice nucleating agent	6
INP	ice nucleating particle	6
CPA	cryoprotective agent	7
DMSO	dimethyl sulphoxide	7
AFP	antifreeze protein	7
MSC	mesenchymal stem cell	8
MD	molecular dynamics	8
SPC	single point charge (water)	9
mW	monatomic water model	9
*TIP4P	rigid planar four-site water model	9
TIP4P-Ew	TIP4P* re-paramerisation for use with Ewald techniques	9
TIP4P/2005	TIP4P* re-paramerisation for condensed phases of water	9
TIP4P/Ice	TIP4P* re-paramerisation for the study of ice	9
NQE	nuclear quantum effect	9
PC	phosphatidylcholine	10
SM	sphingomyelin	10
PE	phosphatidylethanolamine	10
PI	phosphatidylinositol	10
PA	phosphatidate	10
PS	phosphatidylserine	10
PG	phosphatidylglycerol	10
DPG	diphosphatidylglycerol	10
CHL	cholesterol	10

Continued on next page

Continued from previous page

Abbreviation	**Meaning**	**Page**
DPPC	1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine	10
^{2}H NMR	deuterium nuclear magnetic resonance spectroscopy	10
ESR	electron spin resonance	10
DPPI	1,2-dipalmitoyl-*sn*-glycero-3-phosphatidylinositol	11
DMPI	1,2-dimyristoyl-*sn*-glycero-3-phosphatidylinositol	11
OM	outer membrane	12
†LPS	lipopolysaccharide	12
‡KDO	2-keto-3-deoxy-D-mannooctanoic acid	12
Hep	L-glycero-D-manno-heptose	12
Glc	D-glucose	12
Gal	D-galactose	12
ReLPS	LPS† with two KDO‡ molecules per lipid A molecule[§]	12
RaLPS	LPS† with the complete oligosaccharide core	12
GMO	glycerol-1-monooleate	14
POPC	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine	14
POPE	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine	17
POPG	1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol	17
CL	cardiolipin	17
PVCL2	1,10-palmitoyl-2,20-vacenoyl cardiolipin	17
PVPE	1-palmitoyl-2-vacenoyl-*sn*-glycero-3-phosphoethanolamine	17
PVPG	1-palmitoyl-2-vacenoyl-*sn*-glycero-3-phosphoglycerol	17
PBC	periodic boundary condition	18
NVT	fixed number of atoms, volume and temperature	18
NpT	fixed number of atoms, pressure and temperature	19
$NpAT$	fixed number of atoms, pressure, area and temperature	22
EW3DC	Ewald sum in three dimensions with a correction term	20
S_{CD}	deuterium order parameter	22
CDF	cumulative density function	30
PDF	probability density function	30
SMAC	single molecule angle criteria	35
SSR	squared sum of residuals	66
DLS	dynamic light scattering	72
µl-NIPI	microlitre by immersed particle	74
CV	collective variable	79
MT	metadynamics	79
US	umbrella sampling	80
WHAM	weighted histogram analysis method	81

Continued on next page

[§] Known as the “deep rough mutant”.

Continued from previous page

Abbreviation	**Meaning**	**Page**
HSPC	L-α-phosphatidylcholine, hydrogenated (Soy)[¶]	83
DSPC	1,2-distearoyl-*sn*-glycero-3-phosphocholine	83

[¶] Known as “Hydro Soy PC”.

Acknowledgements

I would like to acknowledge a great many people who have assisted me over the course of my doctoral studies. Firstly I must thank my supervisor, Dr Gabriele Sosso, without whom none of this would have been possible, for his guidance and knowledge during the four (and a bit) years I have studied in his group (including my MAS CDT MSc mini-project). I must also thank my secondary supervisor, Dr Ann Dixon, for the role she played in conceiving the project as well as for her experimental insight and access to her group's laboratory and resources.

In addition to my supervisors, I must thank each and every member of the Sosso group over the years for their help, collaboration and friendship. I must also thank the members of the Dixon group for their training and assistance while undertaking the experimental component of my project, in particular Dr Rhiannon Brookes, who trained me on the vesicle preparation as well as on DLS. In addition to the Sosso and Dixon groups, I would like to thank Dr Thomas Whale, who trained me on the µl-NIPI instrument and provided further experimental insight. Structural information for the lipopolysaccharide systems were provided by Dr Pin-Chia Hsu (University of Southampton) and Professor Syma Khalid (University of Oxford). I must also mention Dr Justin Tian (Queen's University Belfast) who proposed the side project covered in chapter 9 and provided experimental insight therein.

I would like to also acknowledge others who have played a role in my PhD: the MAS CDT directors Professor Pat Unwin, Professor Steven Brown and Dr Nikola Chmel, the MAS CDT and Department of Chemistry support staff and the members of my advisory committee: Professor Reinhard Maurer and Professor Corinne Smith. I would also like to acknowledge the other supervisors who I studied under during my first forays into academic research, during my undergraduate programme in the University of Warwick Mathematics Department and my MAS CDT MSc: namely Dr Saul Schleimer, Dr Daan Krammer and Dr Jérôme Charmet.

It would be remiss for me to neglect to also mention those who have supported me in a non-academic sense. First and foremost — my family, and in particular my parents. I would also thank the members of my CDT cohort with whom I formed close friendships during the MSc year; and especially those with a background in Chemistry who were able to help me navigate an MSc in Chemistry, coming from a Mathematics background. I would also mention the many friends I have made from the wider student community

at Warwick, through my involvement with the sports clubs: Warwick Motorsport and Warwick Latin and Ballroom, for a much needed escape from academia.

In additional to individual acknowledgements, this PhD was funded via an Engineering and Physical Sciences Research Council (EPSRC) studentship through the Centre for Doctoral Training in Molecular Analytical Science (MAS-CDT). High performance computing facilities were provided by the Scientific Computing Research Technology Platform (SCRTP) at the University of Warwick, Athena HPC Midlands+, funded via the EPSRC and the national supercomputing service, ARCHER, granted via the UK High-End Consortium for Biomolecular Simulation (HECBioSim).

Declaration of Original Work

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree apart from material regarding shorter DPPC-CHL simulations featuring the analyses detailed in in sections 3.1, 3.3, 3.4 and 5.1 which was previously submitted for the degree of Master of Science. Specifically, much of the work discussed in section 3.4 was taken from my MAS CDT MSc mini-project, however it is pertinent to the subsequent work I have done during the PhD and therefore I have included it.

The work presented (including data generated and data analysis) was carried out by the author except in the cases outlined below:

- Initial configurations for symmetric LPS bilayers were provided by Dr Pin-Chia Hsu (University of Southampton) and Professor Syma Khalid (University of Oxford).
- Some of the HIN analysis code used was written by other members of the Sosso group, in particular the hydrogen bonding analysis code was written by Matthew Warren.

Parts of this thesis have been published by the author:

- C. M. Miles, P.-C. Hsu, A. M. Dixon, S. Khalid and G. C. Sosso, *Phys. Chem. Chem. Phys.*, 2022, **24**, 6476–6491.

Abstract

The transition of water into ice is a fundamental natural process; however nucleation is a rare event in the context of molecular simulations and the underlying mechanism is not well understood. The formation of ice is a key issue in the context of cryopreservation, where cellular damage occurs due to extracellular ice formation at temperatures well above the homogeneous freezing point of pure water. Thus such ice formation must be driven heterogeneously. The primary intention of this thesis is to consider whether lipid bilayers, as found in cellular membranes, may be the agent driving such heterogeneous nucleation.

Via molecular dynamics simulations, a number of phospholipid and lipopolysaccharide bilayers were investigated to see what happens at the interface with supercooled liquid water. While these bilayers do appear, to varying extents, to act as ice nucleating agents, their potency as such does not appear to be strong enough to be the key facilitator in such ice nucleation. In addition to the question as to whether bilayers promote ice nucleation, the structural and chemical reasons are discussed, particularly in comparing the different bilayers with each other. This is a key advantage of molecular simulations as compared to experimental techniques.

In addition to molecular simulations, vesicles were synthesised experimentally and examined using dynamic light scattering to support the validity of the simulation setup. As an aside from the topic of ice nucleation, additional simulations were run of the permeation of small drug molecules through lipid bilayers. These simulations employed enhanced sampling techniques, and resulted in computing a free energy surface which was consistent with experimental data.

Introduction

> *"Water is life's matter and matrix, mother and medium. There is no life without water."*
>
> Albert Szent-Györgyi

The most abundant liquid on Earth and a vital component of all known life, water is one of the most fundamental compounds in the fields of life sciences and organic chemistry. In addition to being abundant and essential to life, water has a number of properties which are unlike other liquids: a well-known example being that its density is maximal in the liquid phase.[1] The nucleation process from water into ice is of great interest to a number of academic and industrial fields. In the field of atmospheric science, ice nucleation plays a key role in the formation of ice clouds and is fundamental to their modelling.[2–5] The emission of potential ice nucleating particles (INPs) from the shipping industry, particularly in arctic regions, has implications for environmental preservation.[6] The discovery of cryoprotective agents (CPAs) and antifreeze proteins (AFPs) also has wide ranging applications, from automotive coolant systems to cryopreservation.[7,8] It is cryopreservation which is the primary motivation behind this thesis. Cryopreservation is the procedure wherein biological material, such as cells, organelles and tissues,[9] is preserved via the employ of supercooling.[10] Particularly when slow-freezing, it is essential to control ice formation during the vitrification procedure[11,12] as ice re-crystallisation during this process accounts for a great deal of cellular damage, often destroying organic matter.[9,13,14]

In homogeneous systems, ice nucleation does not occur at the commonly understood freezing point of 0 °C, yet in our collective experience this is the temperature at which water freezes. In fact, the temperature of homogeneous ice nucleation[1] is 232 K: a supercooling of 41 K. What then causes extracellular ice formation during the cryopreservation process, which typically occurs at far milder supercooling than the onset of homogeneous ice nucleation? Certainly there are many impurities in the aqueous medium surrounding cells which may promote ice nucleation; but it is possible that the biological matter, which is the subject of cryopreservation, could play an active role in promoting such extracellular ice nucleation. In particular, some surfaces are known to act as ice nucleating agents,[15,16] so perhaps the interface between extracellular water and the cellular membrane could function as a site which facilitates ice nucleation.

Despite a wealth of knowledge about the interactions at the interfaces between lipid bilayers and water,[17–20] very little is known about the extent to which lipid bilayers can act as ice nucleating agents (INAs); although cholesterol, a common component of cell membranes, is known to act as an INA in its crystalline form[21–23] and as a monolayer.[24] It remains, however, an open question as to the extent to which phospholipids (which account for the majority of many cellular membranes) can facilitate the formation of ice.

This question is the primary focus of this thesis. By comparing a wide range of lipid bilayers, the effects of bilayer composition, structure and chemistry on interfacial ice nucleation will be elucidated. The following three sets of systems are studied in detail via molecular dynamics (MD) simulations:[*]

- Model cellular membranes, composed of varying quantities of 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC) and cholesterol (CHL).
- Phospholipid membranes with a negative overall charge, composed of varying quantities of DPPC and 1,2-dimyristoyl-*sn*-glycero-3-phosphatidylinositol (DMPI).
- Asymmetric phospholipid-lipopolysaccharide (LPS) membranes, representative of the outer membrane of Gram-negative bacteria. There are three such systems, one with no sugar, one (ReLPS) with two 2-keto-3-deoxy-D-mannooctanoic acid (KDO) molecules per lipid A molecule, and one (RaLPS) with the complete oligosaccharide core.

Altogether these systems exhibit a variety of chemical groups and structural trends as well as a good degree of comparability to published literature and, in the case of the LPS systems, complexity and correspondence with an actual cell membrane component.

DPPC is used as the basis for so many of these systems for a couple of reasons. From a practical perspective, DPPC (often with CHL) bilayers have been the study of many experimental[25–27] and computational[28–33] studies, this wealth of literature is highly valuable to a computational study as it allows for the simulation parameters to be thoroughly verified. DPPC is also a suitable choice for model membrane systems as phosphatidylcholines, the class of lipid to which DPPC belongs, are common constituents of cellular membranes.[34]

It was found, over the course of this work, that the systems studied show some degree of ice nucleation promotion, with some having more of an effect than others. The effects of bilayer topology are considered, as are the chemical differences between the bilayers: in particular the number and accessibility of hydrogen bonding sites, which are known to play and active role in ice formation on the surface of crystalline CHL.[21] From a structural perspective, the introduction of CHL into DPPC bilayers has an especially potent effect on the ordering, which would normally be expected to improve such bilayers' abilities to act as INAs, however at naturally occurring CHL concentrations ($< 50\,\mathrm{mol\%}$)[34–36] it is unclear whether CHL improves the nucleating effect. DMPI may be expected to be a more efficient INA than DPPC due to its greatly increased number of hydrogen bonding sites, although this was not necessarily observed herein. Finally, sugar-coated LPS systems

---

[*] See chapters 2 and 3 for more details of the systems and for structures for the lipids concerned.

appear to act as fairly potent INAs, with the sugars potentially facilitating ice nucleation via the "templating" effect, wherein a surface can act as a template for the growth of $\mathrm{I_h}$, this effect is particularly seen where there are a number of hydroxyl groups arranged with a hexagonal structure.[37,38]

In complement to MD simulations, various mixed DPPC-CHL vesicles were synthesised in the laboratory and their hydrodynamic diameters assessed via dynamic light scattering (DLS). This allows for verification that the two-dimensional simulation setup is representative of real-world bilayers. Despite limitations in the efficacy of DLS for studying rather heterogeneous solutions, the vesicles formed were clearly large enough for slab geometry to be suitable in simulating such systems.

In an aside from the work on ice nucleation, similar bilayer systems were created out of Hydro Soy PC[*] (HSPC). These were simulated at room temperature to study the free energy landscape of the permeation of small drug molecules, specifically felodipine and naproxen, through the bilayer. The enhanced sampling techniques of metadynamics (MT) and umbrella sampling (US) were used for this purpose. MT has limitations for this task and did not produce a good estimate for the free energy, however US proved more successful.

In chapters 1 and 2, relevant topics within ice nucleation and lipid bilayers are reviewed. The main subject of this thesis, that being the computational study of ice nucleation at interfaces with DPPC-CHL/DMPI and LPS bilayers, is organised into four chapters: with the simulation details and analysis methods introduced in chapters 3 to 5, and the results located in chapter 7. Following this, chapters 8 and 9 cover the experimental work and work on drug permeation through HSPC bilayers, respectively. The overall findings of the thesis are then summarised in the conclusions.

---

[*] L-α-phosphatidylcholine, hydrogenated (Soy).

Chapter 1

Ice Nucleation

1.1 The structure of ice

Defining what separates solid ice from supercooled liquid water is of obvious importance to the computational study of ice nucleation. The "ice rules" or "Bernal-Fowler rules", as laid out by Bernal and Fowler[39] as follows:

1. Each oxygen must be covalently bonded to exactly two hydrogens.
2. There must be precisely one hydrogen located roughly in-between each pair of oxygen atoms.

1.1.1 Hexagonal ice

Pauling[40] described the structure of an ideal ice crystal as follows:

1. Each oxygen atom is covalently bonded to two hydrogen atoms, with a bond length of 0.95 Å.
2. Each water molecule is surrounded by four others, oriented in such a way that its two hydrogen atoms are directed towards two of the neighbouring water molecules, forming hydrogen bonds.
3. Such arrangements are restricted by the second requirement of the ice rules, i.e. that each pair of neighbouring oxygens must have one and only one hydrogen approximately in-between. These hydrogen atoms are distanced 0.95 Å from one oxygen atom and 1.81 Å from the other.
4. For an ice crystal of N molecules, there are $(3/2)^N$ possible configurations, satisfying the above requirements.

Ice satisfying these conditions is termed hexagonal ice ($\mathrm{I_h}$) and is the basic form of ice found in nature. The oxygen atoms are arranged in a vaguely two dimensional form[*] on a

---

[*] Somewhat similar to the structure of graphite, but with the tetrahedral bond angles of 109.47° causing some deviation from the two dimensional plane. Additionally these lattices are rigidly bonded to the lattices above and below.

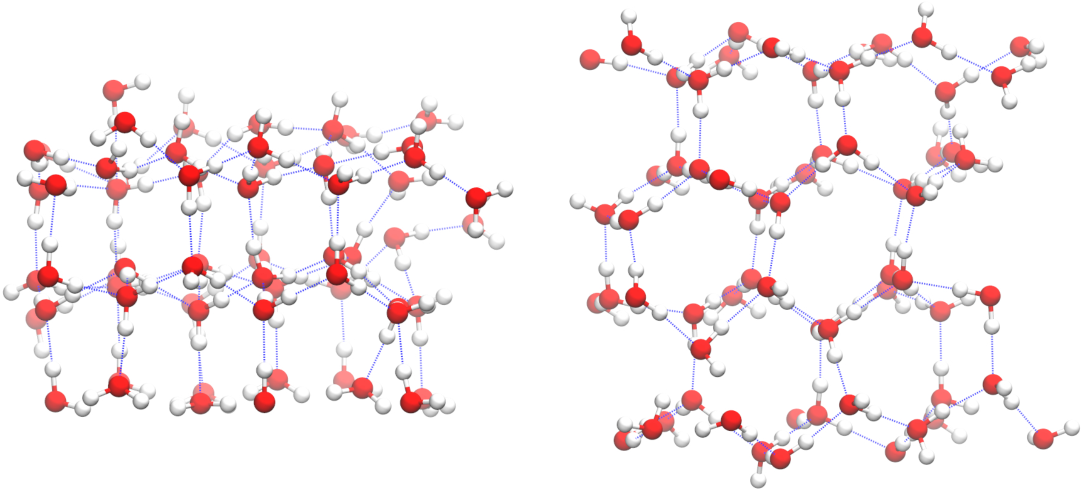


Figure 1: The crystal structure of I_h, shown via two different orientations. Oxygen atoms are coloured red, hydrogen atoms are coloured white and hydrogen bonds are displayed by dashed blue lines.

hexagonal lattice, subject to the tetrahedral angles between these water molecules, with such lattices forming a regular tetrahedral arrangement above and below.[41] This structure is composed of a system of interconnected hexagonal cages (HCs). Such a hexagonal crystal structure is illustrated in figure 1, with oxygen atoms coloured red, hydrogen atoms coloured white and hydrogen bonds displayed by dashed blue lines.

This tetrahedral arrangement cannot be perfectly ordered for both physical and geometrical reasons. One such reason is that while a perfectly ordered state is geometrically possible, and thermodynamically stable below 72 K, the orientation of each water molecule is entirely determined by the configuration of its neighbours; and therefore very difficult to attain (and extremely improbable to occur spontaneously, e.g. in nature).[41] Secondly, the H−O−H angle in a free water molecule is known to be 104.52°, and such molecules are fairly rigid, whereas the O···O···O angles in a perfectly tetrahedral structure should be 109.47°. This slight discrepancy however is not considered significant in the ability of water molecules to form tetrahedral arrangements of ice, although it would make the tetrahedral order necessarily slightly irregular.

1.1.2 The many differing phases of ice

While I_h is the form of ice found in nature, many different solid phases of H_2O are possible, some of which are stable at various temperature/pressure ranges and others are metastable. These other phases are typically grown under high pressure conditions. The various phases of ice are enumerated using Roman numerals, with ice I being I_h and subsequent ice phases being numbered in roughly the order in which they were first observed experimentally.[42–46] Nine of the first twelve phases (all except for IV, IX and XII) have been found to be stable at certain temperatures and pressures and metastable at greatly extended temperature/pressure ranges.[41]

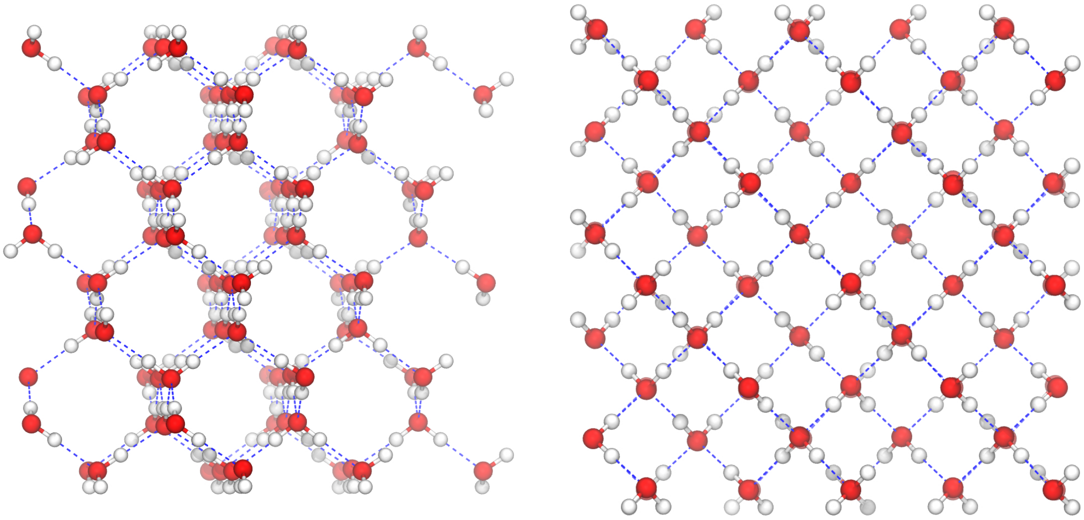


Figure 2: The crystal structure of I_c, shown via two different orientations. Oxygen atoms are coloured red, hydrogen atoms are coloured white and hydrogen bonds are displayed by dashed blue lines.

1.1.3 Cubic ice

Cubic ice (I_c) is a metastable variant of I_h, with the cubic structure seen in diamond rather than the hexagonal structure described above. As with I_h, each water molecule forms four hydrogen bonds to four distinct neighbouring molecules and the density is practically identical. Pure I_c is not seen in nature, however crystals of it can be formed by warming high-pressure phases (II–IX) to temperatures of 120–170 K, where the phase transition to I_c occurs.[47] Such I_c will transition to I_h at around 200 K and I_h will never transition back into I_c under further cooling.[41] This structure is composed of a system of interconnected double diamond cages (DDCs). Such a cubic crystal structure is illustrated in figure 2, with oxygen atoms coloured red, hydrogen atoms coloured white and hydrogen bonds displayed by dashed blue lines.

1.2 Classical nucleation theory

One cannot discuss the theoretical study of ice nucleation without mentioning classical nucleation theory (CNT). This is a simple, yet useful, theory which describes the thermodynamics and kinetics of nucleation quantitatively. In CNT, crystalline structures are treated as homogeneous spheres, surrounded by an infinitely thin interface; this is termed the capillarity approximation; which infers that the difference in free energy between liquid and crystal ($\Delta\mu_\mathcal{V}$) alongside the interfacial free energy ($\gamma_\mathcal{S}$) fully describe the thermodynamics of crystal nucleation.[48] In other words, the free energy of formation ($\Delta G_\mathcal{N}$) of a spherical crystal of radius r is:

$$\Delta G_\mathcal{N} = A(r)\gamma_\mathcal{S} - V(r)\Delta\mu_\mathcal{V} = 4\pi r^2\gamma_\mathcal{S} - \frac{4}{3}\pi r^3\Delta\mu_\mathcal{V}, \tag{1}$$

where $A(r) = 4\pi r^2$ is the surface area of the crystal and $V(r) = \frac{4}{3}\pi r^3$ is the volume of the crystal.

The critical nucleus size, i.e. the size at which the free energy cost of formation matches the free energy difference between liquid and crystal, is obtained by differentiating equation (1), with respect to r, and solving $\Delta G'_{\mathcal{N}} = 0$:

$$\Delta G'_{\mathcal{N}} = 8\pi r \gamma_{\mathcal{S}} - 4\pi r^2 \Delta\mu_{\mathcal{V}} = 4\pi r\left(2\gamma_{\mathcal{S}} - r\Delta\mu_{\mathcal{V}}\right) = 0. \tag{2}$$

Thus $2\gamma_{\mathcal{S}} - r\Delta\mu_{\mathcal{V}} = 0$, and so $r^* = \frac{2\gamma_{\mathcal{S}}}{\Delta\mu_{\mathcal{V}}}$, where r^* is the critical radius.

We define $\rho_{\mathcal{C}}$ to be the number density[*] within the crystal. The expressions[48] for the critical nucleus size (n^*) and free energy barrier for nucleation ($\Delta G^*_{\mathcal{N}}$) are as follows:

$$n^* = \rho_{\mathcal{C}} V(r^*) = \rho_{\mathcal{C}} \cdot \frac{4}{3}\pi r^{*3} = \rho_{\mathcal{C}} \cdot \frac{4}{3}\pi \cdot \left(\frac{2\gamma_{\mathcal{S}}}{\Delta\mu_{\mathcal{V}}}\right)^3 = \frac{32\pi\rho_{\mathcal{C}}}{3} \cdot \frac{{\gamma_{\mathcal{S}}}^3}{{\Delta\mu_{\mathcal{V}}}^3}, \tag{3}$$

$$\Delta G^*_{\mathcal{N}} = 4\pi r^{*2}\gamma_{\mathcal{S}} - \frac{4}{3}\pi r^{*3}\Delta\mu_{\mathcal{V}} = 4\pi\left(\frac{2\gamma_{\mathcal{S}}}{\Delta\mu_{\mathcal{V}}}\right)^2 \gamma_{\mathcal{S}} - \frac{4}{3}\pi\left(\frac{2\gamma_{\mathcal{S}}}{\Delta\mu_{\mathcal{V}}}\right)^3 \Delta\mu_{\mathcal{V}} = \frac{16\pi}{3} \cdot \frac{{\gamma_{\mathcal{S}}}^3}{{\Delta\mu_{\mathcal{V}}}^2}. \tag{4}$$

The central result of CNT is the following equation for the nucleation rate[†] of homogeneous nucleation ($\mathcal{J}$):[48]

$$\mathcal{J} = \rho_{\mathcal{S}} \mathcal{Z} \mathcal{A}_{\text{kin}} \exp\left(-\frac{\Delta G^*_{\mathcal{N}}}{k_B T}\right), \tag{5}$$

where $\rho_{\mathcal{S}}$ is the density of potential nucleation sites, $\mathcal{A}_{\text{kin}}$ is a kinetic pre-factor, k_B is the Boltzmann constant, T is the temperature and $\mathcal{Z}$ is the Zeldovich factor, which is a probability correction factor to account for the fact that a nucleus at the top of the free energy barrier is not guaranteed to form a crystal.[‡][49–51] From an initial impression of equation (5) one might expect[§] that $\mathcal{J}$ increases with temperature; however $\Delta G^*_{\mathcal{N}}$ can generally be assumed to be proportional to $T^3/\Delta T^2$, where ΔT is the supercooling.[48,51,52]

1.2.1 Heterogeneous CNT

CNT can also be used to describe surface heterogeneous nucleation. The nucleation barrier for heterogeneous nucleation is generally much lower than that of homogeneous nucleation; and as such in a heterogeneous system nucleation typically occurs preferentially at the interface(s). Thus, while we were interested equally in the entire surface of the (hypothetically spherical) nucleus for homogeneous nucleation, for heterogeneous nucleation the proximity to the foreign surface must be considered. The expression for heterogeneous

---

[*] I.e. $\rho_{\mathcal{C}} = n(r)/V(r)$, where $n(r)$ is the number of molecules in the spherical crystal with radius r.

[†] The nucleation rate is defined as the number of nucleation events per unit time and volume.

[‡] I.e. $\mathcal{Z}$ is the probability that a nucleus at the top of the free energy barrier will form a crystal, as opposed to re-dissolving.

[§] Perhaps with some alarm, since the nucleation rate should of course decrease with temperature.

free energy of formation $(\Delta G_{\mathcal{N}(\text{hetero})})^*$ is:[48]

$$\Delta G_{\mathcal{N}(\text{hetero})} = \Delta G_{\mathcal{N}(\text{homo})} \cdot f(\theta), \tag{6}$$

where $f(\theta) \leq 1$ is the "shape factor" which balances the three sets of interfacial free energies: $\gamma_{\mathcal{S}(\text{crystal,liquid})}$, $\gamma_{\mathcal{S}(\text{crystal,surface})}$ and $\gamma_{\mathcal{S}(\text{liquid,surface})}$.

1.2.2 Two-step nucleation

One key limitation behind CNT is the assumption that critical or pre-critical nuclei appear spontaneously and fully independent from one another — with the singular requirement being a stochastic overcoming of a single free energy barrier. It stands to reason however, that nucleation may follow a two-step process, wherein a first free energy barrier must be overcome in forming an unstructured cluster, with a second free energy barrier required to overcome to then structure the cluster into a crystalline order. This is not to say that such nucleation theories are truly "one-step" or "two-step" — it is a continuous free energy surface which can be traversed in multiple steps, but the key assumption is having one or two free energy barriers to overcome. Pan *et al.*[53] detailed such a process for the formation of ordered solid phases of proteins. In that work, they described the kinetics for such a two-step nucleation process. Kovács and Christenson[54] described a two-step nucleation process for ice nucleation from vapour, as is of interest in the study of atmospheric ice nucleation and cloud formation.

1.3 Homogeneous ice nucleation

The nucleation of supercooled water into ice occurs homogeneously in nature in certain settings, including that of atmospheric ice formation in clouds. Atkinson *et al.*[55] studied the freezing of pure water droplets with diameters between 4 and 24 µm. They found that volume nucleation[†] was more probable than surface nucleation[‡] for droplets with diameter greater than 6 µm. Sanz *et al.*[56] combined MD simulations with CNT to estimate the interfacial free energy and nucleation free-energy barrier of homogeneous ice nucleation. The conclusion of that work was that homogeneous ice nucleation is virtually impossible at supercooling lesser than 20 K. This appears to be in agreement with Atkinson *et al.*[55] who reported no droplets frozen above 237 K[§], a supercooling of 36 K. Thus much of the ice nucleation in nature would not occur under homogeneous conditions, and relies on heterogeneous factors.

---

[*] With respect to equation (1), relabelling $\Delta G_{\mathcal{N}}$ as $\Delta G_{\mathcal{N}(\text{homo})}$,

[†] Volume nucleation is nucleation within the bulk of a droplet.[55]

[‡] Surface nucleation is nucleation at the surface of a droplet.[55]

[§] According to figure 2 from that paper.

1.4 Heterogeneous ice nucleation

If such a high degree of supercooling is required for homogeneous ice nucleation, this raises the question: what induces ice formation at the much higher temperatures seen in nature? Ice nucleating agents (INAs) take the form of surfaces, molecules or solutes and increase the nucleation rates at higher temperatures by lowering the free energy barrier.[48] Kanji *et al.*[57] list a number of different aerosol ice nucleating particles (INPs) that can promote atmospheric ice nucleation heterogeneously. Reinhardt and Doye[58] used hybrid Monte Carlo simulations[59] to study ice nucleation on a number of different surfaces. Such hybrid simulations work via replacing the regular Monte Carlo moves with short molecular dynamics simulations.[58,59] They found that a Lennard-Jones flat wall did not facilitate nucleation, with ice nuclei more likely to form in the bulk than at the interface. Conversely, they found that rigid ice-like surfaces* were consistently able to promote ice nucleation. In a similar study, Fitzner *et al.*[60] simulated model systems of four generic fcc crystallographic planes: (111), (100), (110) and (211) with water layers of approximately 35 Å. In agreement with the previous study,[58] they found that the surfaces generally promoted ice nucleation.

In addition to computational studies of heterogeneous ice nucleation, often employing MD simulations to probe the nucleation process at an atomistic scale, there are a number of experimental methods which can be used to study nucleation rates. Vali[61] presented a method for quantitative analysis of drop freezing experiments which have been built upon more recently by Whale *et al.*[62]. These methods involve tracking the freezing of microlitre droplets as temperature is ramped downwards towards supercooled conditions and can be used to experimentally compare the efficiency of different particles at nucleating ice. Naturally these cannot provide the atomistic insight that we obtain from molecular simulations, however they are effective at determining the efficiency of different INAs to nucleate ice. This method is discussed in more detail in section 8.2.

Sosso *et al.*[21,24] combined molecular simulations with drop freezing assays[61,62] to study the ice nucleating potential of cholesterol (CHL) structures. They found that CHL crystals[21] exhibited a high degree of ice nucleating ability — acting as even more potent INAs than the mineral feldspar, which is known to be a very effective INA. In a separate paper,[24] it was shown that CHL monolayers also exhibited, albeit to a lesser extent, ice nucleating potential, with such ice nucleating ability directly proportional to the degree of structural order within the monolayer. In these studies, the strong nucleating effect of such CHL structures was attributed to the hydroxyl groups of CHL forming five and six-membered, hydrogen bonded, rings with water molecules which build molecular cages that act as a template for ice nucleation.

---

*I.e. surfaces where the surface structure mimics the structure of ice, as described in section 1.1.1.

1.5 Cryopreservation

At low temperatures, chemical reactions and biological metabolism slow down to such an extent that allows for the long-term preservation of living cells and tissue. This basic principle of cryopreservation is very simple and such biological matter is in fact able to endure extremely low temperatures, below $-180\,^\circ\mathrm{C}$. However, while storage of cryopreserved matter is possible, the freezing and thawing processes pose far more of a challenge.

At sub-zero temperatures above $-15\,^\circ\mathrm{C}$, extracellular ice will begin to form in the aqueous medium, however intracellular ice formation is not observed; this was attributed by Gao and Critser [14] to the plasma membrane blocking the growth of ice crystals. [14] It is the temperature region of -15 to $-60\,^\circ\mathrm{C}$ where cryoinjury[*] occurs. As the extracellular medium is already partially frozen, the chemical potential of the intracellular supercooled water is greater than that of the extracellular water, and thus there is a net movement of water from inside to outside the cells via osmosis. [9,14] The mechanism of cryoinjury depends on the cooling rate. When cooling rapidly intracellular ice will form, leading to an increased cellular volume and resulting in cell death. [63–65] Conversely if cells are cooled too slowly, the increase in volume of the intracellular water due to freezing will be less than the volume loss due to osmosis; leading to a reduction in the cellular volume, which can also result in cryoinjury. [66,67] Thus both rapid and slow freezing techniques can cause cryoinjury and cell death.

Alongside choosing the cooling and thawing rates, the rate of cryoinjury can be reduced by the addition of cryoprotective agents (CPAs). A good CPA should have low toxicity and be capable of penetrating the cells which are to be preserved. [9,10] One example of a very commonly used CPA is glycerol, [68,69] which is widely used in the storage of bacteria and animal sperm. [70] Another CPA, as demonstrated in 1959 by Lovelock and Bishop [71], is dimethyl sulphoxide (DMSO), which is frequently used for the cryopreservation of cultured mammalian cells. [72] One specific class of compound which are used to prevent the growth of ice are antifreeze proteins (AFPs), which are found in nature in extreme cold-dwelling species such as arctic and antarctic fish, birds, amphibians, insects and plants. [73–75] Mitchell *et al.* [76] demonstrated the design and synthesis of an AFP which was non-toxic to red blood cells and greatly reduced the rate of cryoinjury by inhibiting the growth of ice during the thawing process.

In the medical field, there are already many current applications of cryopreservation. One such application is the cryopreservation of embryos for the sake of fertility preservation. Since the first successful pregnancy from a cryopreserved human embryo [77] in 1983, the use of cryopreservation in embryo storage for human-assisted reproduction has become widespread. [78] Riggs *et al.* [79] undertook an extensive analysis of the use of cryopreserved embryos with a total of 11,768 IVF and oocyte donation patients from between 1986 and 2007. They found that the length of time for which the embryos were stored had no adverse effect on post-thaw embryonic survival or pregnancy outcome — indicating that biological

---

[*] I.e. damage to biological matter, caused by low temperatures.

and chemical processes within cryogenically stored embryos are indeed sufficiently slowed for long term storage.

A similar second application, again for fertility preservation, is the cryopreservation of sperm cells and semen. Sperm freezing is used primarily to preserve the fertility of patients undergoing cancer therapy, but also for patients who are unable to ejaculate and for transgender patients who have undergone male-to-female reassignment; wherein estrogen therapy causes temporary or permanent infertility.[80] The first successful human pregnancy using frozen sperm was reported by Bunge and Sherman[81] in 1953, where semen had been stored, for a short period, in dry ice. By 1963, cryopreservation of sperm was becoming widespread, with liquid nitrogen used as the method for long-term storage.[82,83] Szell *et al.*[84] reported in 2013 a successful pregnancy using sperm which had been cryopreserved from 1971–2011, a period of 40 years; this gives a good indication of the ability for sperm cells to survive long-term cryopreservation, over a period that greatly exceeds the typical use of sperm freezing.

Another clinical application of cryopreservation is the preservation of stem cells; particularly mesenchymal stem cells (MSCs); which are found in various tissues, such as bone marrow, adipose tissue, amniotic fluid and umbilical cord blood. MSCs are valuable to the field of regenerative medicine, due to their ability to differentiate into specialised cells from all three lineages: ectoderm, mesoderm and endoderm.[11,85–87] MSCs are capable of differentiating into neurons and have been trialled, on rats and mice, with some promising results, to treat neurodegenerative disorders, such as Parkinson's disease[88–90] and Alzheimer's disease.[90–93] MSCs also have the potential to be used in treating immune disorders, such as type 1 diabetes,[94,95] due to their ability to regulate immune responses; as well as cardiovascular diseases.[96,97] Bahsoun *et al.*[98] investigated the impact that cryopreservation has on bone marrow-derived MSCs, they found that cryopreservation had no effect on the differentiation or proliferation of MSCs, however there were concerns about the ability for cryopreserved MSCs to form colonies, as well as their viability. However the cryopreservation of, and indeed the clinical applications of, MSCs are still in early stages, with novel nanoscale techniques proffering solutions to the issues surrounding cryopreservation.[99]

1.6 Molecular simulations of ice formation

Molecular dynamics (MD) simulations employ Newtonian mechanics to simulate the trajectories of atoms over time based on generalised models of the physics of interatomic interactions Such simulations are extremely valuable as they can provide detail at an atomic scale which is impossible to attain from experimental techniques. The first MD simulations were carried out in the 1950s by Alder and Wainwright[100], but have recently become evermore popular, thanks in no small part to the wide availability of simulation packages, as well as the increase in computing power, leading to a far greater range of possibilities for simulation.[101,102]

In any molecular simulation, there are a great number of decisions to be made. One especially important parameter when configuring a simulation is the choice of force fields. In particular, when we are interested in water dynamics, such as ice nucleation, the choice of water model is key. As simulations involving water have been common since the advent of molecular simulations, there are a huge number of different water models.[103,104] In early simulations it was common to model water as a simple point charge (SPC)[105–107] or even to omit it entirely, however with advances in computational resources, water models with more complex dynamical properties are viable.

One water model which has been frequently used for molecular simulations is the monatomic water model (mW) which manages to coarse grain water molecules into a single atom while retaining the hydrogen bonding structures that water forms via the inclusion of an angular dependent term.[108] This model is the best performing coarse grained water model and in some regards can actually outperform atomistic models.[109] The mW model has been used for studies involving ice nucleation[110,111] due to its speed in comparison with fully atomistic models, however it over predicts the density of ice.[112]

The TIP4P water model[113] is structured with a Lennard-Jones site, O, representing the oxygen atom, with three charge sites: H_1, H_2 and M. There have been many re-parametrisations of the original TIP4P model, with commonly used versions including TIP4P-Ew[114] and TIP4P/2005.[115] The TIP4P/Ice[116] re-parameterisation has been designed specifically to deal with solid-phase properties of water. This ice-focussed parameterisation has been shown to accurately reproduce the properties of ice and supercooled liquid water in multiple studies.[117,118]

The interplay of I_c and I_h during the nucleation process, as well as in stable ice phases is another topic of common discussion. In nature, only I_h is observed,[119] however during molecular simulations of ice nucleation, I_c is often initially more common.[21,120] Buxton *et al.*[121] used molecular dynamics simulations, both classical and employing the path-integral molecular dynamics methodology, to examine the relative stability of I_h and I_c ice phases. They concluded that the higher stability of I_h as compared to I_c could be generally attributed to nuclear quantum effects (NQEs). This provides an explanation for the ubiquity of I_h in nature.

Chapter 2

Lipid Bilayers

The previous chapter introduced the structure of ice, nucleation theory and simulations of ice nucleation. In a similar fashion, this chapter will serve as an introduction to the structure of various lipid bilayers and the dynamics of water at their interfaces. In conjunction, these two chapters discuss the foundations of the theory this work aims to extend. The main focus of this thesis is the ability, or inability, of lipid bilayers to act as ice nucleating agents. The structural and chemical properties of different examples will be examined as will the behaviour of supercooled water at the interface.

2.1 Phospholipid bilayers as cell membranes

The primary components of cell membranes are lipid bilayers, often composed primarily of phospholipids — lipids consisting of a hydrophilic phosphate-containing headgroup, with two hydrophobic fatty acid tails.[34] Most animal cell membranes are approximately 50% phospholipid, by mass. Due to the amphiphilic nature of phospholipids, lipid bilayers are formed spontaneously under suitable conditions, aggregating so as to submerge their hydrophobic tails in the interior of the bilayer, with their hydrophilic heads facing the water. Similarly, where phospholipid bilayers are ruptured, they rearrange autonomously to repair the bilayer structure due to the energetic preference for the hydrophobic tails to be away from water, and vice versa for the hydrophilic heads.[35]

Common phospholipids found within animal cells include phosphatidylcholine (PC), sphingomyelin (SM), phosphatidylethanolamine (PE), phosphatidylinositol (PI), phosphatidate (PA), phosphatidylserine (PS), phosphatidylglycerol (PG) and diphosphatidylglycerol (DPG). For example, human erythrocyte (red blood cell) membranes contain 55 mol% phospholipids[*] and 45 mol% cholesterol (CHL, see figure 3(c)).[36]

One specific phospholipid commonly found in cell membranes is 1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC, see figure 3(a)). As far back as 1961[122] PC-CHL bilayers have been studied, with some early techniques including deuterium nuclear magnetic resonance spectroscopy (^{2}H NMR)[123] and electron spin resonance (ESR)[124], both

---

[*] Specifically, 17 mol% PC, 17 mol% SM, 16 mol% PE and 6 mol% PS.

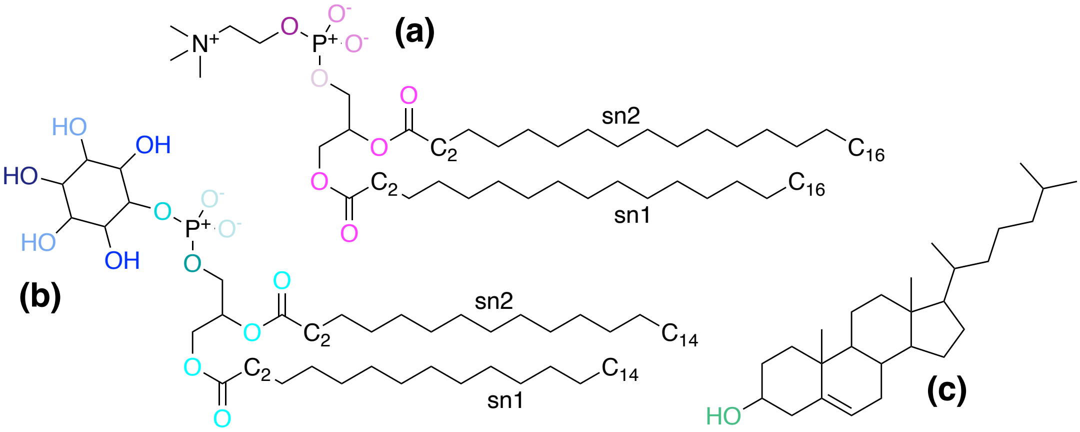


Figure 3: The chemical structures of **(a)** DPPC, **(b)** DMPI and **(c)** CHL. The hydrogen bonding sites are coloured according to the palette reported in figure 24.

of which can provide insight into the mobility of the acyl chains which form lipid tail-groups. DPPC-CHL bilayers have been extensively studied via both simulation[28,30,33,125] and experiment.[126–129]

Another class of phospholipid found, albeit less prominently, in cell membranes are PIs these are known to have an active role within mammalian cell membranes: anchoring proteins and spatially organising trans-membrane signalling pathways.[130,131] Stanishneva-Konovalova and Sokolova[32] studied DPPC-DPPI[*] bilayers, containing 0.8 and 9.4 mol% DPPI, via united atom and coarse-grained molecular simulations. They found that increasing the amount of DPPI lead to increased packing density of the DPPC lipids. In this thesis, specifically the lipid 1,2-dimyristoyl-*sn*-glycero-3-phosphatidylinositol (DMPI, see figure 3(b)) is studied, with regards to the effect it has on DPPC bilayers and their ability to nucleate ice.

2.2 The effect of cholesterol within phospholipid bilayers

As a common constituent lipid of cell membranes, it is important to consider the structural effect the inclusion of CHL has on bilayers. From an experimental standpoint, Hung *et al.*[132] looked at the structure of various PC-CHL bilayers. They used lamellar X-ray diffraction to measure phosphate-phosphate distances across the bilayers. In all cases they found that the thickness of the bilayer increases with CHL concentration while the area per lipid decreases. This thickening of the bilayer is due to the closer packing of lipids, meaning that the tails must be more perpendicular to the central plane of the bilayer. They noted that the amount by which the bilayers thickened was dependent on the fatty acid chains of the PC lipids and also that the maximum thickness was achieved in all cases at around 38 mol% CHL, a concentration considerably below the maximum theoretical solubility of CHL.

---

[*] 1,2-dipalmitoyl-*sn*-glycero-3-phosphatidylinositol.

Leeb and Maibaum[133] studied the effect of CHL concentration on DPPC-CHL bilayers via MD simulations. They found that bilayer thickness increases with CHL concentration up to around 20 mol%, after which a small decline was observed. They also found that the degree of structural order increased with CHL concentration, again up to around 20 mol% where the order plateaued. Again with the most pronounced effect between 0–20 mol% CHL, the surface area per lipid was noted to decrease as CHL concentration increased. This trend was greater than the decrease expected from the smaller size of CHL as compared to DPPC; in fact, adding a single CHL molecule into a (49 lipids per leaflet) pure DPPC bilayer would decrease the total surface area. In simulations with a single CHL molecule per leaflet, they observed similar trends, particularly observing a greater degree of tail order within DPPC tails, the lower the distance from the CHL molecule.

Phospholipid bilayers are in fact two dimensional fluids, with the individual lipids able to diffuse within the plane of the leaflet. The fluidity of such a bilayer depends heavily on its component lipids. CHL in particular has an immobilising effect on neighbouring lipids, due to its its hydroxyl group positioning itself close to the base of the polar phospholipid headgroups, such as PCs, with its rigid plate-like steroid rings stiffening the ends of the phospholipid acyl chains closest to the headgroups.[35] This causes the bilayers to become stiffer and reduces the permeability to both water molecules and small water-soluble molecules such as many drug molecules.[35]

2.3 Lipopolysaccharide membranes

The cell envelope of gram-negative bacteria, such as Escherichia coli (E. coli), consists of a peptidoglycan cell wall, surrounded by an outer membrane (OM). This OM is a lipid bilayer but, unlike the membranes discussed earlier, not a phospholipid bilayer. Instead, the OM is an asymmetric bilayer where the inner leaflet is composed of phospholipids while the outer leaflet is comprised of glycolipids, in particular lipopolysaccharide (LPS).[134] Synthesised LPS membranes consist of three regions: the O-polysaccharide chain, made up of repeating oligosaccharide units; the oligosaccharide core; and Lipid A (see figure 4). “Rough” mutants of LPS (R-LPS) do not have the O-specific chain, with the different classes denoted a–e depending on the degree of completion of the oligosaccharide core.[135] In particular, RaLPS contains the full oligosaccharide core, while ReLPS has only 2 or 3 KDO[*] units linked to each Lipid A molecule (see figure 6).[136,137] See figure 5 for a schematic of such LPS membranes with detail of the oligosaccharide core which was simulated in this work (section 3.2) which consists of two KDO[*] units, four Hep[†] units, three Glc[‡] units and one Gal[§] unit (see figure 6 for structures).

This asymmetric LPS-phospholipid OM is responsible for a number of essential barrier

---

[*] 2-keto-3-deoxy-D-mannooctanoic acid

[†] L-glycero-D-manno-heptose

[‡] D-glucose

[§] D-galactose

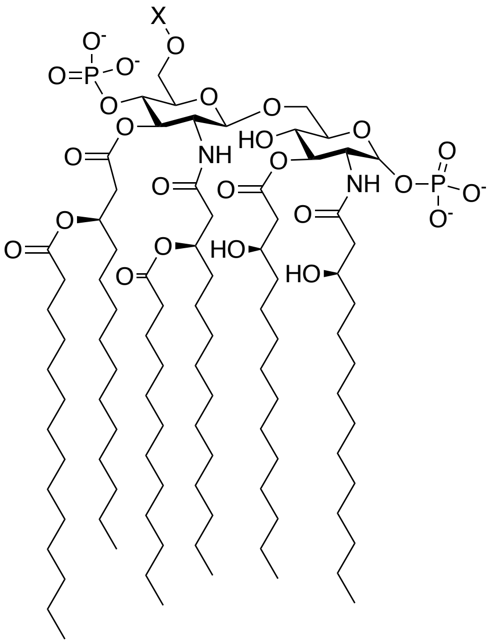


Figure 4: The chemical structure of Lipid A.

functions for the survival of such bacteria. Indeed, the OM protects bacteria against toxic substances, such as antibiotics, while permitting the transport of important nutrients.[138] The effectiveness of the barrier function of LPS can be explained by its amphipathic properties; similarly to phospholipid bilayers, the acyl chains in the lipid A molecules have a hydrophobic nature which prevents the passage of hydrophilic molecules though the leaflet; unlike phospholipids such as PCs however, the oligosaccharide regions display strong hydrophilic properties, thus preventing the passage of hydrophobic molecules alike.[139] The existence of the OM is largely attributed as a key factor in the resistance of gram-negative bacteria to antimicrobials.[139]

2.4 Water at the interface with lipid bilayers

The structure of lipid bilayers would be very different without the presence of water. The "hydrophobic effect", wherein lipid molecules group together — thus preventing contact between water and their hydrophobic tails — is what leads to the creation of "self-assembled lipid bilayers".[140] Åman *et al.*[141] used MD simulations to study the behaviour and interactions of interfacial water on the surface of a DPPC bilayer. They looked at the first and second hydration shells of the lipid headgroups, defined as the regions enclosing 4.5 and 6.9 Å from the phosphorous atom, respectively. A double layer of interacting water was evident, with the first and second shells containing approximately 5 and 16 water molecules, respectively. This corresponds to circa 5 water molecules tightly bound to the carbonyl group, with the remaining circa 11 water molecules around the choline groups. The coordination of the various chemical groups within the headgroups will have a great

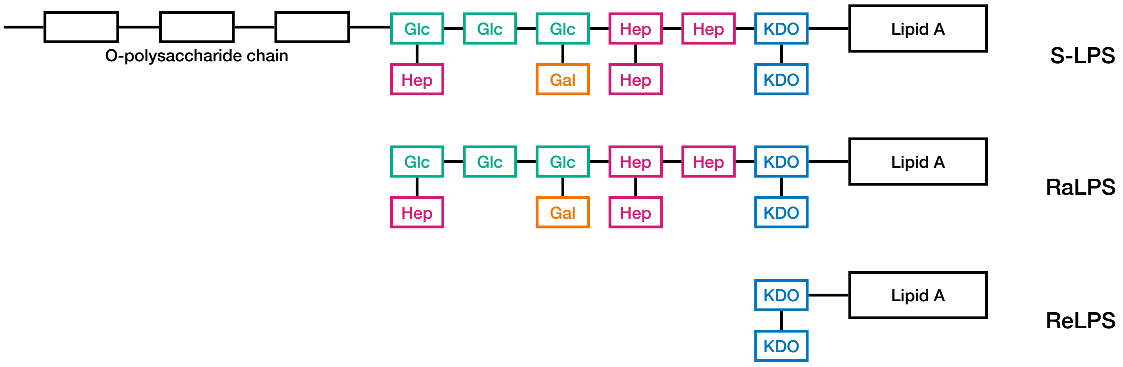


Figure 5: Schematic for three LPS chemotypes.[135] At the top is the complete S-form, with the repeating oligosaccharide units (not simulated in this work). In the middle is the RaLPS mutant, with the complete oligosaccharide core. On the bottom is the ReLPS mutant with two KDO units. The exact constituents of the oligosaccharide core vary with different LPS membranes, this specific variant is the one which has been simulated in the case of the RaLPS and ReLPS systems.

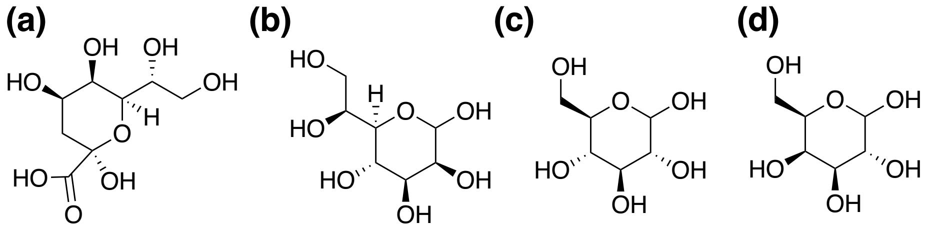


Figure 6: The chemical structures of **(a)** 2-keto-3-deoxy-D-mannooctanoic acid (KDO), **(b)** L-glycero-D-manno-heptose (Hep), **(c)** D-glucose (Glc) and **(d)** D-galactose (Gal).

effect on the analysis later discussed in chapter 7.

The dynamics of water at the interface with lipid bilayers have been extensively examined via molecular simulations. Wilson and Pohorille[17] used molecular dynamics to study the interface between water and a glycerol-1-monooleate (GMO) bilayer. Like DPPC, GMO headgroups are uncharged, they found that water penetrated into the headgroup region but not beyond, they also found that the hydrophilic surface of the GMO bilayer disrupted the interface. Additionally, they suggest that the TIP4P water model significantly underestimates the surface tension of water, an important consideration as if the surface tension is too low this may cause water to spread further from the bilayer, leading to lower overall density. In addition, when simulations have a non-fixed box size, this reduction in water density will lead the in-plane dimensions of the simulation box to shrink — leading to a more compact bilayer or indeed to a gel-phase bilayer at temperatures where it should be liquid crystalline. This is discussed to a greater extent in section 3.4.1, to ensure that these issues do not affect the simulations herein.

Ions within the water can also interact with the bilayer. Böckmann *et al.*[142] described the effect of sodium chloride on pure 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphocholine (POPC) bilayers using a combined experimental/computational approach — coupling fluorescence correlation spectroscopy with MD simulations. They noted a strong interaction between sodium ions and the carbonyl oxygens within the fatty acid tails. This leads to an increased degree of ordering of the lipid tails, resulting in denser and thicker bilayers, with lower in-plane mobility.

Chapter 3

Simulation Details

3.1 DPPC-CHL/DMPI systems

DPPC (see figure 3(a)) bilayers are prototypical models of cellular membranes, having been the subject of many studies;[28,30,143] both computational and experimental in nature, often also containing CHL. Thus, DPPC is an excellent candidate to further our understanding of ice formation in biological matter. DMPI (see figure 3(b)) is another phospholipid with a different headgroup to DPPC, notably with an negative overall charge.[*]

Eleven DPPC-CHL/DMPI lipid bilayer systems were constructed, using CHARMM-GUI,[147–151] with 30 lipids per leaflet (60 per system). The number of each lipid per system, together with the in-plane dimensions of the simulation box, after equilibration and quenching (see section 3.3), are listed in table 1. The systems have square cross section so the x and y dimensions are the same. A TIP4P water layer, approximately 30 Å thick, was placed either side of the bilayers, using the molecular dynamics (MD) package GROMACS[152–158] (see figure 7).

Table 1: Lipid composition of DPPC-CHL/DMPI systems simulated.

System name	**No. lipids per leaflet** **DPPC**	 **CHL**	 **DMPI**	**Box x** [nm]
Pure DPPC	30	-	-	3.56
20 mol% *CHL*	24	6	-	3.56
40 mol% *CHL*	18	12	-	3.37
60 mol% *CHL*	12	18	-	3.37
80 mol% *CHL*	6	24	-	3.35
Pure CHL	-	30	-	3.31
20 mol% *DMPI*	24	-	6	3.91
40 mol% *DMPI*	18	-	12	4.04
60 mol% *DMPI*	12	-	18	4.05
80 mol% *DMPI*	6	-	24	4.20
Pure DMPI	-	-	30	4.12

---

[*] It has been shown in some studies that charges can have a large effect on ice nucleation and can even make the difference between a surface or crystal promoting or inhibiting ice nucleation.[144–146]

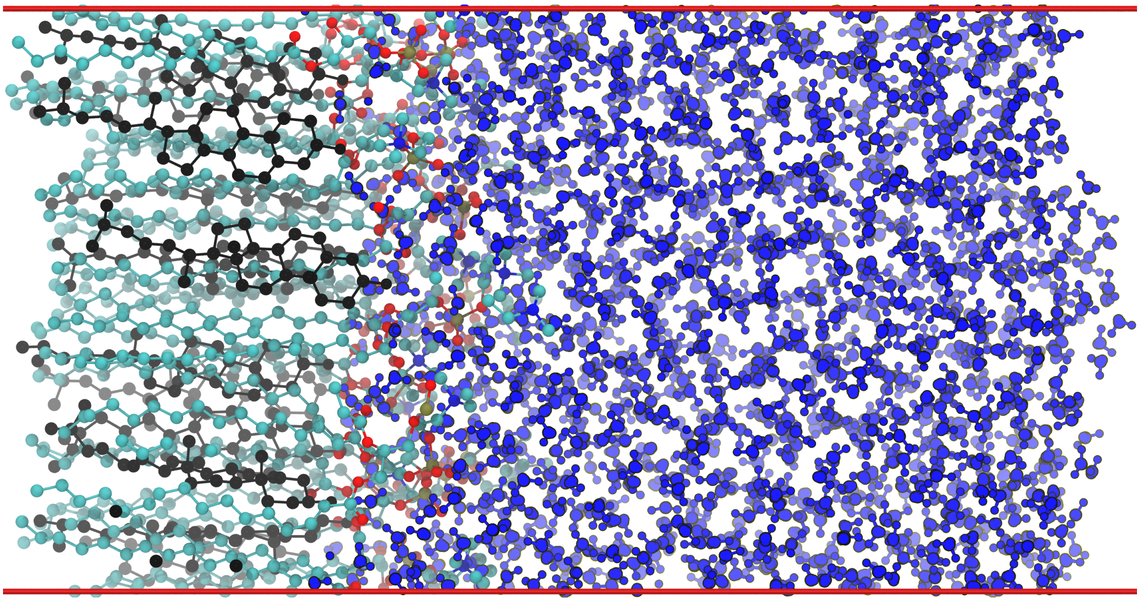


Figure 7: Representative snapshot of (half of) a DPPC-CHL/water system, 40 mol% CHL (coloured in black). The full system is a bilayer with another leaflet and water layer to the left of the figure. Water molecules are coloured in blue. Within the DPPC molecules, carbon, oxygen, phosphorus and nitrogen atoms are coloured in cyan, red, yellow and blue, respectively. Hydrogen atoms belonging to DPPC or CHL molecules are not shown. The water layer is in contact with an extended region of vacuum (see section 3.4). Note that the horizontal axis in this figure is the one labelled z, with the xy-plane containing the vertical axis and the page normal. This convention is shared across every system discussed in this thesis.

3.2 Lipopolysaccharide systems

Three asymmetric phospholipid-LPS systems were simulated. These are representative of the outer membrane of Gram-negative bacteria.[134,159] The three systems consisted of a phospholipid leaflet and a lipid A (see figure 4) leaflet, with varying amounts of sugars: 2-keto-3-deoxy-D-mannooctanoic acid (KDO) L-glycero-D-manno-heptose (Hep), D-glucose (Glc) and D-galactose (Gal); coating the lipid A (see figure 6). Specifically, the three systems studied were Lipid A with no sugars; the "deep rough mutant" ReLPS, with two KDO units linked to each Lipid A molecule; and RaLPS, with the complete oligosaccharide core. The corresponding phospholipid leaflets consist of the three lipids 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoethanolamine (POPE), 1-palmitoyl-2-oleoyl-*sn*-glycero-3-phosphoglycerol (POPG) 1,10-palmitoyl-2,20-vacenoyl cardiolipin (PVCL2) for the Lipid A and ReLPS systems (see figure 8) and the three similar lipids 1-palmitoyl-2-vacenoyl-*sn*-glycero-3-phosphoethanolamine (PVPE), 1-palmitoyl-2-vacenoyl-*sn*-glycero-3-phosphoglycerol (PVPG) and PVCL2 for the RaLPS system. PVPE and PVPG differ from POPE and POPG only in the fact that the C=C double bond is between C_9 and C_{10} on the secondary tail for POPE/G and between C_{11} and C_{12} on the secondary tail for PVPE/G. The exact composition and in-plane dimensions of these bilayers can be found

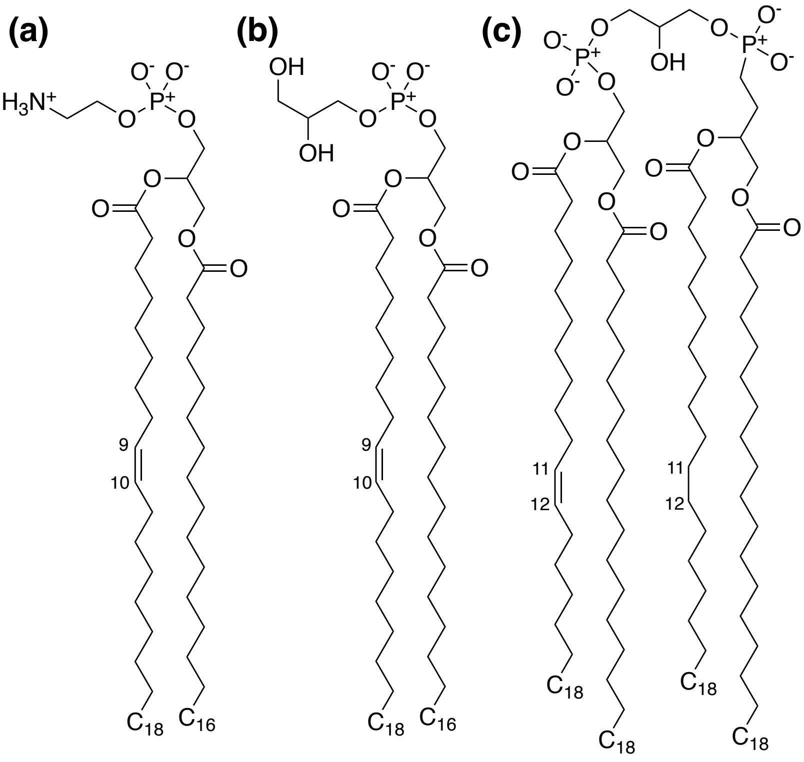


Figure 8: The chemical structures of **(a)** POPE, **(b)** POPG and **(c)** PVCL2. PVPE and PVPG differ from POPE and POPG only in the fact that the C=C double bond is between C_9 and C_{10} on the secondary tail for POPE/POPG and between C_{11} and C_{12} on the secondary tail for PVPE/PVPG.

in table 2. Again, box x and y dimensions are paired.

3.3 Equilibration and MD setup

Simulations were undertaken using the MD package GROMACS.[152–158] The force-field used to model the lipids was CHARMM36,[31,160–164] which was used in particular to take advantage of its CHL parameterisation,[165] and the TIP4P/Ice[116] force field was used for the water molecules. It is worth mentioning that CHARMM36 was originally parameterised to be used with the TIP3P water model, however the combination of CHARMM36 and TIP4P/Ice has been validated and has been used to accurately reproduce the properties of supercooled liquid water and ice in recent studies.[21,117,118] An initial energy minimisation was carried out; using the SETTLE[166] algorithm to constrain the geometry of the water molecules, and the LINCS[167] algorithm to constrain the bilayer geometry. Three dimensional periodic boundary conditions (PBCs) were used. Subsequently, a number of initial equilibration runs (20 ns) were carried out at 323.15 K, sampling the NVT^* ensemble. The fairly high temperature of 323.15 K was chosen to avoid the gel-phase

---

[*] POPE/POPG for Lipid A and ReLPS, PVPE/PVPG for RaLPS.

[*] Fixed number of atoms, volume (i.e. x, y and z box dimensions) and temperature.

Table 2: Composition of LPS systems simulated (No. molecules).

System	**Lipid A**	**ReLPS**	**RaLPS**
POPE†	145	145	90
POPG†	8	8	5
PVCL2	8	8	5
Lipid A	53	53	35
KDO	0	106	70
Hep	0	0	140
Glc	0	0	105
Gal	0	0	35
Box x [nm]	9.14	9.48	7.94

transition for DPPC, which occurs at around 305–315 K.[29]

Following these MD runs, the simulation box was elongated in the z-direction up to 270 Å; creating water-vacuum interfaces, in order to avoid artefacts caused by the slab geometry of the system[168] (see section 3.4). An equilibration run of 20 ns, sampling the NVT ensemble at 323.15 K, was then carried out. Following the NVT run, a longer NpT[‡] run, of 200 ns, was carried out in order to equilibrate the system. Under the NpT ensemble, the x and y (in-plane) box dimensions were coupled and the z box dimension was fixed. Lipid bilayers on their own are capable of expanding and contracting by a reasonably large factor under pressure. A constant surface tension $\gamma_s = 120\,\mathrm{mJ\,m^{-2}}$, was imposed. The choice of this γ_s value is discussed in section 3.4.1.

Following liquid crystalline equilibration, the systems were quenched from 323.15 K to 233.15 K, at a rate of $2.25\,\mathrm{K\,ns^{-1}}$, under constant (ambient) pressure and imposed surface tension ($120\,\mathrm{mJ\,m^{-2}}$). The systems were subsequently equilibrated at 233.15 K, using a 20 ns run under the NPT ensemble with surface tension of $120\,\mathrm{mJ\,m^{-2}}$. The time step used for these simulations was 2 fs, meaning that the 20 ns equilibration runs contained 20 million steps. The ergodic hypothesis is assumed, i.e. that for a sufficiently long simulation, the probability of finding the system in any reachable state is independent of the time, and therefore also independent of the starting configuration.[169]

Once equilibrated at 233.15 K, extended MD runs of 70 ns–3.8 µs, depending on system, under the NVT ensemble were started. The NVT ensemble was preferable for the production runs as the fixed box dimensions make quantitative analyses, particularly those in terms of z, simpler to perform. The DPPC-CHL systems were all run for over 3 µs, however this was deemed unnecessary for the subsequent systems on the basis that the systems are in a good state of equilibration and thus a smaller number of frames adequately samples the configurations. Of course, as these are unbiased simulations of rare events, the phase space is certainly not fully explored, however extended unbiased simulations will not be able to achieve that. These simulations were carried out in GROMACS using the `v-rescale` thermostat, Berendsen barostat with surface tension coupling (where relevant),

---

[‡] Fixed number of atoms, pressure and temperature.

Verlet cutoff scheme, plain cutoff for Van der Waals interactions, with `force-switch`.

3.4 Verification of simulation procedure

Such two-dimensional slab systems pose a particular difficulty in MD simulations as there are trade-offs between the conventional methods for constructing them. In MD simulations, three-dimensional PBCs are preferable, as they do not require the imposition of any "walls" which stop atoms from leaving the system. The issue with using walls is that cause to edge effects, which can be on the scale of nanometers deep into the simulation box. Thus in a water-bilayer system, the water layer must be sufficiently thick enough so that a bulk-like water region can exist beyond the effects of both the wall and the bilayer. A similar issue occurs in this situation if three-dimensional PBCs are employed: as we want to consider an isolated bilayer within bulk water, the water layer must be thick enough that the bilayer cannot interact with itself, and again we desire a bulk-like region within the water layer.

As we are limited by computing resources, neither of the aforementioned setups are ideal; extending the water layer naturally has the effect of increasing the computational costs. Thankfully, there is another option, as described by Bostick and Berkowitz [168]. As detailed in the previous section, we have an extended vacuum region on either side of the water, A combination of an Ewald sum and imposed surface tension constrains the water with only minimal interfacial effects at the water-vacuum interface. This is called the Ewald sum in three dimensions with a correction term (EW3DC). In the following subsections, this simulation setup will be scrutinised to determine whether it is valid.

3.4.1 Surface tension

The electrostatic potential which is imposed when using the EW3DC method is not sufficient on its own to stop water from spreading out into the vacuum region. Just as the behaviour of water in the real world is greatly affected by its high surface tension, imposing surface tension is necessary here to make the system act as it should when using the NpT ensemble. According to literature[170] results: the ice-water surface tension of TIP4P/Ice water is $29.8\,\mathrm{mJ\,m^{-2}}$. Assuming that the water-bilayer and water-vacuum surface tensions are similar, one can therefore expect to get the correct surface tension value by multiplying this by the number of interfaces (4). Thus a value of $\gamma_s \approx 120\,\mathrm{mJ\,m^{-2}}$ is expected.

Using the pure DPPC system, following the 323.15 K NVT run, a number of 20 ns NpT runs were commenced also at 323.15 K with varying values for this imposed surface tension. The resultant water densities are illustrated in figure 9. A linear regression was carried out, omitting the point at $\gamma_s = 150\,\mathrm{mJ\,m^{-2}}$ as this point is a clear outlier, the regression model has $R^2 = 0.9088$: strong correlation. This linear regression gives the water density with $\gamma_s = 120\,\mathrm{mJ\,m^{-2}}$ as $0.989\,\mathrm{g\,cm^{-3}}$, the correct density of water at 323.15 K. Therefore $\gamma_s = 120\,\mathrm{mJ\,m^{-2}}$ was chosen, in agreement with the literature value

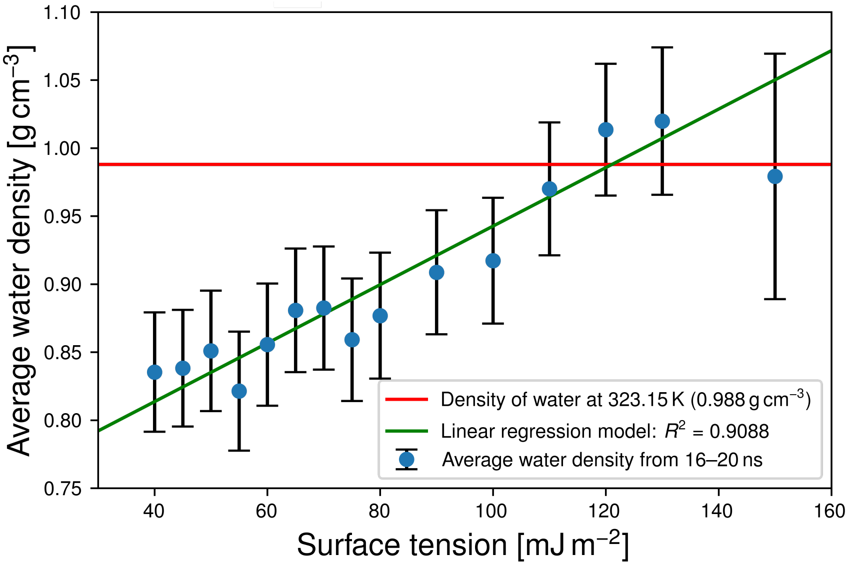


Figure 9: Bulk water density for different imposed surface tensions. The error bars show one standard deviation. The red line is the literature value for water density at 323.15 K. The green line shows the linear regression model, excluding the outlier at $\gamma_s = 150\,\text{mJ m}^{-2}$.

for surface tension. To further verify the structure when simulating with imposed surfaces tension, the surface area per lipid can be compared to literature data. For example, for pure DPPC after equilibration at 323.15 K, the box x and y dimensions were 4.24 nm, giving an average area per lipid of 0.599 nm^2 which is in good agreement with the data given by,[133] as are the 20 and 40 mol% CHL systems (see table 3).

Table 3: Average (mean) area per lipid after NVT and NPT equilibration at 323.15 K (liquid crystalline phase) compared to reference data from Leeb and Maibaum[133] (data extracted from figure 4 of that paper).

System	Box x [nm]	Mean area per lipid [nm^2]	Reference [nm^2]
Pure DPPC	4.24	0.599	0.60
20 mol% *CHL*	3.59	0.429	0.43
40 mol% *CHL*	3.52	0.413	0.41

3.4.2 Energy continuity

The MD package GROMACS[152–158] has an in-built function, `gmx energy`, which extracts energy values from MD trajectories. Using this function, the potential energy, kinetic energy, total energy, temperature, pressure and box dimensions were extracted from the equilibrium trajectories and it was confirmed that the values over time matched with what would be expected. Any jumps in energy between separate trajectory files would indicate a continuation error. Additionally, looking at these energies gives an impression of whether the simulations were long enough for the system to have reached an equilibrium, i.e. whether the energies appear to have converged. In particular, this was used to decide the length of the equilibrium trajectories: after an initial 20 ns NpT run these energies, and in particular the box dimensions, did not appear to converge, thus the longer equilibration run length of 200 ns was chosen.

3.4.3 Deuterium order parameter

The structure of lipid membranes can be studied experimentally using ^{2}H NMR to quantify the orientational mobility of C−H bonds* along the acyl chains of the lipids.[171,172] These ^{2}H NMR results, which can be very accurate, can be compared to molecular simulations via the deuterium order parameter (S_{CD}), which is defined as follows:[173,174]

$$S_{\mathrm{CD}} = \left\langle \frac{3\cos^2(\theta) - 1}{2} \right\rangle , \tag{7}$$

where θ is the (time and position dependent) angle between the C−D bond and a reference axis (in this case, the positive z-axis) and the brackets $\langle\cdot\rangle$ denote an ensemble average. Defined as such, the value of S_{CD} depends on not only the order within the structure but also the global orientation.

Figure 10 shows the values of S_{CD} along the acyl chains of DPPC in the pure DPPC simulations, in the liquid crystalline phase at 323.15 K, compared to reference data,[31] including both experimental S_{CD} values for a DPPC bilayer as well as those computed from a $NpAT^\dagger$ simulation under the CHARMM36 force field, as used herein. Note that the overall trends are the same, although all of the values are slightly higher than the reference data. This can be explained by the fact that the surface area per lipid in these pre-quenched simulations is slightly higher than in the reference data.[31] This should not be any cause for concern as the surface area per lipid will reduce during the annealing process, as the bilayer transitions from the liquid crystalline to the gel phase. Note that the imposed surface tension will not prevent this from occurring as it does not have a direct effect on the contraction (or expansion) of the box dimensions — the imposed surface tension value would be the same regardless of the in-plane dimensions of the system. Nonetheless, the matching shape of this S_{CD} distribution provides further evidence for the validity of the simulation setup.

---

[*] Naturally, for ^{2}H NMR, these are carbon-deuterium bonds.

[†] Fixed number of atoms, pressure, cross-sectional area and temperature.

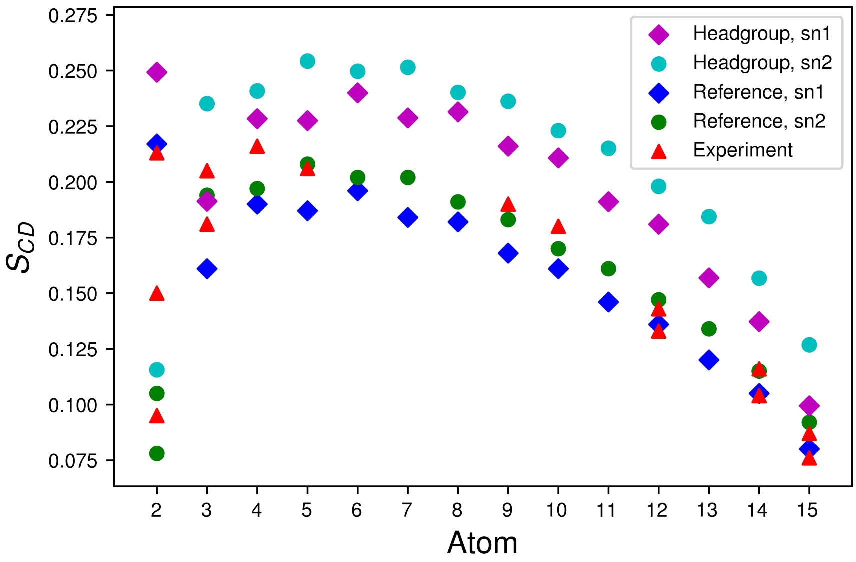


Figure 10: Deuterium order parameter (S_{CD}) values for acyl chains in pure DPPC bilayer simulations at 323.15 K, plotted as magenta diamonds and cyan circles; compared to reference data, all taken from the paper by Klauda *et al.*[31], plotted as blue diamonds and green circles for reference simulation data and red triangles for reference experimental data.

Chapter 4

Order Parameters for Water

The first challenge in analysing the occurrence of ice nucleation from a simulation trajectory is distinguishing between liquid water and ice. In the case of nucleation studies, using unbiased molecular simulations, this task is made all the more difficult by the fact that we are looking for small, pre-critical ice nuclei; which persist for a few nanoseconds, at most. One way of distinguishing between different phases of water is with geometric order parameters. Such parameters are designed to encode some geometric property of each molecule into a numerical value. Depending on the parameter in use, as well as the analytical intent, we can either use these values to filter down to only molecules of particular interest or we can study the distribution of values, perhaps in relation to time or some spacial axis. In this chapter, a number of useful order parameters are defined, as well as a method for condensing multiple parameters into one dimension.

4.1 Steinhardt bond order parameters

One such set of order parameters, which are commonly used for the purpose of studying ice nucleation, are the (Steinhardt) bond order parameters.[175] These work via an averaging of the spherical harmonics of a chosen order, over the molecules within the first coordination shell. Commonly the order chosen is 3, 4 or 6 for identifying cubic and hexagonal crystal structure. The Steinhardt parameter, of order l, for a molecule i, is defined as follows:

$$q_l(i) = \sqrt{\frac{4\pi}{2l+1}\sum_{m=-l}^{l}|q_l^m(i)|^2} \in \mathbb{R} \tag{8}$$

with the complex sub-parameters $q_l^m(i)$ defined thus:

$$q_l^m(i) = \langle Y_l^m(\mathbf{r_{ij}})\rangle_{j\in\mathbf{1}_i} \in \mathbb{C} \tag{9}$$

where $\langle\cdot\rangle_{j\in\mathbf{1}_i}$ denotes an ensemble average over the first coordination shell, *excluding i itself*, $\mathbf{r_{ij}}$ is the vector from molecule i to molecule j and Y_l^m indicate the spherical harmonics. Note that for the Steinhardt parameters, molecules are considered as single points in space — we use the position of the oxygens as their centres. Equations (8) and (9) are defined in such a way as to make $q_l(i)$ rotationally invariant by the properties of the spherical harmonics. The normalisation constant $4\pi/(2l+1)$ can be split into two parts. The numerator 4π comes from the 0-th order spherical harmonic[†] $Y_0^0(\mathbf{r_{ij}}) = 1/\sqrt{4\pi}$. The denominator $2l+1$ exists to average the $|q_l^m(i)|^2$ over $m \in \{-l, \ldots, l\}$.

4.1.1 Averaged local bond order parameters: *Lechner and Dellago*

Lechner and Dellago[176] proposed an alternative version ($\overline{q}_l(i)$) of these parameters, averaging the complex vectors $q_l^m(i)$ over a particle's first coordination shell as follows:

$$\overline{q}_l(i) = \sqrt{\frac{4\pi}{2l+1} \sum_{m=-l}^{l} |\overline{q}_l^m(i)|^2} \in \mathbb{R} \tag{10}$$

with

$$\overline{q}_l^m(i) = \langle q_l^m(k) \rangle_{k \in \overline{\mathbf{1}}_i} \in \mathbb{C} \tag{11}$$

where $\langle \cdot \rangle_{k \in \overline{\mathbf{1}}_i}$ denotes an ensemble average over the first coordination shell *including i itself*, note that this is *different* to the ensemble average from equation (9), which did not include the molecule i.[‡] The advantage to these parameters as compared to the regular Steinhardt parameters is that, by taking into account the second shell around each molecule, they can provide a more accurate determination of different structures. The do however have the downside of coarsening the resolution, as they take into account molecules further away and thus are excellent for categorising bulk materials, but less good for smaller clusters.[176]

4.1.2 Averaged local bond order parameters: *Li et al.*

Li *et al.*[177] also proposed an enhanced version of the Steinhardt parameters, similarly to Moore *et al.*[178]. They are defined as follows:

$$\widetilde{q}_l(i) = \Re\left(\left\langle \frac{\vec{q}_l(i) \cdot \vec{q}_l^{\,*}(k)}{|\vec{q}_l(i)||\vec{q}_l^{\,*}(k)|} \right\rangle_{k \in \mathbf{1}_i}\right) \in \mathbb{R} \tag{12}$$

where $\vec{q}_l(i)$ is the $(2l+1)$-dimensional vector:

$$\vec{q}_l = (q_l^{-l}, q_l^{-l+1}, \ldots, q_l^{l-1}, q_l^{l}) \in \mathbb{C}^{2l+1}, \tag{13}$$

$\vec{q}_l^{\,*}(i)$ is the corresponding vector of complex conjugates $q_l^m(i)^*$ and $\Re(z)$ is the real part of the complex vector z. These parameters average the dot products of q_l^ms amongst the first coordination shell in order to take into account orientational order of the first

---

[†] The (constant and real) value of the 0-th order spherical harmonic can in turn be deduced from the fact that $\int_0^{2\pi} \int_0^{\pi} Y_0^0(\theta,\phi) Y_0^0(\theta,\phi)^* \sin(\theta) \mathrm{d}\theta \mathrm{d}\phi = 1$, where z^* means the complex conjugate of z, in this case the same since $Y_0^0(\theta,\phi) \in \mathbb{R}$.

[‡] Under this notation, $\mathbf{1}_i$ is the first coordination shell of molecule i, while $\overline{\mathbf{1}}_i = \mathbf{1}_i \cup \{i\}$.

coordination shell in a similar way to the previous averaged parameters. Computing these three variants of the Steinhardt bond order parameters is discussed in section 6.1.

4.2 A comparison of the different bond order parameters

A visual comparison for the three sets of parameters can be found in figure 11. It is worth noting that this figure is produced using model systems of fully coordinated water, with liquid water simulated at 300 K, ice at 265 K and clathrate water at 230 K. Under-coordination has a large effect on the bond order parameters, with proximity to an interface having an especially large impact on the two sets of averaged local bond order parameters. This is due to their definition effectively using two coordination shells.

We are particularly interested in good separation between liquid water and ice. In the case of the regular versions q_l, the distribution of values for ice are fairly narrow and could reasonably be used for filtering, however it would leave a lot of non-icy molecules as the distribution of values for liquid water are very broad. The second set of parameters, $\overline{q}_l$ parameters perform much better at distinguishing liquid water from ice, particularly $\overline{q}_3$ and $\overline{q}_6$. Finally, the third set of parameters $\widetilde{q}_l$ also show good separation between liquid water and ice, particularly $\widetilde{q}_3$ and $\widetilde{q}_6$. The fourth order parameter $\widetilde{q}_4$ shows the best separation between the four states: liquid, clathrate, $\mathrm{I_h}$and $\mathrm{I_c}$for an individual parameter, with clathrate water typically overlapping some other state in other cases. Interestingly, for the averaged parameters clathrate water appears to overlap with liquid water, while for the regular q_l parameters, the distribution of values for clathrate water looks very similar to that of ice. This could potentially be leveraged to separate the four states more clearly by selecting one regular and one averaged parameter in a two-dimensional parameter.

4.3 Lattice order parameters

Two additional order parameters that can be used to look at the structure of water are the three and four-body lattice order parameters, F_3 and F_4; defined respectively by Báez and Clancy[179] and Rodger *et al.*[180]

4.3.1 Three-body lattice order parameter

The three-body lattice order parameter, $F_3(i)$, for a given water molecule i, considers the O···O···O angles between triplets of water molecules, with the given molecule being central. It is defined by the following equation:[179]

$$F_3(i) = \sum_{\Lambda \in \mathbf{1}_i^2} \left(|\cos(\theta_{i\Lambda})| \cos(\theta_{i\Lambda}) - \cos^2(109.47°) \right)^2, \tag{14}$$

where $\mathbf{1}_i^2$ is the set of all pairs of *distinct* water molecules in the first coordination shell of i, *excluding i itself*, and given such a pair Λ, $\theta_{i\Lambda}$ is the angle between the oxygen of the

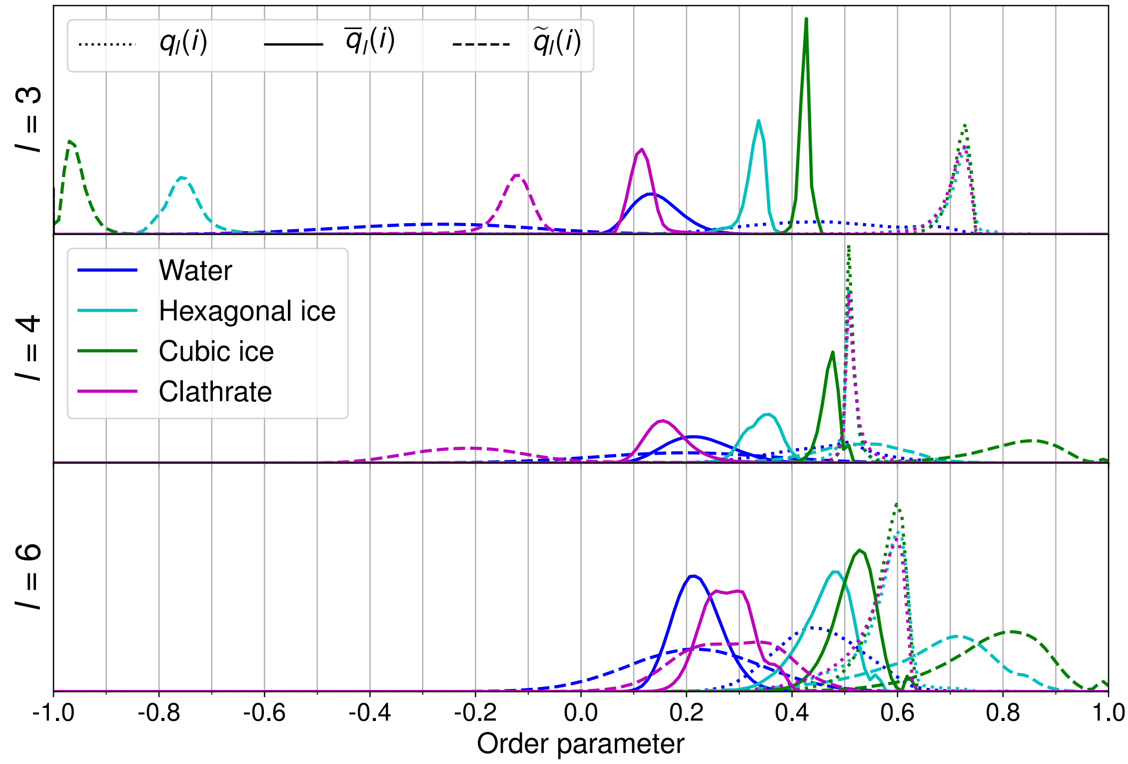


Figure 11: A comparison, via probability density plots, of the 3rd, 4th and 6th order Steinhardt parameters (dotted lines), with the averaged versions proposed by Lechner and Dellago[176] (solid lines) and the averaged versions proposed by Li *et al.*[177] (dashed lines). Blue indicates liquid water (at 300 K), cyan hexagonal ice (at 265 K), green cubic ice (at 265 K) and magenta clathrate water (at 230 K).

central molecule i and the oxygen atoms of Λ. Note that 109.47° is the angle found in a perfect tetrahedral lattice, thus a perfect ice structure will yield F_3 values of exactly zero.

4.3.2 Four-body lattice order parameter

The four-body lattice order parameter, $F_4(i)$, for a given water molecule i, considers the H−O⋯O−H torsional angles between pairs of water molecules. It is defined by the following equation:[180]

$$F_4(i) = \langle \cos(3\varphi_{ij}) \rangle_{j \in \mathbf{1}_i}. \tag{15}$$

Here, $\langle \cdot \rangle_{j \in \mathbf{1}_i}$ indicates an ensemble average across the first coordination shell of i, *excluding i itself* and φ_{ij} denotes the H−O⋯O−H torsional angle for water molecules i and j, choosing the outermost hydrogen atoms from each molecule.

4.3.3 Lattice order parameters in combination

By treating F_3 and F_4 as a two-dimensional order parameter, we are able to achieve good separation between liquid water, ice and clathrate (see figure 12) — the distinction between liquid and clathrate in particular being something that the bond order parameters struggle with. This two dimensional parameter is fairly good at distinguishing between liquid, clathrate and ice — as with many of the Steinhardt parameters, the distribution of values for water is fairly broad, however the distributions for clathrate and ice are much more focussed and distinct. As such, filtering to the ice (or clathrate) region would remove the almost all of the clathrate (or ice) molecules and a large proportion of the liquid water.

It is F_4 alone which is responsible for the clear separation between ice and clathrate, however F_3 will help such a filtering algorithm by removing most of the water. It is important however to recognise that F_3 is especially sensitive to coordination — as it is defined as a sum, rather than an ensemble average. Of all the parameters defined, F_4 is likely the least affected by coordination.

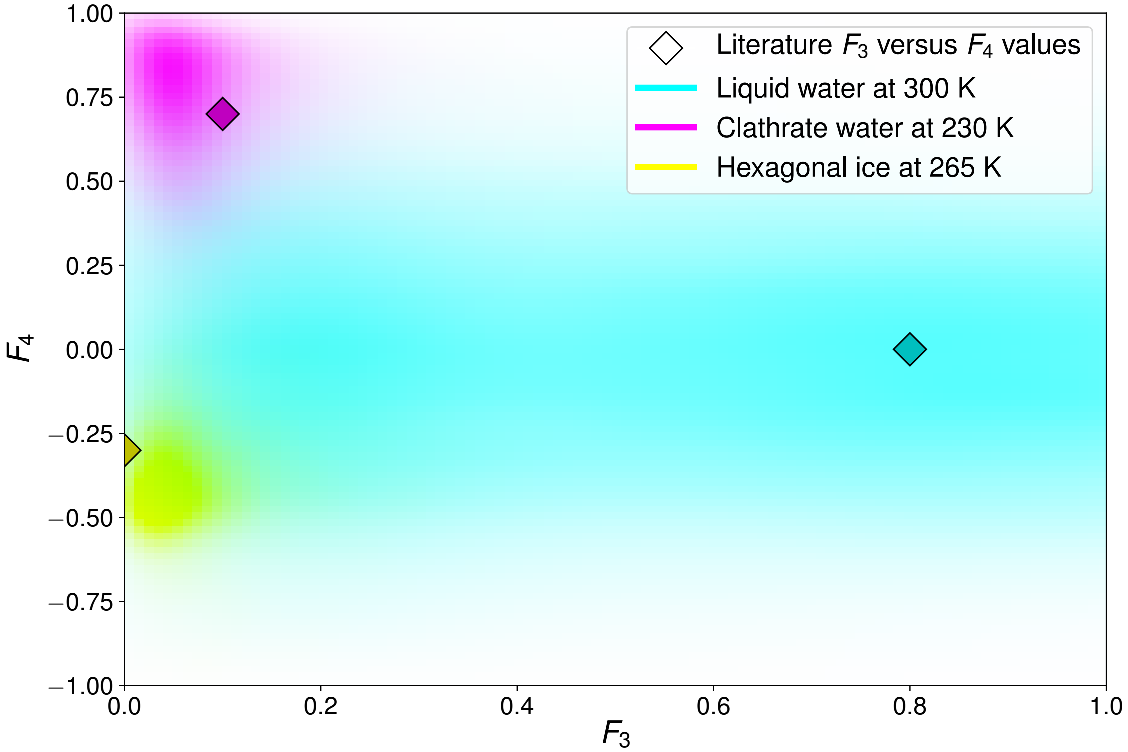


Figure 12: Two dimensional distribution of lattice order parameters F_3 and F_4 for model water systems. Literature values, as reported by Parui and Jana[181], are marked with diamonds.

4.4 Orientational order parameter

The orientational order parameter, θ, for a given water molecule, corresponds to the angle defined between the dipole moment (acting from positive to negative charge density) and the bilayer normal, i.e. the vector pointing outwards from the bilayer at a 90° angle from the bilayer plane. An angle of $\theta = 0°$ indicates that the dipole moment is pointing perpendicularly away from the bilayer, while an angle of $\theta = 180°$ indicates where the dipole moment points toward the bilayer (see figure 13).

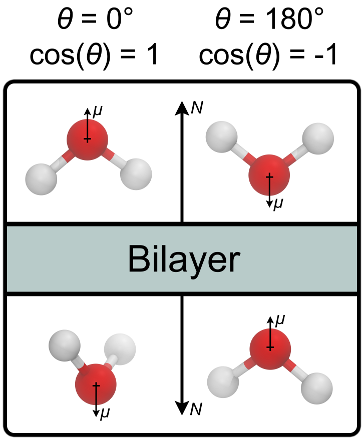


Figure 13: Visual schematic of water dipole orientation at $\theta = 0°$ and $\theta = 180°$. The dipole moment (μ) for each water molecule and the bilayer normal (N) on either side are displayed.

4.4.1 Orientation of ideal water

One can mathematically derive the distribution of θ values for ideal water (i.e. a water box where one is equally likely to find any possible orientation when picking a random water molecule). In a mathematical sense, for the water to be perfectly ideal, this box must be assumed to be infinite in either space or time.[*] To derive this distribution, we define a coordinate system with some water molecule[†] (i) as the origin, the xy-plane being the tangential plane of the bilayer and the z axis being normal to the bilayer. Now, we label the unit vector pointing out of the molecule in the direction of its dipole moment $\mathbf{r}(i)$.[‡] All possible such vectors map out a unit sphere, $S = \{\mathbf{x} \in \mathbb{R}^3 \mid \|\mathbf{x}\| = 1\}$, surrounding the molecule.

As orientation is a continuous parameter, we must consider the probability density rather than the probability of a given orientation. As explained above, we are considering the unit sphere S to be the set of all possible orientations; furthermore it is instinctively evident that any two regions on the sphere with the same surface area should correspond to two ranges of orientations with the same total probability. Thus if we have a (two-dimensional) range of orientations which maps onto a region, $\Pi \subset S$, the probability that any given molecule, i, is oriented within that same range is:

$$\mathbb{P}(i \in \Pi) = \frac{A(\Pi)}{A(S)} = \frac{A(\Pi)}{4\pi} \tag{16}$$

where $A(\Gamma)$ denotes the area of a surface Γ.

We can parameterise these normalised vectors using spherical coordinates $\theta \in [0, \pi]$,

---

[*] In fact, given that such an ideal system is both homogeneous and at equilibrium, infinite space and infinite time, or indeed infinite space *and* time, are mathematically equivalent (see appendix A).

[†] We are considering the water molecule as a single point.

[‡] I.e. $\mathbf{r}(i) = \frac{\mu}{\|\mu\|} \in S$, where μ is the dipole moment for molecule i.

$\phi \in [0, 2\pi)$, where θ is the same orientational order parameter as defined earlier and ϕ is the angle of the projection of $\mathbf{r}(i)$ onto the xy-plane.[*] We are not, in fact, interested in the value of ϕ, so will define regions using only θ, it is implied that each region allows all possible values of ϕ. Let a and b be values such that $0 \leq a \leq b \leq \pi$ and let Π be the region of the sphere defined by $\Pi = \{\mathbf{r}(\theta, \phi) \in S \mid a \leq \theta \leq b\}$. Then we have the following expression for the area of Π:

$$A(\Pi) = \iint_{\Pi} \left\| \frac{\partial \mathbf{r}}{\partial \theta} \times \frac{\partial \mathbf{r}}{\partial \phi} \right\|. \tag{17}$$

We write $\mathbf{r}$, and its partial derivatives, in terms of θ and ϕ:

$$\mathbf{r}(\theta, \phi) = \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix}, \quad \frac{\partial \mathbf{r}}{\partial \theta}(\theta, \phi) = \begin{pmatrix} \cos\theta\cos\phi \\ \cos\theta\sin\phi \\ -\sin\theta \end{pmatrix}, \quad \frac{\partial \mathbf{r}}{\partial \phi}(\theta, \phi) = \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix}.$$

Now, applying equation (17), we get the following expression for the area of Π:

$$\begin{aligned} A(\Pi) &= \int_a^b \int_0^{2\pi} \left\| \begin{pmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{pmatrix} \times \begin{pmatrix} -\sin\theta\sin\phi \\ \sin\theta\cos\phi \\ 0 \end{pmatrix} \right\| \mathrm{d}\phi\,\mathrm{d}\theta \\ &= \int_a^b \int_0^{2\pi} \left\| \begin{pmatrix} \sin^2\theta\cos\phi \\ \sin^2\theta\sin\phi \\ \cos\theta\sin\theta \end{pmatrix} \right\| \mathrm{d}\phi\,\mathrm{d}\theta = \int_a^b \int_0^{2\pi} \sqrt{\sin^2\theta}\,\mathrm{d}\phi\,\mathrm{d}\theta \\ {}^{\dagger} &= \int_a^b \int_0^{2\pi} \sin\theta\,\mathrm{d}\phi\,\mathrm{d}\theta = 2\pi \int_a^b \sin\theta\,\mathrm{d}\theta = 2\pi(\cos a - \cos b) \end{aligned}$$

And thus we arrive at a final expression for the probability for a random molecule having orientation $a \leq \theta \leq b$:

$$\mathbb{P}(a \leq \theta \leq b) = \mathbb{P}(i \in \Pi) = \frac{\cos a - \cos b}{2}. \tag{18}$$

4.4.2 The probability density function of θ

From equation (18), we can derive the cumulative density function (CDF):

$$F(\theta) = \frac{\cos 0 - \cos\theta}{2} = \frac{1}{2} - \frac{1}{2}\cos\theta, \tag{19}$$

and from this we can differentiate to give the probability density function (PDF):

$$f(\theta) = \frac{\mathrm{d}F}{\mathrm{d}\theta}(\theta) = \frac{1}{2}\sin\theta. \tag{20}$$

---

[*] Note that in this section we use angles in radians rather than degrees to simplify the derivation.

[†] We know $\sin\theta \geq 0$ since $0 \leq \theta \leq \pi$.

For the purpose of using orientation as an order parameter, it would be preferable to be able to understand the results without comparison to ideal water; thus we shall define a new normalised order parameter ϑ, such that θ and ϑ are in one-to-one correspondence and so that the PDF, $g(\vartheta)$ is constant.

Finding such a parameter is in fact rather easy, we define $\vartheta = \cos\theta$, which is clearly a strictly decreasing one-to-one correspondence (bijection) on the domain $\theta \in [0, \pi]$. If α and β are values such that $-1 \leq \alpha \leq \beta \leq 1$, then from equation (18) we obtain the expression for the probability that ϑ falls between α and β:

$$\mathbb{P}(\alpha \leq \vartheta \leq \beta) = \frac{\beta - \alpha}{2}, \tag{21}$$

noting that $\alpha \leq \vartheta \leq \beta$ is equivalent to $\cos^{-1}\alpha \geq \theta \geq \cos^{-1}\beta$, since cos is a strictly decreasing function on the domain $[0, \pi]$. Thus we obtain the CDF and PDF, respectively, as follows:

$$G(\vartheta) = \frac{\vartheta - 0}{2} = \frac{\vartheta}{2}, \quad g(\vartheta) = \frac{\mathrm{d}G}{\mathrm{d}\vartheta}(\vartheta) = \frac{1}{2}. \tag{22}$$

Note that, for this revised order parameter, $\vartheta = -1$ corresponds to the dipole moment pointing into the bilayer, $\vartheta = 1$ corresponds to the dipole moment pointing out of the bilayer and $\vartheta = 0$ corresponds to the dipole moment pointing parallel to the plane of the bilayer (see figure 13). The parameter $\cos\theta$ has in fact been used in a number of studies, albeit often with no explanation for its preference over θ.[182,183]

4.5 Compound order parameters

Often a single parameter can be good at separating between two phases, but ineffective at separating between others. Since all the parameters defined so far are one-dimensional, the likelihood of overlap between different regions is great. If we take two of the more promising parameters and treat them as a single two-dimensional parameter, as we did in section 4.3.3 with F_3 and F_4, we often find that we can obtain better separation. This comes with its downsides however: visualising two-dimensional data is difficult; particularly if we wish to look at the trends over time or a spacial axis. Thus if we could reduce the parameter back down to one dimension, while maintaining the separation, this is very desirable. There are many methods to perform such a dimensionality reduction, from fields such as linear algebra[184] and machine learning;[185] however, one can perform a simpler reduction from two dimensions to one by hand. There are a few natural methods for this; one example is to simply rotate the system and then project onto the x-axis.

In figure 11, we saw that the 3rd and 6th order local bond parameters $\overline{q}_3$ and $\overline{q}_6$ were particularly good at distinguishing between ice and water and $\overline{q}_3$ was reasonably good at separating the hexagonal and cubic ice phases. Figure 14 shows these two parameters on a two-dimensional grid. It appears that such method of rotation followed by projection could work fairly well, however this does not appear to be the best method for separation preserving dimensionality reduction. Instead, notice that we can draw two lines (as shown in the figure) splitting the domain into four regions. Numbering counter-clockwise, starting with the right hand region, region 1 contains a great majority of cubic ice, region 2 contains the majority of hexagonal ice and region 3 the majority of liquid water and clathrate.

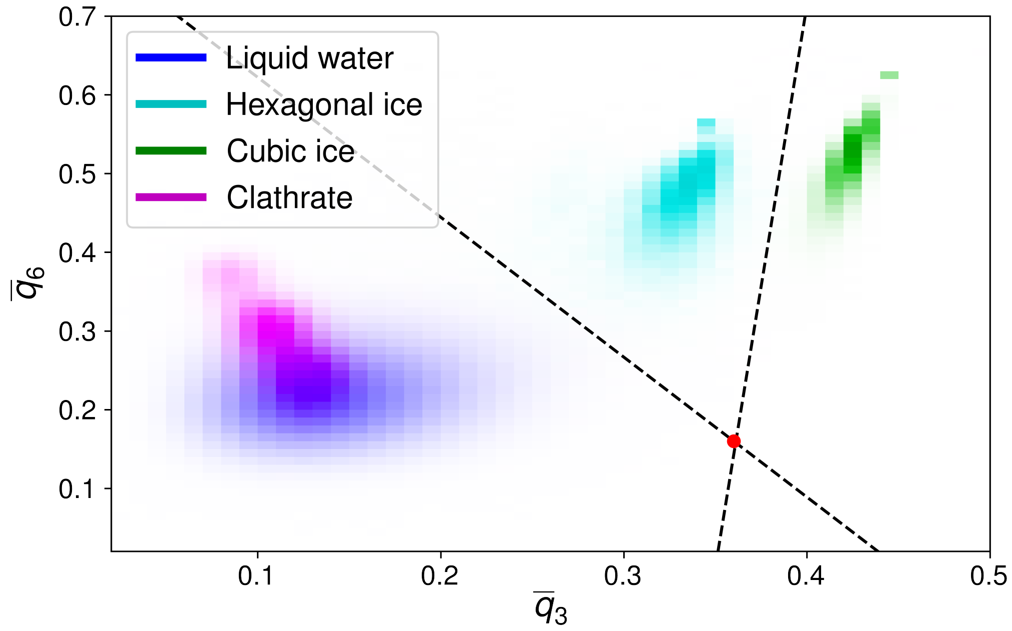


Figure 14: Two-dimensional distribution of $\overline{q}_3$ and $\overline{q}_6$ values for model water systems. Two lines are drawn, separating the hexagonal and cubic ice regions and the liquid/clathrate region. The intersection of these two lines $(0.36, 0.16)$ is marked in red.

Now we have produced this separation, we can reduce to a single dimension, by defining a new parameter, ϕ, which is based on the angle of the vector going from the intersection point between the two lines (plotted in red, figure 14) to $(\overline{q}_3, \overline{q}_6)$:

$$\phi(i) = \frac{1}{\pi}\cos^{-1}\left(\frac{\overline{q}_3(i) - 0.36}{\sqrt{(\overline{q}_3(i) - 0.36)^2 + (\overline{q}_6(i) - 0.16)^2}}\right). \tag{23}$$

Note that ϕ is defined in such a way that it takes values between 0 and 1, although this is not a necessity. The distribution of ϕ values for model systems is shown in figure 15. In fact, this compound parameter has an error rate lower than 0.1% in differentiating between ice and non-icy water, in such model systems. It is also good at differentiating between hexagonal and cubic ice.

It is worth however noting that for the model systems, liquid water was simulated at 300 K, ice at 265 K and clathrate at 230 K. It would be worth studying the effect of temperature on these distributions. In particular, supercooled liquid water does show a greater degree of local bond order than under standard conditions, albeit still lesser than ice. For these specific systems, the compound parameter does achieve a better degree of separation, however this is definitely as a result of overfitting and water within real systems is far more difficult to classify, hence the loose requirements for water to be considered "icy" in section 5.1. As such, the increased computational cost to compute such a compound parameter is unlikely to be worth it.

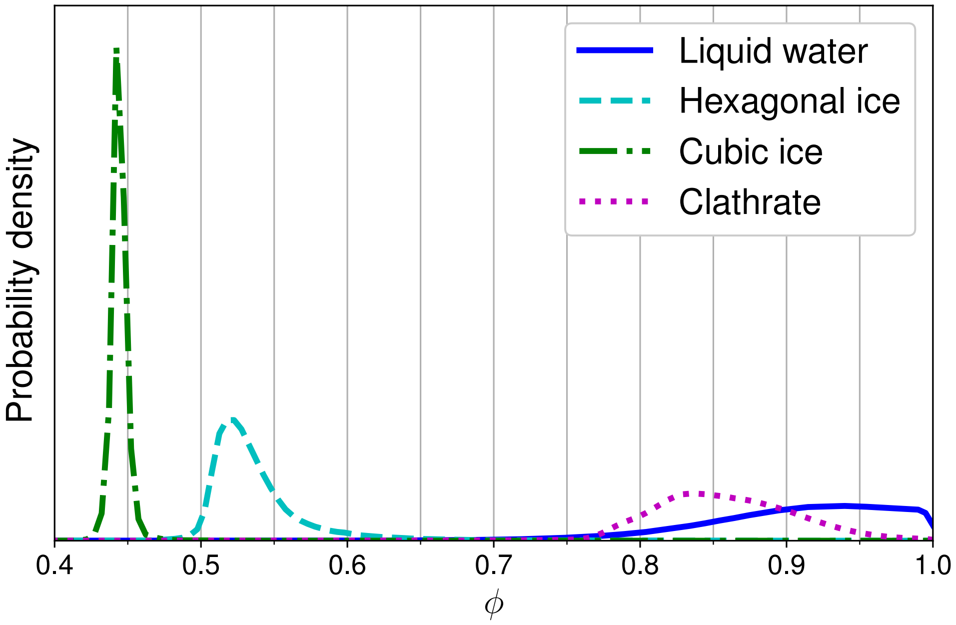


Figure 15: Distribution of compound parameter ϕ values for model water systems.

Chapter 5

Analysis Methods

5.1 Largest icy cluster per frame

The largest "icy cluster" per frame was computed using the PLUMED 2.3.0 software, compiled with the crystallization module.[186–188] These clusters were computed as follows:

1. Filter all water molecules by PLUMED's implementation of a local q_6 parameter (see section 4.1).[*]
2. Compute contact matrix between molecules left after filtering.
3. Use the contact matrix to build a list of clusters via a depth first search.
4. Determine which cluster is the largest.

Note that for the DPPC-CHL and DPPC-DMPI systems, the two water layers are treated as one so it is truly the largest cluster per frame; for the LPS systems, which are asymmetric, the two water layers are treated separately, so there are two largest clusters per frame, one from each side. Looking at pre-critical icy clusters, in this way, has been shown to be a valid method for quantitative analysis of the ice nucleating ability of a given particle or surface; as evidenced by a number of previous studies.[24,192,193] An example of a PLUMED input to compute these is given in appendix B. Note that these clusters are pre-critical and an individual cluster rarely persists for more than a frame (4 ps) or two.

5.2 Hydrogen bonds

In a recent study, Sosso *et al.*[21] found that CHL molecules' role as ice nucleating agents is partially due to their amphoteric hydroxy groups, which are capable of both donating and accepting hydrogen bonds. In that work it was observed that ice nuclei would form, with those hydroxy groups being a member of the cluster, due to their chemical similarities to water. Clearly then the formation of, or lack thereof, hydrogen bonds between icy clusters

---

[*] While other parameters described in chapter 4, for example q_3, could also be good options for this step, q_6 has been used in numerous previous works for this purpose.[177,189–191]

and a bilayer is an important factor to consider. Thus the number of hydrogen bonds in each frame, between the largest icy cluster (as defined in the previous section) and the bilayer, was computed. A hydrogen bond X−H···Y was defined via to the geometric criteria: $d(\mathrm{X}, \mathrm{Y}) \leq 3.2\,\text{Å}$ and $\theta \geq 150°$, where θ is the angle between X, H and Y.

It is worth noting that the lipids in question have varying numbers of hydrogen bonding sites. CHL has only one, although it can both donate and accept hydrogen bonds, but as previously mentioned is known to be active in the nucleation process. DPPC has a total of eight bonding sites, although these are all only capable of accepting hydrogen bonds, while DMPI has thirteen bonding sites, five of which can donate or accept hydrogen bonds. Another factor to consider is the accessibility of these sites, while the CHL hydroxy group is active in the case of CHL crystals[21] and monolayers,[24] in mixed DPPC-CHL bilayers, the hydroxy group is often poorly hydrated and therefore not close enough to any icy clusters to form hydrogen bonds.

5.3 Bilayer ordering

The degree of structural order within a bilayer is likely to play a large role in the ice nucleating potential. Sosso *et al.*[24] found that for CHL monolayers, the ice nucleating ability was directly proportional to the degree of structural order. From a chemical perspective, it is also clear that the way in which headgroups are packed will have an effect on the potential to form hydrogen bonds, thereby improving or lessening a bilayer's effect on ice nucleation. In this section, two measures for structural order within lipid bilayers are defined. The first, the single molecule angle criteria (SMAC), is a more novel approach to studying bilayers; while the second, the Voronoi accessible area per lipid, has been frequently employed in similar computational studies of bilayers.[30,133,194,195]

5.3.1 SMAC collective variable

The degree of order within the bilayers can be assessed in multiple ways. The SMAC parameter is a collective variable (CV) which measures orientational order with respect to a defined molecular axis.[196] In this case, we define SMAC so that it is high, approaching 1, for systems where lipid tails are close to parallel; and lower for systems where lipids are more chaotically oriented. To define the SMAC parameter s_i, for a particular lipid molecule i, first one must define a molecular axis. In the case of these simulations, the axis is chosen to follow the (first) tail of each lipid in the bilayer with the exception of lipid A, where we use each of the four primary chains and PVCL2, where we use both of the primary chains. For example, in the case of DPPC, the axis is defined as being the vector from the central carbon atom* to the penultimate carbon in the primary tail (C_{15}, see figure 3). The choice of these axes was intended to reduce noise from movement and

---

*The carbon which is bonded to the headgroup, primary tail oxygen and secondary tail carbon.

packing of the tails and the numbers of axes were intended to balance the fact that some of the lipids in question have more tails than others.

We define a switching function f, which acts on the distance r_{ij} between two lipids i and j:

$$f(r_{ij}) = \frac{1}{1+(2r_{ij})^6}, \tag{24}$$

and set $n_i = \Sigma_{i\neq j} f(r_{ij})$. Now we define another switching function ψ, this time acting on the size of the coordination shell:

$$\psi(n_i) = \exp\left(-\frac{n_i}{6}\right). \tag{25}$$

Finally we define a Gaussian kernel function K, acting on the angle θ_{ij} defined between the molecular axes of two lipids i and j:*

$$K(\theta_{ij}) = \exp\left(-\frac{\theta_{ij}^2}{2\sigma^2}\right), \tag{26}$$

where σ is the width of the Gaussian; here we used $\sigma = 0.58$. Now we are ready to define the SMAC parameter:

$$s_i = (1-\psi(n_i)) \cdot \frac{\Sigma_{i\neq j} f(r_{ij}) K(\theta_{ij})}{n_i}. \tag{27}$$

The average SMAC parameter was computed using the SMAC collective variable from the PLUMED 2.3.0 software, compiled with the crystallization module.[186–188] This CV has typically been used in the past for nucleation studies and molecules in solution rather than for membrane ordering.[197] The PLUMED input for computing the SMAC parameter is given in appendix C.

5.3.2 Voronoi area per lipid

A second measure for the spatial ordering of bilayers is to compare the average surface area per lipid. This is achieved by computing the accessible area for each lipid via the construction of a Voronoi tessellation.[133,194] In the case of CHL, the lone oxygen is used as the vertex, whereas for DPPC and DMPI the central glycerol carbon and the first carbon in each tail are used as vertices; this is to account for the difference in size between cholesterol and the phospholipids. These accessible areas were not computed for the LPS systems, as the leaflet we are interested in (lipid A with or without sugars) is homogeneous in-plane. For these systems, the average area per lipid A is computed by dividing the total surface area by the number of lipid A molecules. The accessible areas were computed using the built in area per lipid function of the FATSLiM package.[198]

---

[*] I.e. $\cos(\theta_{ij}) = \mathbf{u}\cdot\mathbf{v}$, where $\mathbf{u}$ and $\mathbf{v}$ are unit vectors in the direction of the molecular axes of molecules i and j respectively (as defined at the start of section 5.3.1).

5.4 Geometric analysis

The method detailed in section 5.1 is highly effective in assessing the pre-critical nucleation of ice, without requiring immense computational resources. However if one wants to examine these clusters in more detail, there are various geometrical criteria that can be used. Ice nuclei are formed of two key cage structures: double diamond and hexagonal cages, the building blocks of cubic and hexagonal ice, respectively. Both of these cage structures are formed of six-membered, hydrogen bonded rings.[199] Thus a natural first geometry to consider is the prevalence of six-membered (hydrogen bonded or not) rings. One can then look at whether these rings assemble any double diamond or hexagonal cages. In addition to looking for ice structures, one can look for the five and six-membered ring cage structures, characteristic of clathrates.[200]

5.4.1 Rings and cages analysis

Using the R.I.N.G.S. code,[201] the number of five and six-membered rings within the largest icy cluster was computed for each frame. Using code developed within the Sosso group (see chapter 6),[202] which I have extensively modified, the number of these rings which were wholly hydrogen bonded and the number of double diamond and hexagonal cages formed by the six-membered rings was computed.

5.4.2 Clathrate clustering analysis

Using the five and six-membered rings identified by the R.I.N.G.S. code,[201] an algorithm to look at clathrate cages was implemented into code (see chapter 6).[202] This was based on the method described by Bi and Li[200] where they first looked for 5^6, $5^6 6^1$ and $5^6 6^2$ half-cages, and then looked for such half-cages which could form a full cage of either 5^{12}, $5^{12}6^2$, $5^{12}6^3$ or $5^{12}6^4$ structure. These full cage structures are illustrated in the form of two-dimensional Schlegel diagrams in figure 16(a)–(d).

The code written does not compute these half or full cages, instead looking for “555”, “655” and “6556” “part-cages,” (see figure 16(e)–(g)) however it would be a fairly simple extension to use these to compute the half or full cages. Nonetheless, locating these three types of part-cages allows for rendering such cages using a visualisation package and the number of such part-cages provides a good indication of the amount of clathrate-like water in the system. Perhaps unsurprisingly, for the systems studied in this work there were not many part-cages found; hence why this particular analytic tool was not extended further.

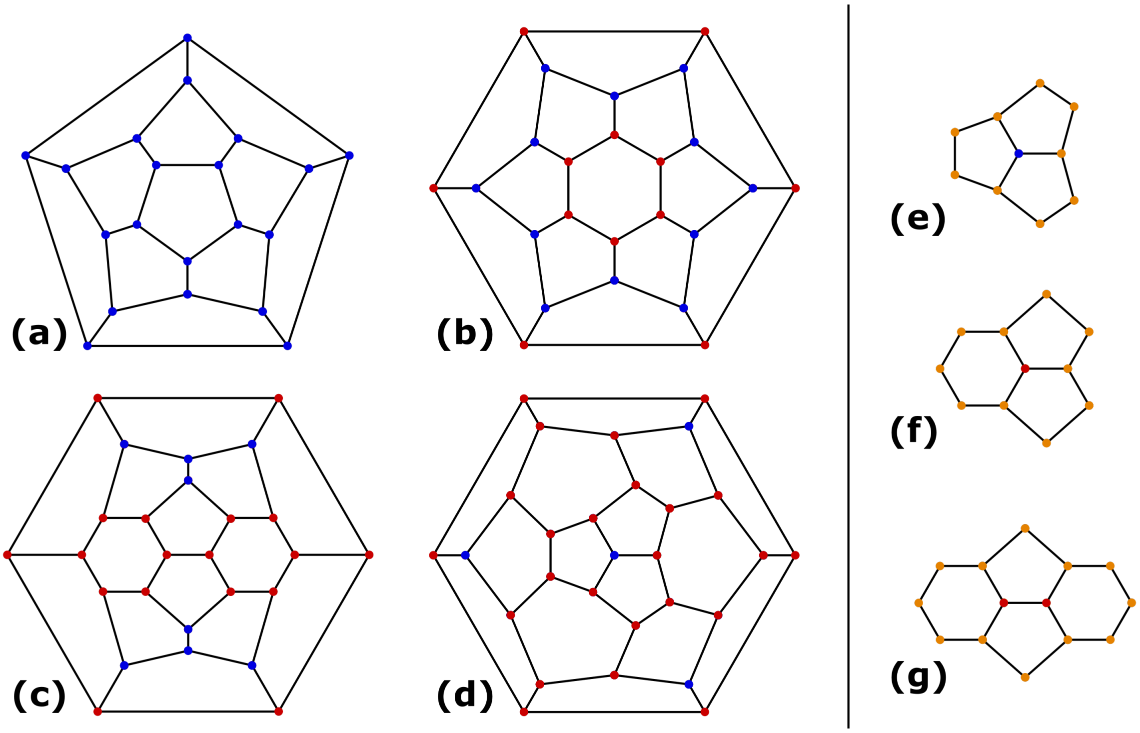


Figure 16: The left panel shows two-dimensional Schlegel diagrams for the clathrate cage structures: (a) 5^{12}, (b) $5^{12}6^2$, (c) $5^{12}6^3$ and (d) $5^{12}6^4$. Nodes are coloured blue if located in the centre of a 555 part-cage and red if located in the centre of a 655 part-cage. The right panel shows the three types of part-cage: (e) 555, (f) 655 and (g) 6556.

Chapter 6

Analysis algorithms

During the course of this PhD, I have written and edited a large body of code in the programming languages of FORTRAN and Python. In this chapter I detail some of the more complex algorithms which I have used within my analysis code. All of the code snippets within are in pseudocode for readability.

A large amount of the analysis content of this thesis pertains to a suite of analysis code "HIN", which was developed by various members of the Sosso group and is available on GitHub.[202] Within this code I have personally written the following modules:

- `bondorder`, which computes the Steinhardt parameters and averaged versions $q_l(i)$, $\overline{q}_l^m(i)$ and $\widetilde{q}_l^m(i)$ (see sections 4.1 and 6.1),
- `clathrates`, which computes the lattice order parameters F_3 and F_4 (see sections 4.2 and 6.2).

I have also extensively rewritten and in cases extended the modules:

- `order`, which computes the orientational order parameter θ (see section 4.4),
- `rings`, which computes the number of rings within a cluster or region of the system and the number of cages (see section 5.4.1).

In addition to those analytical components, I completely redesigned the input file system, changing it from a very long (and constantly growing as new modules were added) fixed format input file to a much more concise and humanly readable input format, also adding the option to display a command line progress bar.

6.1 Computing the Steinhardt bond order parameters

The three variants of Steinhardt bond order parameter, defined in section 4.1, were computed via a FORTRAN code. In this section, the algorithm is detailed. Full FORTRAN source code is available on GitHub.[202] The first step in computing any of the Steinhardt parameters, for an individual water molecule i, is to compute $q_l^m(i)$ for every value of $m \in \{-l, \dots, l\}$, see equation (9).

```
function qlm(i) {
    complex sigma = 0 + 0i # complex variable

    # loop over first coord shell (excluding i itself)
    for j in [first coord shell]:
        real (r, theta, phi) = position(j) - position(i) # spherical coordinates
        sigma += Ylm(l, m, theta, phi) # add the l,m spherical harmonic to the sum

    return sigma / [size of first coord shell] # divide sum by count to get avg.
}
```

Computation of the regular Steinhardt parameter, $q_l(i)$, see equation (8), is now simple with the following algorithm:

```
function ql(i) {
    real sigma = 0 # real variable

    # loop over m from -l to l
    for m in [-l,...,l]:
        sigma += |qlm(i)|^2 # add absolute square of qlm to the sum

    return sqrt((4pi*sigma) / (2l+1))
}
```

6.1.1 Computing the *Lechner and Dellago* local bond order parameters

To compute the second set of bond order parameters, first we need to compute $\overline{q}_l^m(i)$ for every value of $m \in \{-l, \ldots, l\}$, see equation (11), as we did for $q_l^m(i)$ in the previous section. This is achieved by simply summing the $q_l^m(j)$ for every molecule j within the fist coordination shell (including i itself). Once all the $\overline{q}_l^m(i)$ are computed, calculating $\overline{q}_l(i)$ is much the same as for $q_l(i)$:

```
function qlb(i) {
    real sigma = 0 # real variable

    # loop over m from -l to l
    for m in [-l,...,l]:
        sigma += |qlmb(i)|^2 # add absolute square of q^-lm to the sum

    return sqrt((4pi*sigma) / (2l+1))
}
```

6.1.2 Computing the *Li et al.* local bond order parameters

The third set of parameters are defined somewhat differently from the first two, see equation (12). The algorithm for computing $\widetilde{q}_l(i)$, for a water molecule i, is detailed below:

```
function qlt(i) {
    complex sigma = 0 + 0i # complex variable
    real qi_squared = 0 # real variable

    # loop over m from -l to l
    for m in [-l,...,l] :
        qi_squared += |qlm(i)|^2

    # loop over first coord shell (excluding i itself)
    for j in [first coord shell]:
        complex qi_dot_qj = 0 + 0i # complex variable
        real qj_squared = 0 # real variable

        # loop over m from -l to l
        for m in [-l,...,l]:
            qi_dot_qj += qlm(i) * (qlm(j)^*)
            qj_squared += |qlm(j)|^2

        sigma += qi_dot_qj/sqrt(qj_squared)

    return Re(sigma) / ([size first coord shell]*sqrt(qi_squared))
}
```

6.1.3 Computing the spherical harmonics

In the implementation of the code, the third, fourth and sixth spherical harmonics are hard-coded for performance reasons; however, if other orders were desired they can be computed via the equation:[203]

$$Y_l^m(\theta,\phi) = \left(\frac{2l+1}{4\pi}\frac{(l-m)!}{(l+m)!}\right)^{\frac{1}{2}} P_l^m(\cos\theta)e^{im\phi}, \tag{28}$$

where P_l^m are the associated Legendre polynomials, which obey and are fully described by the following four identities:

$$P_l^{-m}(x) = (-1)^m\frac{(l-m)!}{(l+m)!}P_l^m(x), \text{ for } l \in \mathbb{N},\ m \in \{1,\dots,l\}; \tag{29}$$

$$P_l^m(x) = (-1)^m(2m-1)!!(1-x^2)^{m/2}, \text{ for } m = l \in \mathbb{N}; \tag{30}$$

$$P_l^m(x) = x(2m+1)P_m^m(x), \text{ for } l \in \mathbb{N},\ m = l-1; \tag{31}$$

$$(l-m)P_l^m(x) = x(2l-1)P_{l-1}^m(x) - (l+m-1)P_{l-2}^m(x), \text{ for } l \in \mathbb{N},\ m \in \{0,\dots,l-2\}; \quad (32)$$

where $n!!$ indicates the double factorial of n, i.e. $n!! = 3 \cdot 5 \cdots (n-2) \cdot n$ for odd n and $n!! = 2 \cdot 4 \cdots (n-2) \cdot n$ for even n. Thus the associated Legendre polynomials can be computed recursively as follows:

```
recursive function Plm(l,m,x) {
    if m < 0: return (-1)^m * ((l+m)! / (l-m)!) * Plm(l,-m,x)
    else if m = l: return (-1)^m * (2m-1)!! * (1-x^2)^(m/2)
    else if m = l-1: return x * (2m+1) * Plm(m, m, x)
    else: return [x * (2l-1) * Plm(l-1,m,x) - (l+m-1)*Plm(l-2,m,x)] / (l-m)
}
```

Note that a negative m is immediately replaced with a positive m in the above algorithm and a positive m is always increased until it reaches l, whereby P_l^l is well defined. Thus this algorithm will converge for any positive integer l, and $m \in \{-l,\dots,l\}$.

6.2 Computing the lattice order parameters

The three and four-body lattice order parameters F_3 and F_4, see equations (14) and (15) respectively, were also computed via a FORTRAN code. Here those algorithms are detailed. As with the Steinhardt bond order parameters section, full FORTRAN source code is available on GitHub.[202]

6.2.1 Computing the three-body lattice order parameter

The three-body lattice order parameter, F_3, was computed, according to equation (14), using the equivalence $\mathbf{r}_{ij} \cdot \mathbf{r}_{ik} = |\mathbf{r}_{ij}||\mathbf{r}_{jk}| \cos(\theta_\Lambda)$, as detailed below:

```
function F3(i) {
    real sigma = 0 # real variable

    # loop over molecules j,k in first coord shell (with i,j,k distinct)
    for j,k in [first coord shell]:
        real cos2_num = r_ij.r_ik * |r_ij.r_ik| # |cos(th)|cos(th) numerator
        real cos2_den = |j|^2 * |k|^2 # |cos(th)|cos(th) denominator
        sigma += ( cos2_num/cos2_den - cos^2(109.47) )^2

    return sigma
}
```

6.2.2 Computing the four-body lattice order parameter

The four-body lattice order parameter, F_4, was computed, according to equation (15). The largest challenge in computing $F_4(i)$ for a water molecule i is calculating the H–O···O–H torsional angles, φ. The torsional angle for water molecules i and j is defined as being the angle formed between the two planes Π and Γ, so that $\mathbf{o}_i, \mathbf{o}_j, \mathbf{h}_i \in \Pi$ and $\mathbf{o}_i, \mathbf{o}_j, \mathbf{h}_j \in \Gamma$; where $\mathbf{o}_i$, $\mathbf{o}_j$, $\mathbf{h}_i$ and $\mathbf{h}_j$ are the positions of the oxygen atoms and outermost hydrogen atoms from molecules i and j. The intersection of these planes is the line L, containing both $\mathbf{o}_i$ and $\mathbf{o}_j$:

$$L = \{\mathbf{o}_i + \alpha \mathbf{o}_j \mid \alpha \in \mathbb{R}\}, \tag{33}$$

Now, the angle between planes Π and Γ is the same as the angle between two vectors $\mathbf{u} \in \Pi$ and $\mathbf{v} \in \Gamma$, chosen such that both are perpendicular to L.

For simplicity, we define the coordinate system such that $\mathbf{o}_i = \mathbf{0}$, i.e. $\mathbf{o}_j$, $\mathbf{h}_i$ and $\mathbf{h}_j$ are vector positions relative to $\mathbf{o}_i$, now $L = \{\alpha \mathbf{o}_j \mid \alpha \in \mathbb{R}\}$. Since the vectors $\mathbf{u}$ and $\mathbf{v}$ are perpendicular to L, we must have that $\mathbf{u} \cdot \mathbf{o}_j = 0$ and $\mathbf{u} \cdot \mathbf{o}_j = 0$. Note now that since setting $\mathbf{o}_i = \mathbf{0}$, we have:

$$\Pi = \{\alpha \mathbf{h}_i + \beta \mathbf{o}_j \mid \alpha, \beta \in \mathbb{R}\}, \quad \Gamma = \{\alpha \mathbf{h}_j + \beta \mathbf{o}_j \mid \alpha, \beta \in \mathbb{R}\}, \tag{34}$$

and since we are only interested in angles and not magnitudes, we can assume that $\mathbf{u}$ and $\mathbf{v}$ are of the form:

$$\mathbf{u} = \mathbf{h}_i + \lambda \mathbf{o}_j, \quad \mathbf{v} = \mathbf{h}_j + \mu \mathbf{o}_j, \quad \lambda, \mu \in \mathbb{R}. \tag{35}$$

Now

$$\mathbf{u} \cdot \mathbf{o}_j = (\mathbf{h}_i + \lambda \mathbf{o}_j) \cdot \mathbf{o}_j = \mathbf{h}_i \cdot \mathbf{o}_j + \lambda \mathbf{o}_j \cdot \mathbf{o}_j = \mathbf{h}_i \cdot \mathbf{o}_j + \lambda |\mathbf{o}_j|^2 = \mathbf{0}, \tag{36}$$

$$\mathbf{v} \cdot \mathbf{o}_j = (\mathbf{h}_j + \mu \mathbf{o}_j) \cdot \mathbf{o}_j = \mathbf{h}_j \cdot \mathbf{o}_j + \mu \mathbf{o}_j \cdot \mathbf{o}_j = \mathbf{h}_j \cdot \mathbf{o}_j + \mu |\mathbf{o}_j|^2 = \mathbf{0}. \tag{37}$$

Finally, we get values for λ and μ:

$$\lambda = -\frac{\mathbf{h}_i \cdot \mathbf{o}_j}{|\mathbf{o}_j|^2}, \quad \mu = -\frac{\mathbf{h}_j \cdot \mathbf{o}_j}{|\mathbf{o}_j|^2}, \tag{38}$$

and we can calculate the cosine of the torsional angle φ:

$$\cos(\varphi) = \frac{\mathbf{u} \cdot \mathbf{v}}{|\mathbf{u}||\mathbf{u}|}, \tag{39}$$

using the values obtained using $\mathbf{u} = \mathbf{h}_i + \lambda \mathbf{o}_j$, $\mathbf{v} = \mathbf{h}_j + \mu \mathbf{o}_j$. The algorithm for computing F_4 is given below in pseudocode:

```
function F4(i) {
    real sigma = 0 # real variable

    # loop over first coord shell (excluding i itself)
    for k in [first coord shell]:

        # get position vectors for oxygen atoms and outer hydrogen atoms
        vector oi = (0, 0, 0) # all positions relative to oxygen from molecule i
        vector oj = position([oxygen from molecule j])
```

```
        vector hi = position([hydrogen from molecule i, furthest from oj])
        vector hj = position([hydrogen from molecule j, furthest from oi])

        real lambda = -hi.oj / |oj|^2
        real mu = -hj.oj / |oj|^2

        vector u = hi + lambda*oj
        vector v = hj + mu*oj

        real cos_phi = u.v / |u||v|
        sigma += 4*cos_phi^3 - 3*cos_phi

    return sigma
}
```

Chapter 7

Results

In this chapter, the results from the DPPC-CHL/DMPI and LPS systems, supercooled to 233.15 K, are discussed. As introduced in sections 3.1 and 3.2, these are water-bilayer-water systems with a variety of different bilayers: DPPC-CHL bilayers from 0–100 mol% CHL, DPPC-DMPI bilayers from 0–100 mol% DMPI and asymmetric bilayers with one lipid A leaflet and one phospholipid leaflet. Extracting quantitative information from MD simulations is a difficult task, however the analytical tools described in chapters 4 and 5 can provide insight into the structure of the bilayers, as well as the amount of nucleation events occurring and the location of these occurrences.

7.1 Pre-critical ice nuclei

In section 5.1, methodology was detailed for computing the largest "icy cluster" occurring in each frame of the simulation trajectories. Figure 17 shows the distribution of sizes of these clusters for all the different systems. For the DPPC-CHL/DMPI systems, the largest cluster across both sides is considered as the bilayers are symmetrical. For the LPS systems, the largest cluster from each side is considered as the bilayers are asymmetrical (i.e. the two sides are treated as separate systems).

Note that the LPS systems are considerably larger, so one would expect to see larger clusters on average. This is due to both volume (or number of molecules) and area based factors. From a volumetric perspective, regardless of the bilayers ability or inability to promote ice nucleation, a larger number of water molecules makes it statistically more likely to see larger clusters. While it is true that splitting the LPS systems in half reduces this effect, each half of these systems still has around three to four times as many water molecules as the smaller systems. From the perspective that the LPS systems have a larger surface area, it follows that if the bilayer does promote ice nucleation, then the interface is larger and therefore one would expect to see more "icy" molecules and therefore larger clusters. It would theoretically be possible to normalise the data in order to account for this; however the effect of system size on the distribution of cluster sizes is a rather complex question and the mismatching scales are not a particular issue for the analyses

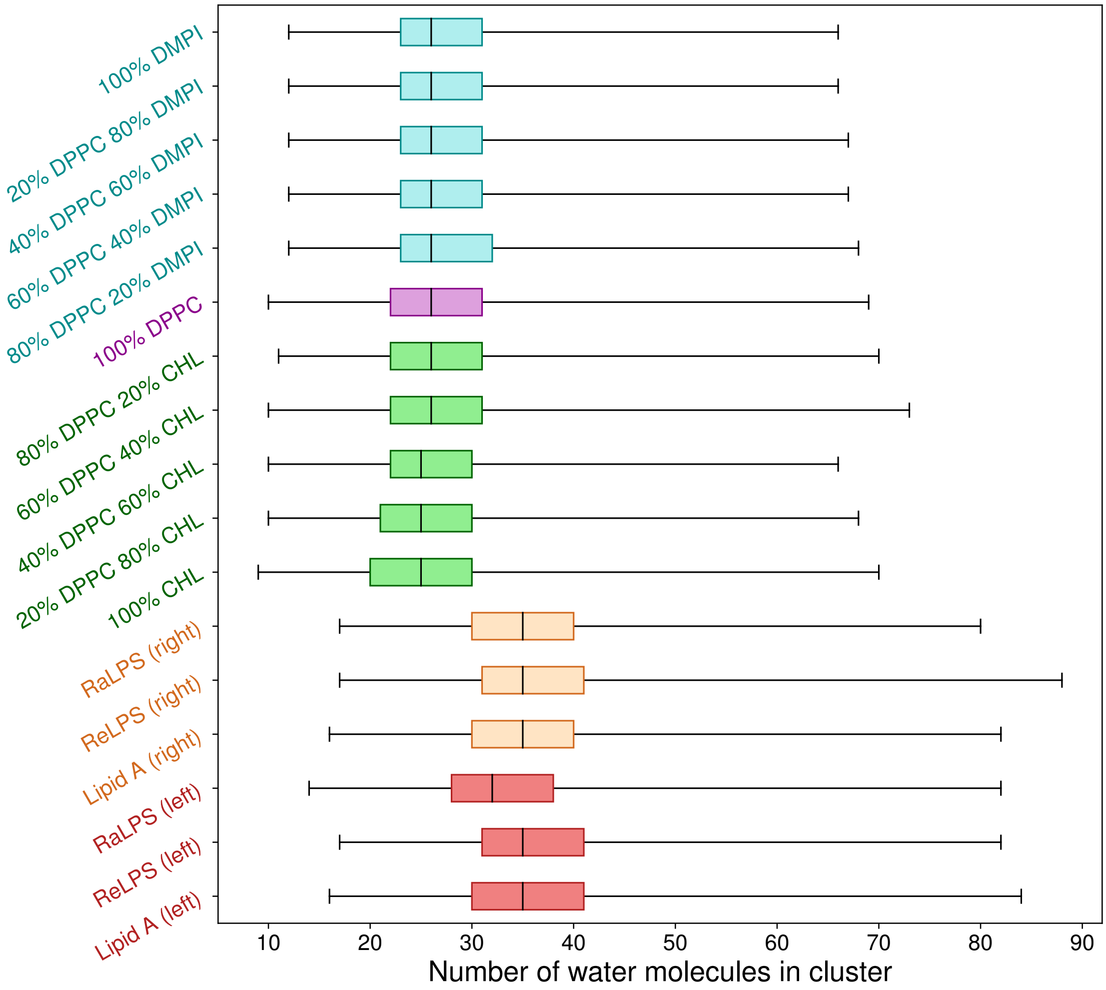


Figure 17: Box plots of the sizes of the largest icy cluster per frame for the various systems. The asymmetric LPS systems are split into “left”: the side containing a mixture of phospholipids, and “right”: the side containing Lipid A with (ReLPS, RaLPS) or without (Lipid A) lipopolysaccharide chains. The boxes extend from the upper to lower quartile, with a line at the median. The whiskers encase the entire range of values, excluding outliers; which are not plotted. Outliers are defined as being values falling above the 99.9th percentile. While the analyses described in sections 7.1.1 and 7.1.2 only consider such clusters which contain at least 30 water molecules, the box plots include all largest clusters.

discussed herein, many of which are more qualitative than quantitative.

From figure 17, it is noted that the inter-quartile ranges and medians are broadly similar for all eleven DPPC-CHL/DMPI systems. They are also similar (but higher) across the six sides of the LPS simulations. In the LPS systems, the slightly lower values for RaLPS can be explained by the fact that the simulation box is smaller than that of the ReLPS and Lipid A systems. The same can likely be said of the slightly lower values observed in the high CHL systems.

7.1.1 Choice of minimum cluster size

Figure 18 shows the difference in probability density, when filtering by a minimum size of 30 or 40 or not filtering, for the DPPC-CHL systems. When using a minimum of 40 molecules per cluster, around 93.5 % of all frames are omitted for the DPPC-CHL systems. Filtering by a minimum of 30, on the other hand, retains around 30 % of frames in those same systems. Filtering by this minimum size of 30 does however counter the emergence of a peak at the water-vacuum interface; caused by smaller icy clusters forming preferentially at this interface, as compared to in the bulk water region.

Gasparotto *et al.*[204] reported that supercooled water molecules, i.e. molecules of water cooled below 0 °C but still in a liquid state, are relatively mobile at water-vacuum interfaces, with the effect that hydrogen bonds are formed and broken more quickly than in bulk water. This could explain the observed phenomenon of clusters forming preferentially at the vacuum interface; while the fleetingness of these hydrogen bonded structures provides an explanation for why these clusters are generally fairly small.

7.1.2 Ice nuclei at the lipid-water interface

The rate of emergence of pre-critical icy clusters at the interface between the bilayers studied herein and the water layer provides a unique insight into the ability of each bilayer to act as an INA. Figures 19 to 21 show the distribution of largest pre-critical icy nuclei for the DPPC-CHL, DPPC-DMPI and LPS systems, respectively. In these graphs, only clusters with at least 30 water molecules are included, as discussed in section 7.1.1.

In the case of DPPC-CHL (see figure 20), some degree of increased ice nucleation was observed for all six systems. It is interesting to note that the pure DPPC bilayer appears to have a similar ice nucleating potential to pure CHL, which is known to be a good INA in both crystalline[21] and monolayer[24] form, despite the fact that the bilayer-water interface is much more diffuse in the case of DPPC than that of CHL. These DPPC bilayers appear to be comparable to CHL monolayers, which are known[24] to facilitate ice nucleation below −15 °C, in their ability to act as INAs; this suggests that while such bilayers can act as INAs, they are weak in comparison to biological INAs such as crystalline CHL,[21] Pseudomonas syringae[205,206] and pollen.[207–210] For the mixed bilayers, containing both DPPC and CHL, the effect on ice nucleation appears to decrease as CHL concentration increases, noted by a less prominent peak in the probability density at the bilayer-water

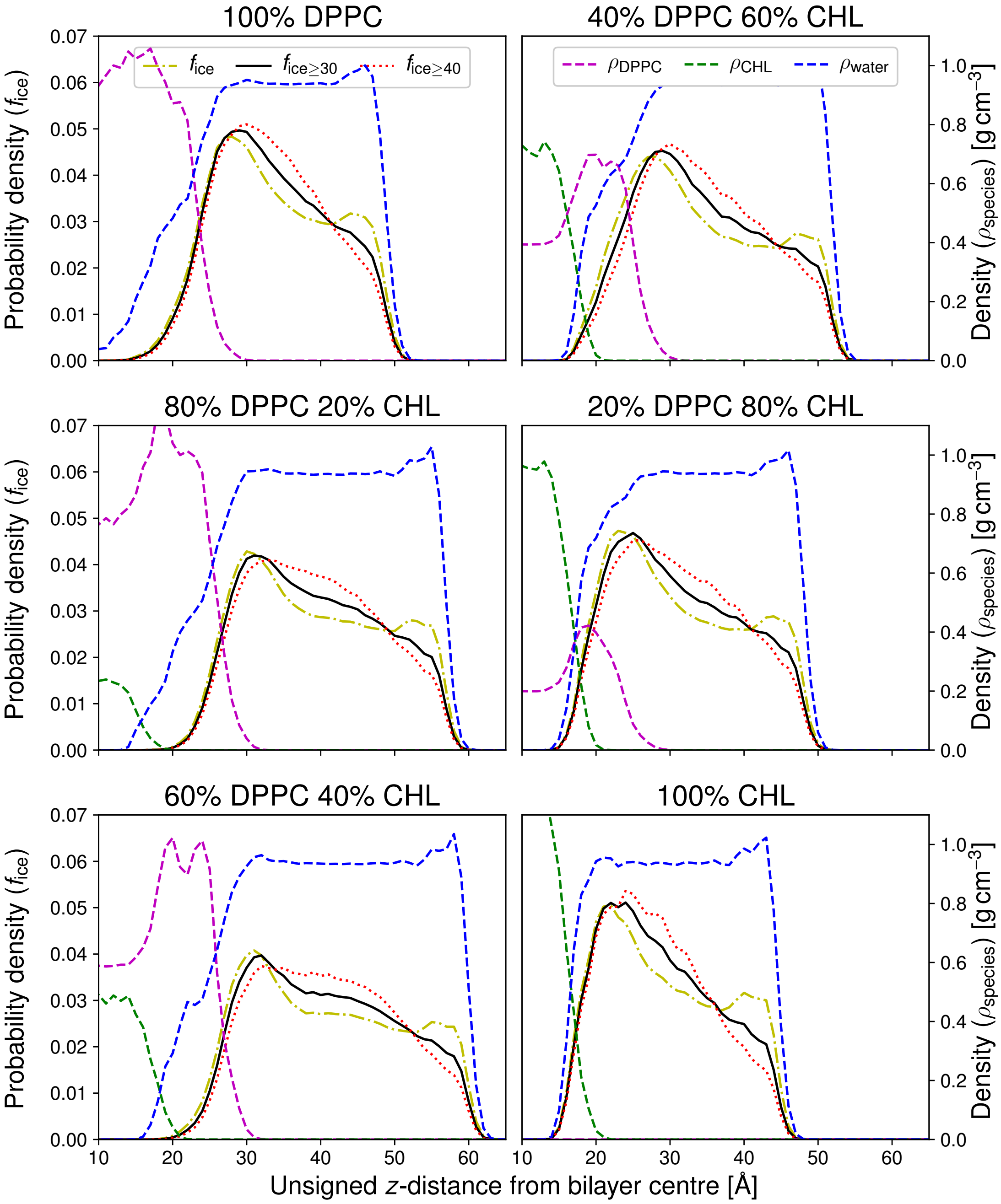


Figure 18: Probability density $f_{\text{ice}}(z)$ (black, solid) for (unsigned) z-distance of water molecules, within a largest icy cluster, from the centre of mass of the bilayer. DPPC, CHL and water densities: $\rho_{\text{DPPC}}(z)$, $\rho_{\text{CHL}}(z)$ and $\rho_{\text{water}}(z)$ are displayed with dashed magenta, green and blue lines, respectively. Corresponding PDFs only including clusters with minimum 30 or 40 molecules are shown by yellow and red dotted lines, respectively.

interface, up until 40 mol% CHL, wherein it increases once more until 100 mol% CHL. This can likely be attributed at least in part to the accessibility of the CHL hydroxyl group for low CHL systems and the fact that any water which is surrounding these groups is under-coordinated, as discussed later in section 7.3; while the hydrogen bonding sites on the DPPC molecules become more tightly packed (see section 7.2) and fewer in number.

Similar trends are observed in the case of the DPPC-DMPI systems (see figure 20). It appears that pure DMPI appears is even more active than pure CHL, showing a pronounced peak in the probability density at the bilayer–water interface. It is also of note that the decrease of the peak with increased DMPI concentration is less pronounced than was seen with the DPPC-CHL systems.

Finally, we consider the LPS systems (see figure 21). These are asymmetric systems, the "left" side of the bilayer features a phospholipid-water interface, whilst the "right" side of the bilayer is composed of lipid A (plus coating sugars in the case of ReLPS and RaLPS). Focusing first on the phospholipid leaflet alone (left side), we observe a small increase of the probability density for the ice nuclei to form within the interfacial region. Moving onto the right side of the membranes, it appears that the Lipid A system (top panel of figure 21) has little, if any, ice nucleating potential, with no noticeable peak in the probability density for icy clusters to form at the bilayer-water interface. However, this is not the case when we look at the ReLPS and RaLPS systems.

The PDF for RaLPS appears intriguing, as we see a substantial number of ice nuclei even within the extended sugar-water interface, where the water density is much lower than in the bulk of the water layer. This is indicative of high potency at nucleating ice of the RaLPS system but not inconsistent with the fact that ice does not appear in under-coordinated water.[*] This is because water is not evenly dispersed within the sugar layer and, while the average water density is around 0.6 g cm^{-3}, this is arranged in pockets of fully coordinated water, see figure 22 for a visualisation of the RaLPS sugar distribution.

We have looked at the emergence of pre-critical icy clusters around the bilayer-water interface, this is a great measure of a surface's tendency to promote ice nucleation. Next, we look at some structural properties of the bilayer itself. In a previous study,[24] it was shown that there is a directly proportional relationship between the degree of structural order within CHL monolayers and their ability to promote ice nucleation. This is not directly comparable with the result herein, as the systems differ in both structure and chemistry, however the fact that better ordered systems may be better at promoting ice nucleation, with all other factors equal, should provide an interesting insight to the relationship between pre-critical cluster emergence and structural order.

---

[*] As described in section 1.1, water molecules within ice crystals must necessarily be surrounded by four other water molecules. Similarly, the procedure described in section 5.1 for computing the icy clusters will be strongly affected by the coordination of the water molecules.

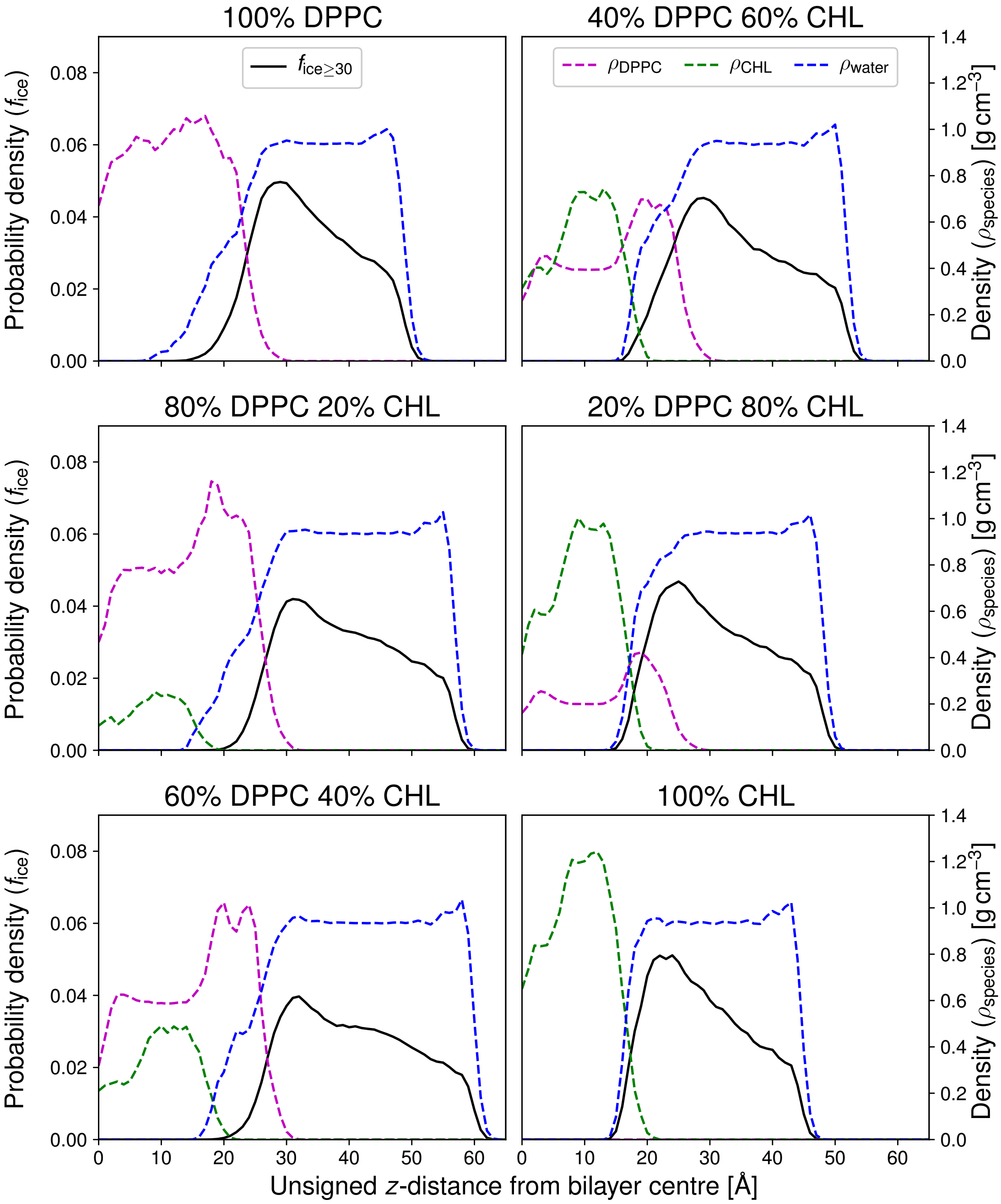


Figure 19: Probability density $f_{\mathrm{ice}}(z)$ (black, solid) for (unsigned) z-distance of water molecules, within a largest icy cluster, from the centre of mass of the bilayer. DPPC, CHL and water densities: $\rho_{\mathrm{DPPC}}(z)$, $\rho_{\mathrm{CHL}}(z)$ and $\rho_{\mathrm{water}}(z)$ are displayed with dashed magenta, green and blue lines, respectively.

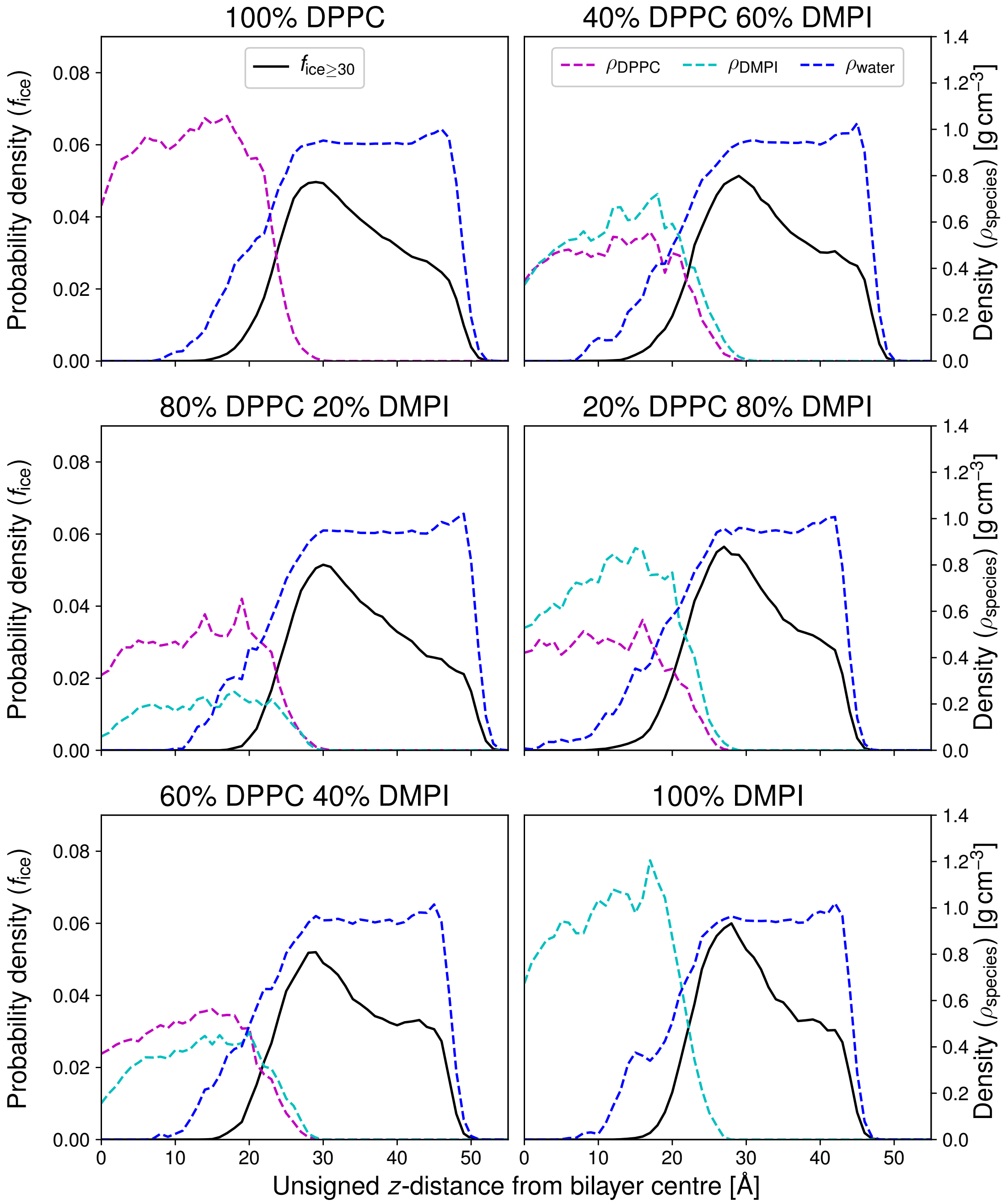


Figure 20: Probability density $f_{\text{ice}}(z)$ (black, solid) for (unsigned) z-distance of water molecules, within a largest icy cluster, from the centre of mass of the bilayer. DPPC, DMPI and water densities: $\rho_{\text{DPPC}}(z)$, $\rho_{\text{DMPI}}(z)$ and $\rho_{\text{water}}(z)$ are displayed with dashed magenta, cyan and blue lines, respectively.

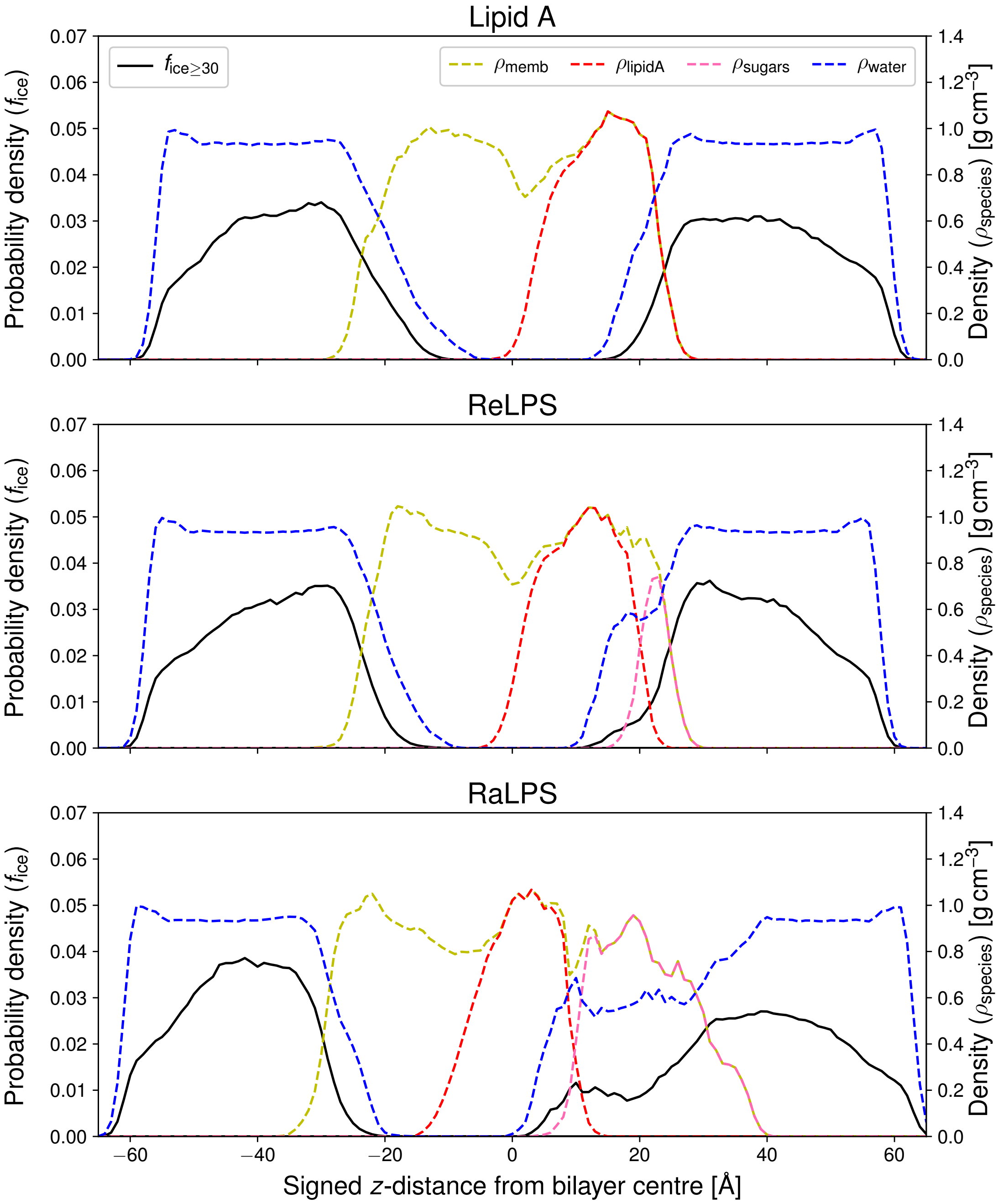


Figure 21: Probability density $f_{\mathrm{ice}}(z)$ (black, solid) for (signed) z-distance of water molecules, within a largest icy cluster, from the centre of mass of the bilayer (including the sugars). Total membrane, lipid A, sugar and water densities: $\rho_{\mathrm{memb}}(z)$, $\rho_{\mathrm{lipidA}}(z)$, $\rho_{\mathrm{sugars}}(z)$ and $\rho_{\mathrm{water}}(z)$ are displayed with dashed red, yellow, pink and blue lines, respectively.

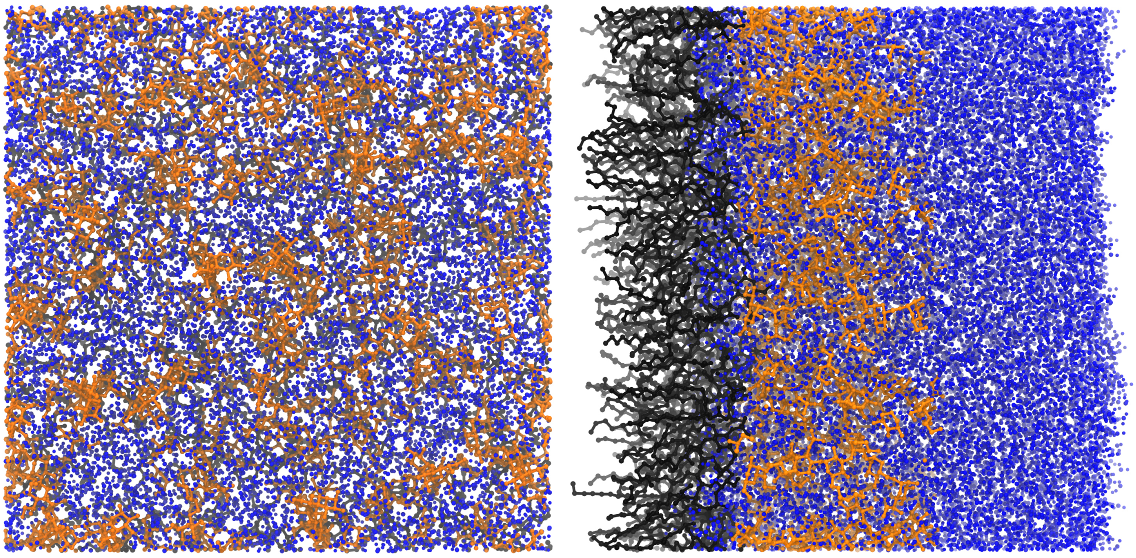


Figure 22: Distribution of sugars in the RaLPS system. Left panel shows xy-plane cross section. Right panel shows the system across the z-axis. Lipid A is coloured black, sugars are coloured orange, water molecules are coloured blue. Hydrogen atoms are not drawn for Lipid A and sugars.

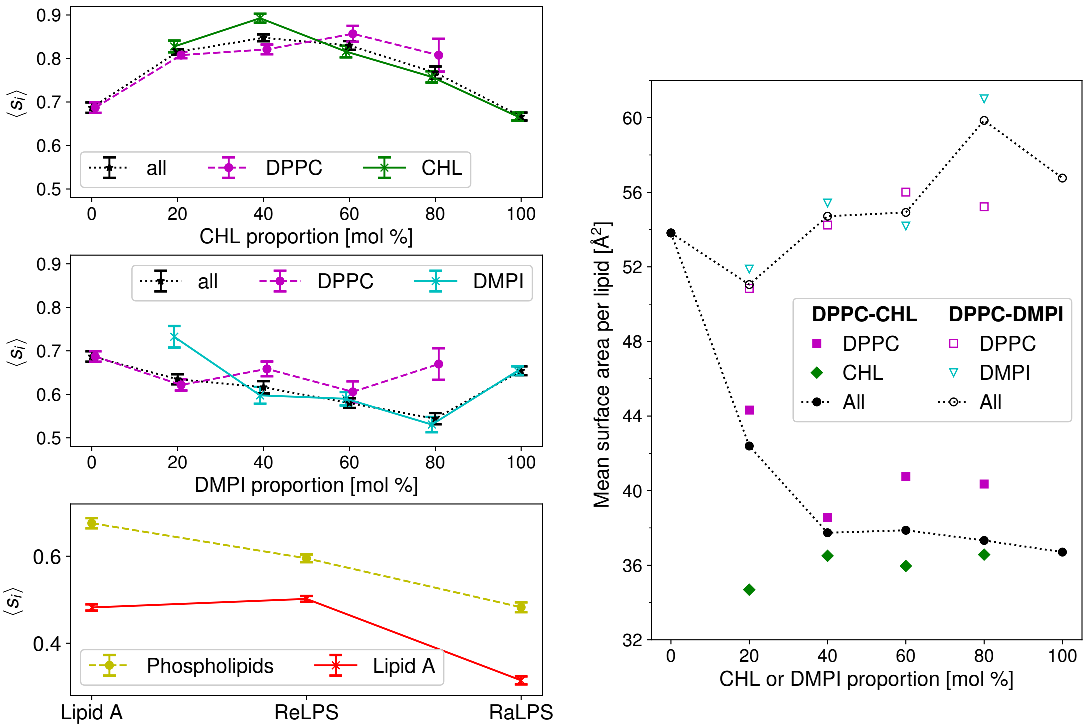
7.2 Structural order within the bilayers

To explain the trends observed in the simulations, it is vital to have an understanding of the structural properties of the different bilayers and how the lipids interact with each other. In section 5.3.1 the SMAC parameter was defined, this provides a measure of order within the bilayers, taking into account both local density and orientation. Higher values of SMAC correspond with a greater degree of order between the lipid tails. The average SMAC values for each system are given in figure 23(a). For the DPPC-CHL/DMPI systems, the average SMAC values are also reported for each lipid type; while for the LPS systems, the SMAC values for the two leaflets are reported separately.

Adding CHL into DPPC bilayers we see a progressive increase in structural order up to 40 mol% CHL, with both DPPC and CHL lipids appearing the most ordered at around this concentration. The ordering and condensing effect that CHL has on DPPC bilayers is well known[25,30,133] and here appears to decline at concentrations higher than 40 mol% CHL.* Remarkably, the pure DPPC and pure CHL bilayers report a very similar degree of structural order via the SMAC parameter, a value lower than that of any of the mixed systems. It is interesting to note that phospholipid bilayers found in nature typically have CHL content ranging from 0–50 mol%.[34,35] Comparing these trends with those from figure 19, the most ordered of the bilayers appear to be those with the least ice nucleating potential. This appears surprising as previous studies[24] on CHL monolayers found that structural order was in directly proportional to the ice nucleating potential. One should

---

*There is not much literature relating to bilayers with greater than 50 mol% CHL owing to their non-existence in nature. For the same reason however, it seems unsurprising to see the structural ordering decreasing above these concentrations.



(a) Average SMAC parameter

(b) Voronoi area per lipid

Figure 23: (a) Average SMAC parameter $\langle s_i \rangle$ for lipid tails in different systems. The top panel shows SMAC for DPPC-CHL bilayers, with mol% CHL along the horizontal axis, the middle panel shows SMAC for DPPC-DMPI bilayers, with mol% DMPI along the horizontal axis, while the bottom panel shows SMAC for the three LPS bilayer systems. The error bars show one standard deviation. The points are slightly offset along the x-axis for improved readability. The y-scale is identical for DPPC-CHL and DPPC-DMPI for purposes of comparison, the scale is different for LPS due to the greatly lower degree of order in the lipid A tails. (b) Average (Voronoi) surface area per lipid. DPPC-CHL systems have solid points, DPPC-DMPI have unfilled points.

note this apparent contradiction, but also note that the systems here are affected by chemical factors as well as structural factors, as these are heterogeneous bilayers.

In DPPC-CHL bilayers, we saw mixed bilayers which were better ordered than the pure bilayers. The opposite is true in DPPC-DMPI bilayers — here the systems with the greatest degree of structural order are the pure DPPC and pure DMPI bilayers, while the mixed systems are less ordered. Adding DMPI into DPPC bilayers induces a degree of structural disorder, largely due to disorder between the DMPI tails, interestingly the DPPC tails retain a similar degree of order across all systems. In comparing DPPC-CHL to DPPC-DMPI we should recall that CHL is a much smaller lipid than DPPC, while DMPI has a slightly larger headgroup than DPPC, but is of a much more similar size. Considering the fact that the trends in figures 19 and 20 are similar, this poses interesting questions about the relevance of structural order to a system's ability to act as an INA. In section 7.3, the inter-molecular hydrogen bonds are considered, which will help to explain these trends.

The final set of systems to compare the SMAC parameter for are the LPS systems. Here the addition of sugars from Lipid A, with no sugar, to ReLPS, with two KDO units linked to each lipid A residue, appears to result in very little, if any increase in the SMAC parameter, and thus the degree of structural order within the bilayer. The story is rather different when we look at the RaLPS system, with the complete oligosaccharide core on each lipid A residue, this system appears to display a much lower degree of structural order amongst the (four major) lipid A tails. In all three systems, the lipid A tails show rather lower values of SMAC than any component within the DPPC-CHL/DMPI systems, however we should remember that SMAC measures the density of tails as well as the orientational symmetry.

More surprising is the trend seen in the SMAC parameters for the phospholipid leaflets; here it appear that the phospholipids become less well-ordered as we add sugar to the lipid A leaflet. This can be partially attributed to the difference in density between the three systems.[*] Overall, the LPS systems do seem to be less well ordered than the DPPC-CHL/DMPI systems, which is perhaps unsurprising due to the far greater level of complexity in these systems. We saw in figure 21 that the Lipid A system was not a particularly good INA, however the ReLPS and RaLPS systems did display a good amount of ice nucleating potential. Again, this seems to contradict previous results that showed better ordered monolayers to be better INAs than less well ordered ones, however the differences can be attributed to the "rough" structure of the sugar coated lipid A leaflets — causing more accessibility to hydrogen bonding sites, as discussed in the next section.

The second measure of bilayer order discussed is the accessible area per lipid. This was computed using a Voronoi tessellation, as described in section 5.3.2. Figure 23(b) shows these Voronoi accessible areas per lipid for the DPPC-CHL/DMPI systems. These values are higher for the DPPC-DMPI systems than for their DPPC-CHL counterparts — unsurprising as DMPI has a slightly larger headgroup than DPPC, while CHL is a substantially smaller lipid. In the case of DPPC-CHL, the trends in Voronoi area are in good agreement with the SMAC results, with DPPC taking up the least space in the 40 mol% CHL system. Interestingly, CHL accesses a similar amount of space in all systems, with the exception of the 20 mol% CHL system, where the CHL molecules average area per lipid is slightly lower, this can be attributed to the tight packing between DPPC and CHL lipids, similarly to as described by Leeb and Maibaum[133].

In the DPPC-DMPI systems, the trends are broadly linear, with area per lipid increasing with DMPI concentration. This is what one would expect given the relative sizes of the two lipid headgroups and appears to suggest, as did the SMAC results, that the two lipids interact with each other in a fairly similar manner to which they interact with other lipids of their same type. Voronoi areas were not computed for the LPS systems

---

[*] Note that, by equation (27), s_i should scale with density by a factor of $(1-\psi(n_i))$, as the right hand side of equation (27) is the mean of $f(r_{ij})K(\theta_{ij})$, which is unaffected by density. Thus, by equation (25), the SMAC parameter increases with density if all other factors are unchanged.

as the difference between these systems is in the z-direction rather than in the xy-plane. The average area per lipid A residue can however be calculated by simply dividing the cross-sectional area by the number of lipid A molecules. This gives average area per lipid for the Lipid A, ReLPS and RaLPS systems to be 157.6, 169.7 and 180.2 Å^2, respectively.

Having looked at both the icy clusters and the degree of structural order between the different bilayers, a number of apparent contradictions have arisen. The next section, regarding the hydrogen bonds which form between the bilayers and the icy clusters will go some way to reconciling these differences and further elucidating the interplay between structural and chemical factors in what makes a lipid bilayer good or bad at promoting ice nucleation.

7.3 Hydrogen bonds

Figure 24 shows the average number of hydrogen bonds between each bilayer and the largest icy cluster per frame, as a function of the bilayer composition. For the DPPC-CHL/DMPI systems, the hydrogen bonds are categorised by bonding site (see figure 3) and whether the bond is donated or accepted by the bilayer. For the LPS systems, the hydrogen bonds are categorised with reference to each residue instead, as there are far too many potential bonding sites to illustrate them all separately.

Looking first at the DPPC-CHL systems, the number of hydrogen bonds between the largest icy clusters and the CHL molecules is roughly proportional to the CHL concentration. However, for systems below 40 mol% CHL, we only see a negligible number of such hydrogen bonds; this can be explained by the inaccessibility of the CHL hydroxyl group to the water — in such systems, the CHL molecules tend to sit close to the centre, with the DPPC headgroups extending much further out. Clearly illustrating this point is figure 19, where in the 20 and 40 mol% CHL systems there is very little overlap, along the z-direction, between CHL density and the location of the icy clusters. This is due to the small degree of overlap between CHL and water density.

It is less easy to quantify the trends for hydrogen bonds between DPPC and the icy clusters, as these appear to fluctuate rather than showing any clear correlation to the DPPC concentration. DPPC has eight hydrogen bonding sites (see figure 3(a)) with five of these being located especially deep within the bilayer-water interface (the four “tailgroup oxygens” and the innermost phosphate oxygen). As the DPPC concentration decreases we see an increase in hydrogen bonds forming at these five sites, due to an increase in water density compared to the higher DPPC systems; this trend peaks at 60 mol%, after which the reduction in total bonding sites leads to a reduction in hydrogen bonds. While those five sites show the clearest trend, similar trends are observed for the other bonding sites, with none of the sites being as accessible as the CHL hydroxy group in pure CHL environments such as monolayers, crystals or the pure CHL bilayer system studied here.

In contrast, the DPPC-DMPI systems present a very different picture. Owing to the fact that these two lipids are far more alike than DPPC and CHL, the trends appear

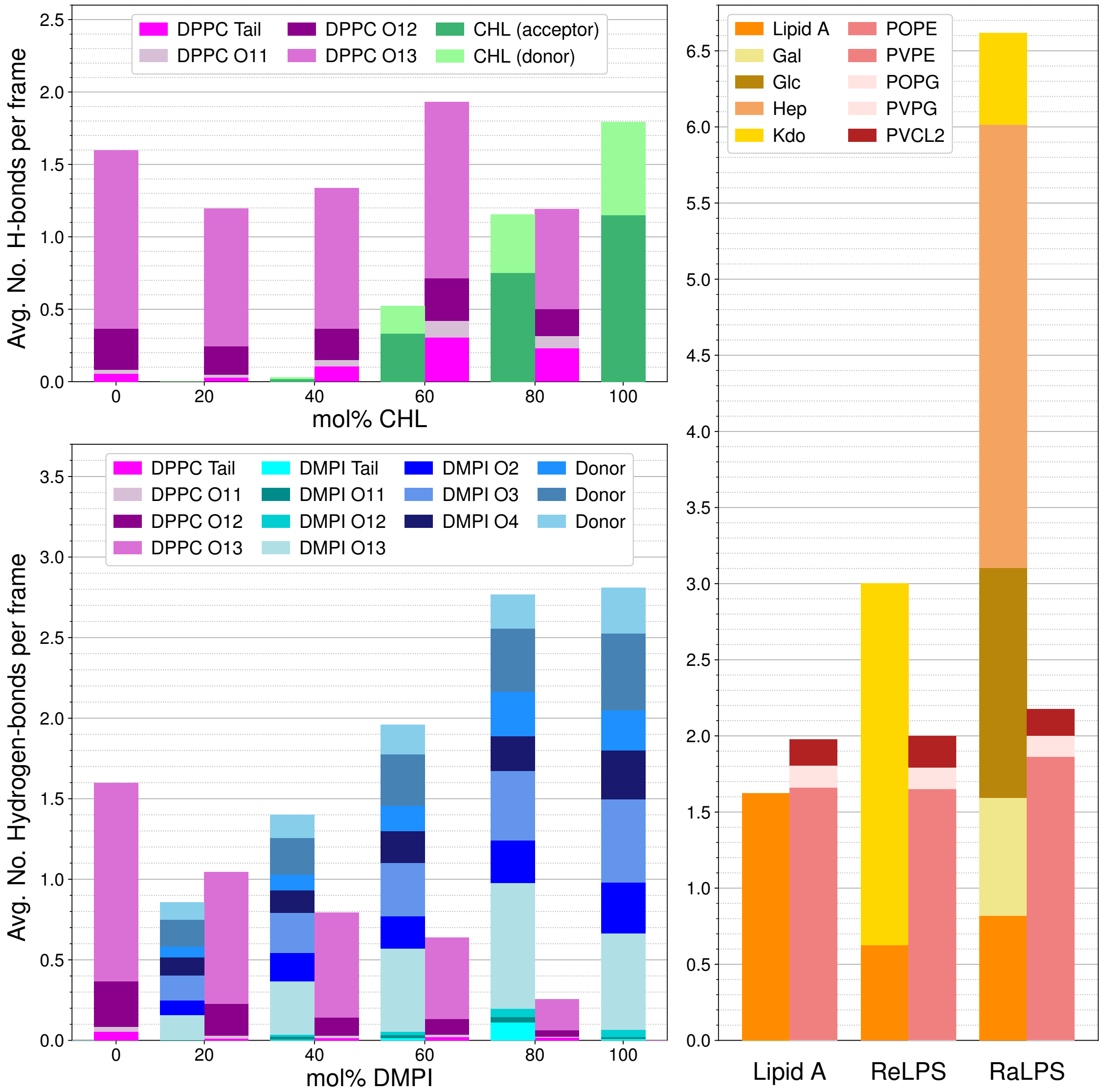


Figure 24: Average number of hydrogen bonds per frame from the largest icy cluster to DPPC-CHL (top left), DPPC-DMPI (bottom left) and LPS (right) bilayers. For DPPC, CHL and DMPI, the different possible bonding sites are coloured in the same way as in figure 3. DMPI has four bonding sites (O2, O3 and O4) which can accept or donate hydrogen bonds; the third and fourth columns in the key correspond to those three sites, with acceptors on the left and donors on the right. For the LPS systems, bonding sites are grouped by molecule and not distinguished between donors and acceptors.

much more straightforward. For DPPC the number of hydrogen bonds appears to be approximately linearly proportional to the concentration of DPPC. For DMPI, the same seems to be true up until 80 mol% DMPI, after which the number of hydrogen bonds appears to plateau. It seems reasonable to suggest, considering the structure of DMPI (see figure 3(c)), that interfacial icy clusters are already fully “saturated” with available hydrogen bonding sites in the 80 mol% DMPI bilayer. Interestingly, this 80 mol% DMPI system has the most hydrogen bonds when we include those with both lipids, perhaps this can be accounted for by the “roughness” due to the lower degree of structural order in this system, as compared to the 60 and 100 mol% DMPI systems (see figure 23).

It is evident also that DMPI, on average, forms a greater number of hydrogen bonds with the icy clusters than DPPC. This can be attributed to the greater number and accessibility of hydrogen bonding sites in the PI headgroup as compared to PC — thirteen, five of which can both accept and donate hydrogen bonds, as compared to eight for DPPC, all of which may only accept bonds. Interestingly, these results appear to be consistent with the slightly higher ice nucleating effect of pure DMPI bilayers in comparison to pure DPPC bilayers (see figures 19 and 20). In comparing these trends it is important to recall that, conversely to what is seen in DPPC-CHL bilayers, mixed DPPC-DMPI bilayers are less well ordered than their pure counterparts (see figure 23). This makes it all the more striking that we see such ice nucleating potential from the mixed DPPC-DMPI systems; likely in part due to the higher potential DMPI has to form hydrogen bonds with water molecules and the order induced by such hydrogen bonding having an effect on ice nucleation.

The final set of systems to consider is the LPS systems. Here the number of hydrogen bonds between the phospholipid leaflet and the icy clusters at the phospholipid-water interface are fairly consistent across all three systems. It would be surprising to see any specific trends in this case as the only difference between those leaflets are minor. Comparing the hydrogen bonds formed with the three lipid A leaflets, strong trends are observed. In figure 21, we saw that lipid A alone does not appear to facilitate ice nucleation very much if at all, this appears to correspond with the bar for the Lipid A system in figure 24. In contrast, the ReLPS and RaLPS systems do form a large number of hydrogen bonds with the icy clusters, especially with the most accessible of the sugar residues.

In RaLPS, one might be surprised to see so many bonds forming between lipid A and the icy clusters, and surprised to see the density of icy clusters where the water density is low. Certainly one shouldn’t see many icy clusters in under-coordinated water; however the nature of the systems is that, while the average water density is around $0.6\,\mathrm{g\,cm^{-3}}$, the water is assembled in pockets of fully coordinated regions. This means that the effective surface area of the RaLPS system, and to a lesser extent the ReLPS system, is far greater than the cross-sectional area, hence the ability for clusters to form so many hydrogen bonds. In other words, due to the sugar residues extending well into the water region and coming into contact with pockets of fully coordinated water, there are a vastly greater number of possible sites where hydrogen bonds might form between icy clusters and the bilayer.

7.4 Water geometry

7.4.1 Water orientation

It is widely accepted that the structure of liquid water can be greatly affected by contact with hydrophilic and hydrophobic surfaces.[211] Such effects have been posed as a leading factor with regard to the ability of different surfaces to nucleate ice.[212,213] One way to look at the hydrophilic or hydrophobic properties of surfaces is by looking at the distribution of orientations of water molecules, as a function of their distance from the surface. In section 4.4, we defined two different orientational order parameters: θ, defined as the angle between the dipole moment of a water molecule and the bilayer normal, and ϑ, defined as the cosine of θ; with the point of ϑ being that the probability density function for ideal water is a constant 1/2.

In figure 25, the values of ϑ are plotted for a bulk-like region of the pure DPPC system. Here, we see that the distribution of probability density for the bulk-like water is close to $\frac{1}{2}$, as expected for ideal water. This has been plotted as a colourmap over z in order to demonstrate that the distribution of ϑ values is invariant across this region. Thus this water region is indeed bulk-like. We do see a number of horizontal lines in this plot, these can be explained as artifacts of the binning procedure for histogram production, note that they are broadly constant across the z range. The distribution of θ values, as a function of the z-distance from the centre of the bilayer, are given as colour maps in figures 27 and 29 and similar plots for ϑ are given in figures 28 and 30, with water density also shown on

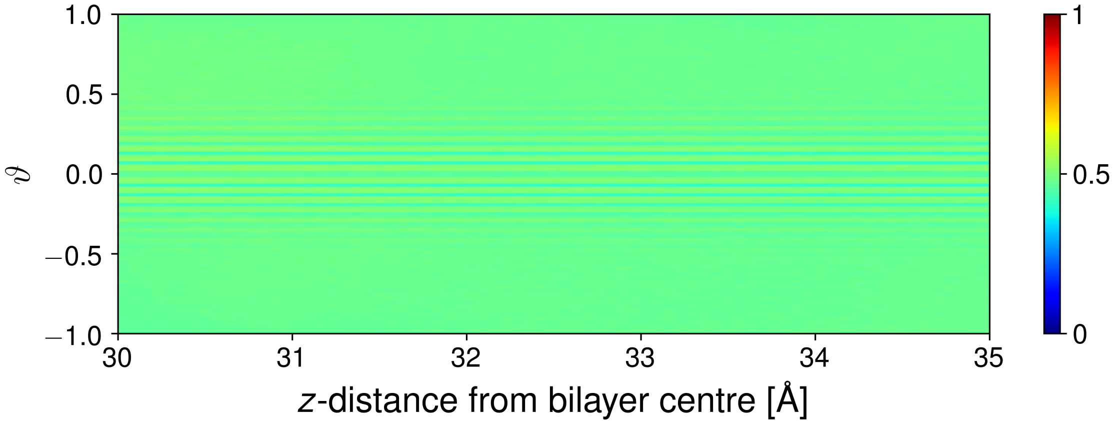


Figure 25: Colourmap of normalised orientational order parameter $\vartheta = \cos(\theta)$ distribution for a bulk-like region of water, taken from the pure DPPC system. The scale of the colour bar is the probability density with respect to ϑ. Since the range of ϑ is $[-1, 1]$, a uniform distribution will produce the probability density value of 1/2. This has been plotted as a colourmap over z in order to demonstrate that the distribution of ϑ values is invariant across this region.

each graph for reference.

In figure 27, we see a trend across the DPPC-CHL systems where θ values below 90° are seen near the interface with the high DPPC systems, with a slight inclination for θ values in the same region to rise above 90° when we have a higher proportion of CHL. This trend is made all the more clear in figure 28 when we look at the ϑ distribution; again here we see well defined peaks around $\vartheta = 1$ for the high DPPC systems, with a corresponding dip in the probability density around $\vartheta = 1$ in the case of the high CHL systems. This trend can be explained by the topology of the headgroups of DPPC and CHL. In the case of DPPC, there are a number of hydrogen bonding sites, all of which can only accept hydrogen bonds, this could explain why it is that we see a greater number of water molecules with their dipole moments pointing away from the bilayer. For a visualisation of this, see figure 26 which shows a snapshot of water molecules at the surface of a DPPC bilayer. While this is a case of averages, it is apparent from the figure that more water molecules have their dipoles pointing away from the bilayer than towards it. Conversely, CHL has one bonding

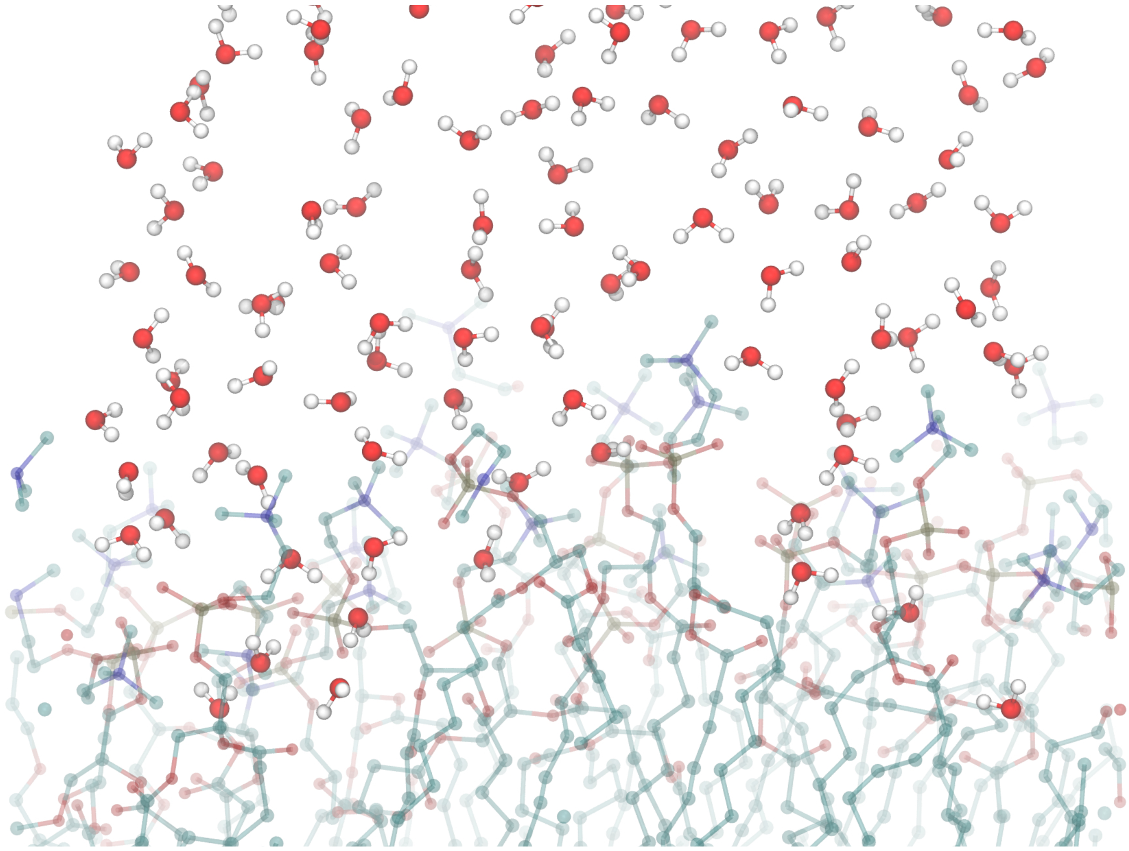


Figure 26: A snapshot of water molecules at the surface of a DPPC bilayer. The horizontal axis is the xy-plane, while the vertical axis is the z-axis. Water molecules are displayed solidly with their oxygen atoms coloured red and their hydrogens coloured white. DPPC molecules are transparent with carbon, oxygen, phosphorus and nitrogen atoms coloured in cyan, red, yellow and blue, respectively. Hydrogen atoms are not displayed for the DPPC molecules. On average, more water molecules near the interface have their dipoles pointing away from the bilayer than towards it.

site which can either donate or accept hydrogen bonds. The fact that we see a slight preference for water molecules' dipole moments to point into the bilayer might initially seem at odds with figure 24 where we see around twice as many water molecules donating hydrogen bonds to the CHL hydroxy group than accepting them. However, when one examines the area in which we see water molecules' dipole moments pointing into the bilayer, these are in fact water molecules which are deep within the headgroups, therefore many of these molecules may have their dipole moments pointed away from CHL hydroxy groups.

Turning our attention now to the right hand panels of figures 27 and 28, we observe a much weaker trend for the DPPC-DMPI systems. This is somewhat expected as both the structure and the chemistry of DPPC and DMPI are much more alike than that of DPPC and CHL. In all six systems we see an increased concentration of θ below 90° around the diffusion layer, however this is less pronounced for the high DMPI than it is for the high DPPC systems. Again, this difference can be explained by the geometry of hydrogen bonding sites, as DMPI has an increased number in comparison to DPPC, however the additional sites can act as both acceptors and donors, once again implying that we would not expect to see such a strong preference for the water dipoles to point outwards from the bilayer. When we look at the distribution of ϑ for DPPC-DMPI, as shown in the right hand panel of figure 28, we can see that as stated with θ and explained by the hydrogen bonding sites, DMPI has less of an orienting effect on water than DPPC, although water molecules in the diffusion layer are more likely to be oriented with their dipole pointing away from the bilayer. We can also see more clearly than in figure 27 that the orienting effect of DPPC is more substantially retained at lower DPPC concentrations than in the case of the DPPC-CHL systems. This can be explained by the fact that DPPC and DMPI are more similar to one another in both chemistry and size than DPPC and CHL are.

Finally, we discuss figures 29 and 30. For the diffusion layer between water and the phospholipid leaflet, there is a slight tendency in θ to be below 90° as we saw with DPPC and DMPI. For the Lipid A leaflet, we see a similar trend, which is unsurprising due to the chemical similarities between the headgroup of lipid A molecules and the PC, PI, PE, PG and CL phospholipid headgroups. Looking at ϑ, the information can be more easily seen. For the Lipid A system, we see a large peak around $\vartheta = -1$, corresponding to a concentration of water molecules with their dipoles pointing into the bilayer at the edge of the bulk water. This is interesting as it is the opposite trend to which we saw for phospholipid bilayers and cannot be explained by hydrogen bonding, as lipid A, like the phospholipids is more likely to accept hydrogen bonds than to donate them, thus some other form of inter-molecular interaction must be responsible for this phenomenon. In the case of the ReLPS and RaLPS, we once again see a higher number of interfacial water molecules oriented with their dipoles pointing outwards and the distribution of water molecules upon the phospholipid leaflet is rather similar to the distribution we saw for DMPI, which is unsurprising.

Overall, analysis of the normalised parameter ϑ has been remarkably successful in comparison to θ, with details impossible to see in the θ plots becoming very pronounced when considering ϑ. One might initially be skeptical of the drastic increase in the clarity of the peaks, but when considering the shape of the sinusoidal wave that $y = \cos(x)$ produces from 0 to 180° this is less surprising. It is also important to note that there is geometric legitimacy to this normalisation, as explained in section 4.4.2, and thus these more pronounced peaks are also credible.

7.4.2 Rings and cages

The number of 3–9 membered rings within each icy cluster was computed for the DPPC-CHL bilayers, in particular, the number of six-membered rings is of interest as these are the basic building blocks of both cubic and hexagonal ice. Extraction of detail from these results is not a simple task. For each frame, we have the number of six-membered rings within the largest icy cluster which were counted to give a probability distribution Ω_N, for $N = 1, 2, \ldots$ defined as such:

$$\Omega_N = \frac{\text{Number of frames with } N \text{ six-membered rings}}{\text{Total number of frames}}. \tag{40}$$

Note that N starts at 1 and not 0. This is because in practice the R.I.N.G.S. code always finds at least one six-membered ring within the largest icy clusters. Plotting the values of Ω_N gives an insight as to how one might fit the data. Specifically, when we plot using a logarithmic scale, the distribution looks linear, suggesting that one might be able to fit a line:

$$\ln\big(\mathbb{P}[n_R = N]\big) = mN + c, \tag{41}$$

where ln indicates the natural logarithm, $\mathbb{P}[n_R = N]$ is the expected probability of having N six-membered rings and m and c are constants. This can be written equivalently as:

$$\mathbb{P}[n_R = N] = Ae^{mN}, \tag{42}$$

where A is the constant $A = e^c$. One possible method for fitting the line is using a least-squares regression. However, that method leads to values for m and c which do not satisfy the probability axiom of unit measure. In other words, we must require that:

$$\sum_{N=1}^{\infty} \mathbb{P}[n_R = N] = 1. \tag{43}$$

Well,

$$\sum_{N=1}^{\infty} \mathbb{P}[n_R = N] = \sum_{N=1}^{\infty} Ae^{mN} = A\sum_{N=1}^{\infty} e^{mN}, \tag{44}$$

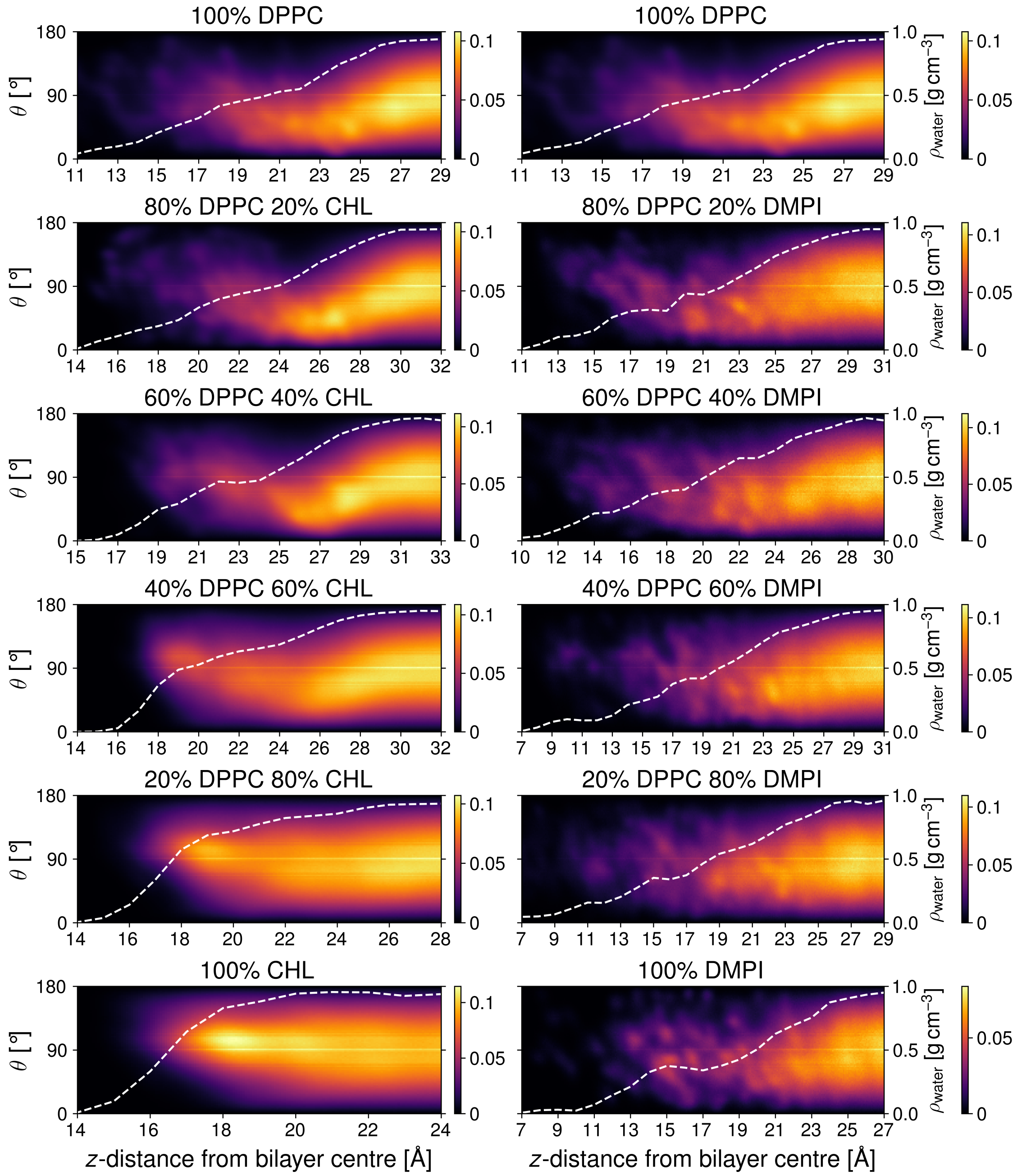


Figure 27: Orientational order parameter (θ) colour maps for DPPC-CHL/DMPI systems (see section 4.4). Average water density is shown as a dashed white line. The scale for the colour map is count per frame per Å^3 per 180°. A value of $\theta = 0°$ corresponds to the dipole moment pointing away from the plane of the bilayer (along the normal) while a value of $\theta = 180°$ corresponds to the dipole moment pointing towards the plane of the bilayer.

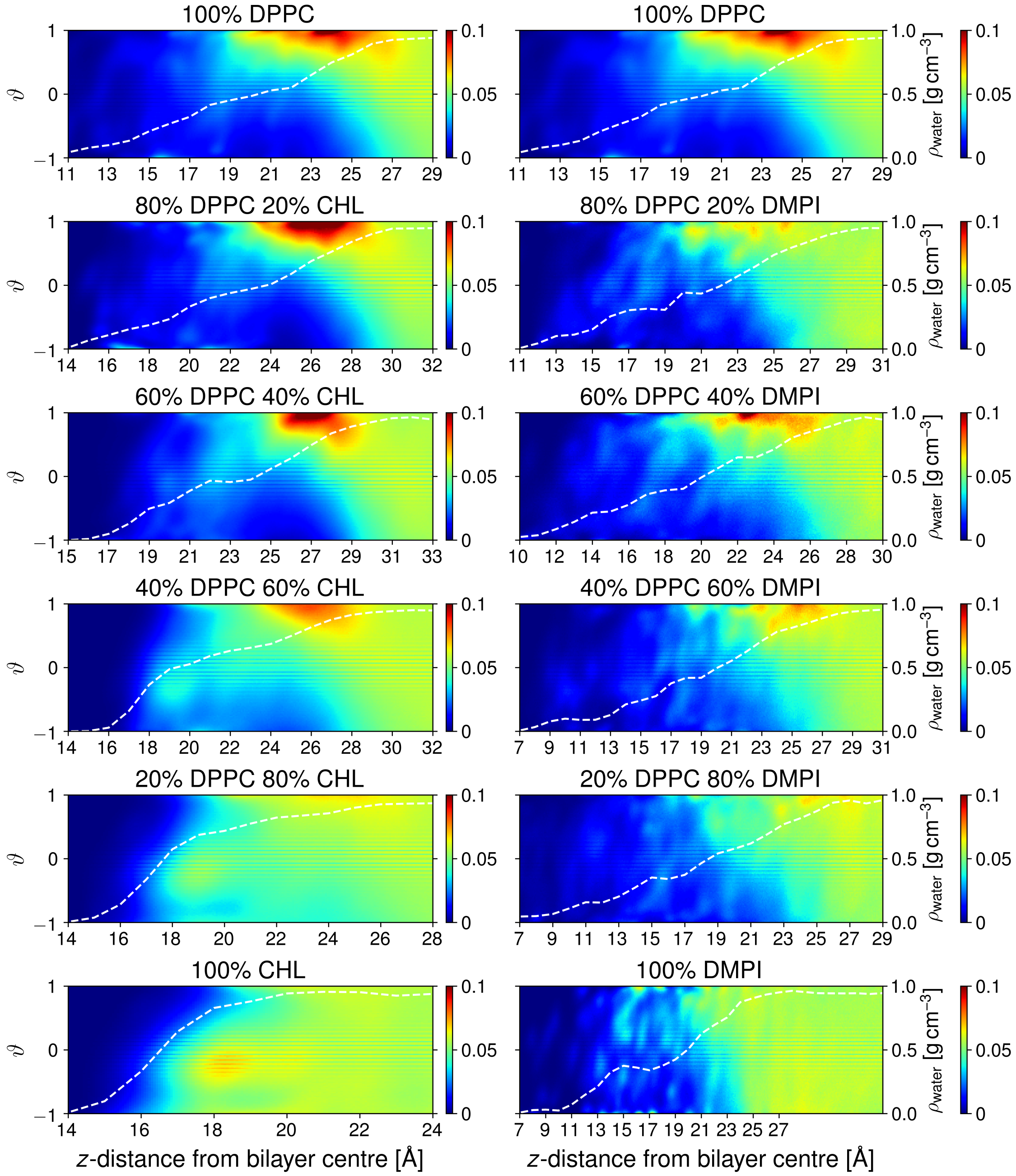


Figure 28: Normalised orientational order parameter ($\vartheta = \cos(\theta)$) colour maps for DPPC-CHL/DMPI systems (see section 4.4). Average water density is shown as a dashed white line. The scale for the colour map is count per frame per Å^3. A value of $\vartheta = 1$ corresponds to the dipole moment pointing away from the plane of the bilayer (along the normal) while a value of $\vartheta = 0$ corresponds to the dipole moment pointing towards the plane of the bilayer.

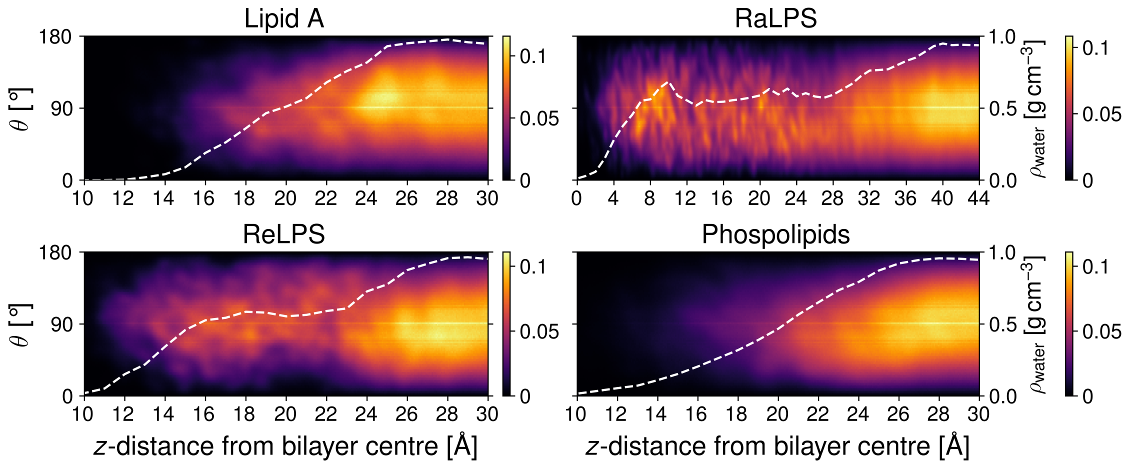


Figure 29: Orientational order parameter (θ) colour maps for LPS systems (see section 4.4). Average water density is shown as a dashed white line. The scale for the colour map is count per frame per Å^3 per 180°. A value of $\theta = 0°$ corresponds to the dipole moment pointing away from the plane of the bilayer (along the normal) while a value of $\theta = 180°$ corresponds to the dipole moment pointing towards the plane of the bilayer.

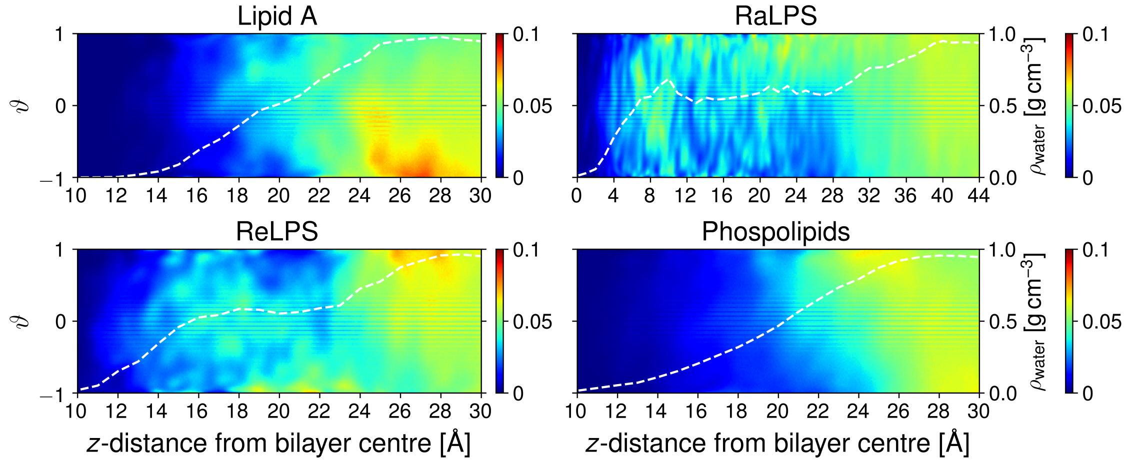


Figure 30: Normalised orientational order parameter ($\vartheta = \cos(\theta)$) colour maps for LPS systems (see section 4.4). Average water density is shown as a dashed white line. The scale for the colour map is count per frame per Å^3. A value of $\vartheta = 1$ corresponds to the dipole moment pointing away from the plane of the bilayer (along the normal) while a value of $\vartheta = 0$ corresponds to the dipole moment pointing towards the plane of the bilayer.

which is a geometric series, since $m < 0$, thus $0 < e^m < 1$. And therefore

$$\sum_{N=1}^{\infty} \mathbb{P}[n_R = N] = \frac{Ae^m}{1 - e^m} = 1. \tag{45}$$

This can be rearranged to discover that

$$A = e^{-m} - 1. \tag{46}$$

Therefore there is only one unknown to choose to fit the curve, i.e.

$$\mathbb{P}[n_R = N] = (e^{-m} - 1)e^{mN}. \tag{47}$$

A reasonable choice for m might be the one so that $\mathbb{P}[n_R = 1] = \Omega_1$, then

$$\Omega_1 = \mathbb{P}[n_R = 1] = (e^{-m} - 1)e^m = 1 - e^m, \tag{48}$$

thus $e^m = 1 - \Omega_1$, and so

$$m = \ln(1 - \Omega_1), \tag{49}$$

and

$$A = e^{-m} - 1 = e^{-\ln(1-\Omega_1)} - 1 = \frac{1}{1 - \Omega_1} - 1 = \frac{\Omega_1}{1 - \Omega_1}. \tag{50}$$

We know that $\Omega_1 < 1$, therefore this logarithm is definitely defined.

With the exception of the 20 mol% CHL system, this model agrees with the data from the DPPC-CHL systems fairly well for smaller values of N. For larger values ($N \sim 10$) equations (42), (49) and (50) give an overestimate for the probability for approximately the range $10 \geq N \geq 20$ and an underestimate for greater values of N. However, the fit is fairly good and the nature of the logarithmic scale exaggerates the discrepancy. We can verify how well these curves fit by computing the squared sum of residuals (SSR), i.e. the sum of the squares of the differences between the actual data values and the curve:

$$\text{SSR} = \sum_{N=1}^{\infty} (Ae^{mN} - \Omega_N)^2. \tag{51}$$

The computed values of A and m for the six DPPC-CHL systems and the SSRs are given in table 4 and the values of Ω_N, along with the fitted curves $\mathbb{P}[n_R = N] = Ae^{mN}$ are displayed in figure 31. There is a reverse exponential relationship between A and the expected number of rings, i.e. the larger the value of A, the fewer rings are expected on average. Thus it appears that we see more rings in the lower CHL systems, with the greatest number of rings being seen around the 20–40 mol% CHL systems. Note that the 20 mol% CHL system is an outlier.

A similar process was done for the number of DDCs and HCs in the systems, with the main difference being that most frames contain zero cages, thus the procedure must be

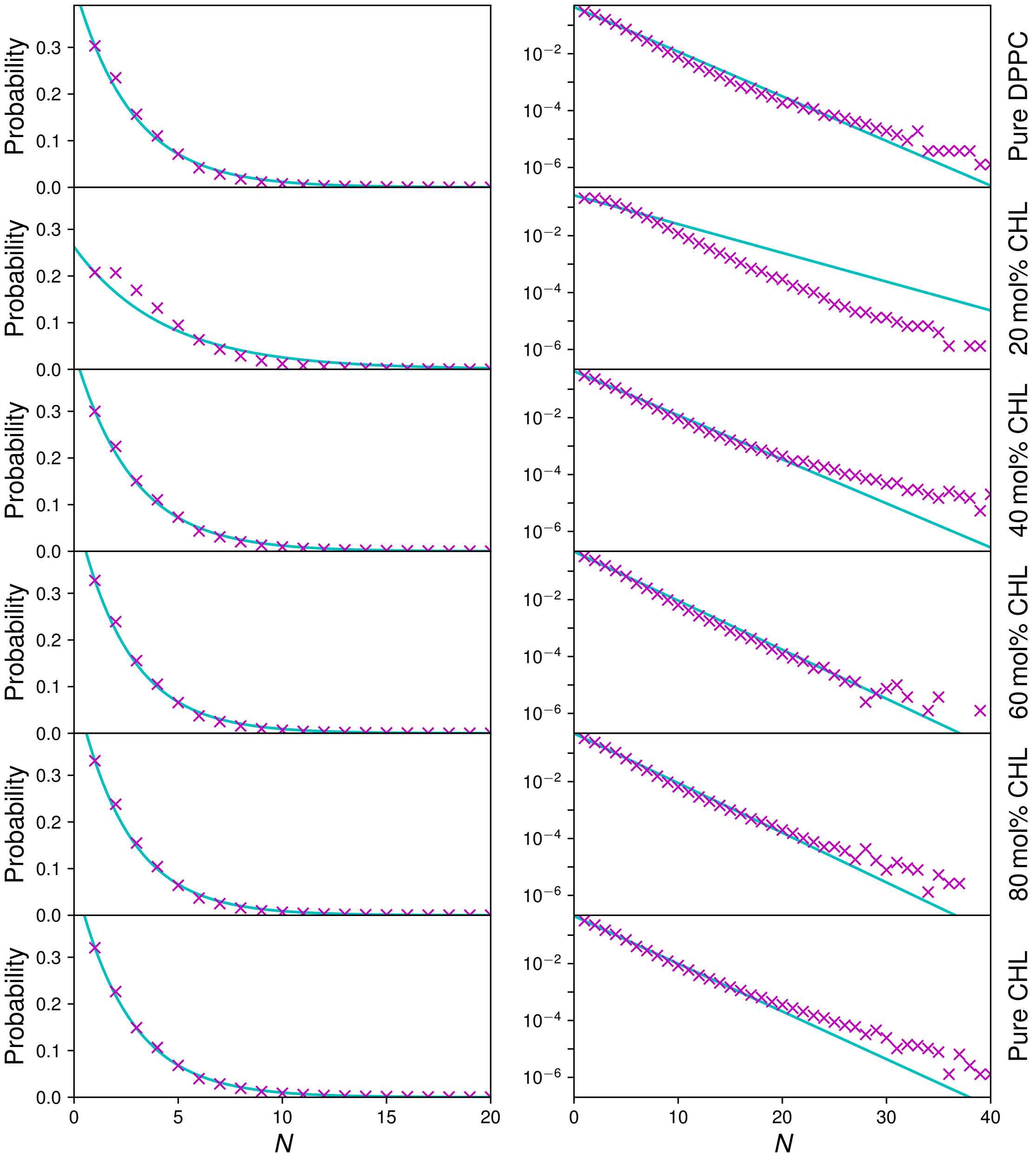


Figure 31: Probability distribution of number of rings per frame for DPPC-CHL systems. Values from simulations Ω_N are shown as magenta crosses, curves fitted of the form $\mathbb{P}[n_R = N] = Ae^{mN}$ are shown in cyan, values for A and m are given in table 4. The left hand panel shows the values with a linear scale, while on the right hand side, a semi-log scale is used, with the probability values plotted with a logarithmic scale, this allows us to see more detail for higher values of N.

Table 4: Values of A and m for fitted curves of the form $\mathbb{P}[n_R = N] = Ae^{mN}$. The squared sum of residuals (SSR) is also given.

System	A	m	SSR
Pure DPPC	0.435	−0.361	9.04×10^{-4}
20 mol% CHL	0.262	−0.233	5.37×10^{-3}
40 mol% CHL	0.429	−0.357	4.17×10^{-4}
60 mol% CHL	0.487	−0.397	5.83×10^{-4}
80 mol% CHL	0.495	−0.402	4.66×10^{-4}
Pure CHL	0.472	−0.386	1.80×10^{-4}

altered so as to include Ω_0. However, the exponential appearance only seems to start at $N = 1$ like with the rings, so equation (43) becomes:

$$\sum_{N=1}^{\infty} \mathbb{P}[n_C = N] = 1 - \Omega_0, \tag{52}$$

and equations (49) and (50) become, respectively:

$$m = \ln\left(1 - \frac{\Omega_1}{1 - \Omega_0}\right) \quad \text{and} \quad A = \frac{(1 - \Omega_0)\Omega_1}{1 - \Omega - \Omega_1}. \tag{53}$$

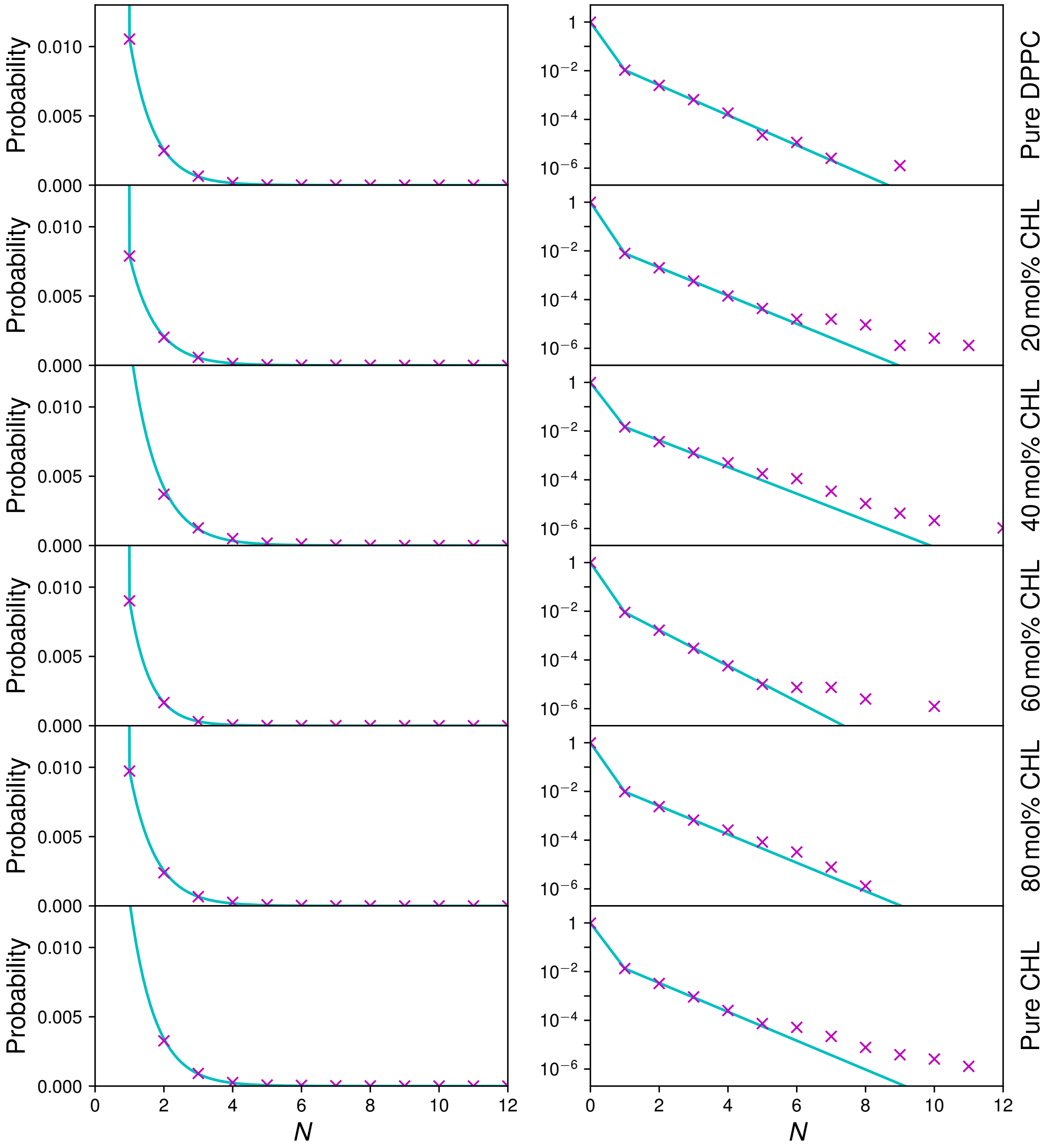
The final form of the approximate distribution is then given:

$$\mathbb{P}[n_C = N] = \begin{cases} \Omega_0 & N = 0 \\ Ae^{mN} & N > 0 \end{cases}. \tag{54}$$

The values for Ω_0 and the computed values for A and m for DDCs and HCs are given in table 5. The values and fitted curves for DDCs and HCs are shown in figures 32 and 33, respectively. Note that for both DDCs and HCs, $\mathbb{P}[n_C = 0] \approx 1$ for all systems, and the maximum numbers of cages seen are much lower than for the rings. Nonetheless, the fit is reasonably good for the DDCs, for HCs, the slope is steeper still and is hard to analyse.

Table 5: Values of A and m for fitted curves of the form given in equation (54).

System	DDCs			HCs		
	Ω_0	A	m	Ω_0	A	m
Pure DPPC	0.986	0.0436	−1.42	0.989	0.509	−3.85
20 mol% CHL	0.989	0.0298	−1.33	0.991	0.714	−4.36
40 mol% CHL	0.979	0.0519	−1.26	0.986	0.326	−3.17
60 mol% CHL	0.989	0.0485	−1.68	0.991	0.529	−4.09
80 mol% CHL	0.987	0.0373	−1.34	0.991	0.538	−4.16
Pure CHL	0.982	0.0523	−1.36	0.987	0.514	−3.74



$\Omega 0$

Figure 32: Probability distribution of number of DDCs per frame for DPPC-CHL systems. Values from simulations Ω_N are shown as magenta crosses, curves fitted of the form given in equation (54) are shown in cyan, values for A and m are given in table 5. The left hand panel shows the values with a linear scale, while on the right hand side, a semi-log scale is used, with the probability values plotted with a logarithmic scale, this allows us to see more detail for higher values of N. Note that the model was defined differently for $N = 0$ and for $N > 0$ in equation (54), which is why the logarithms of the curves are not linear.

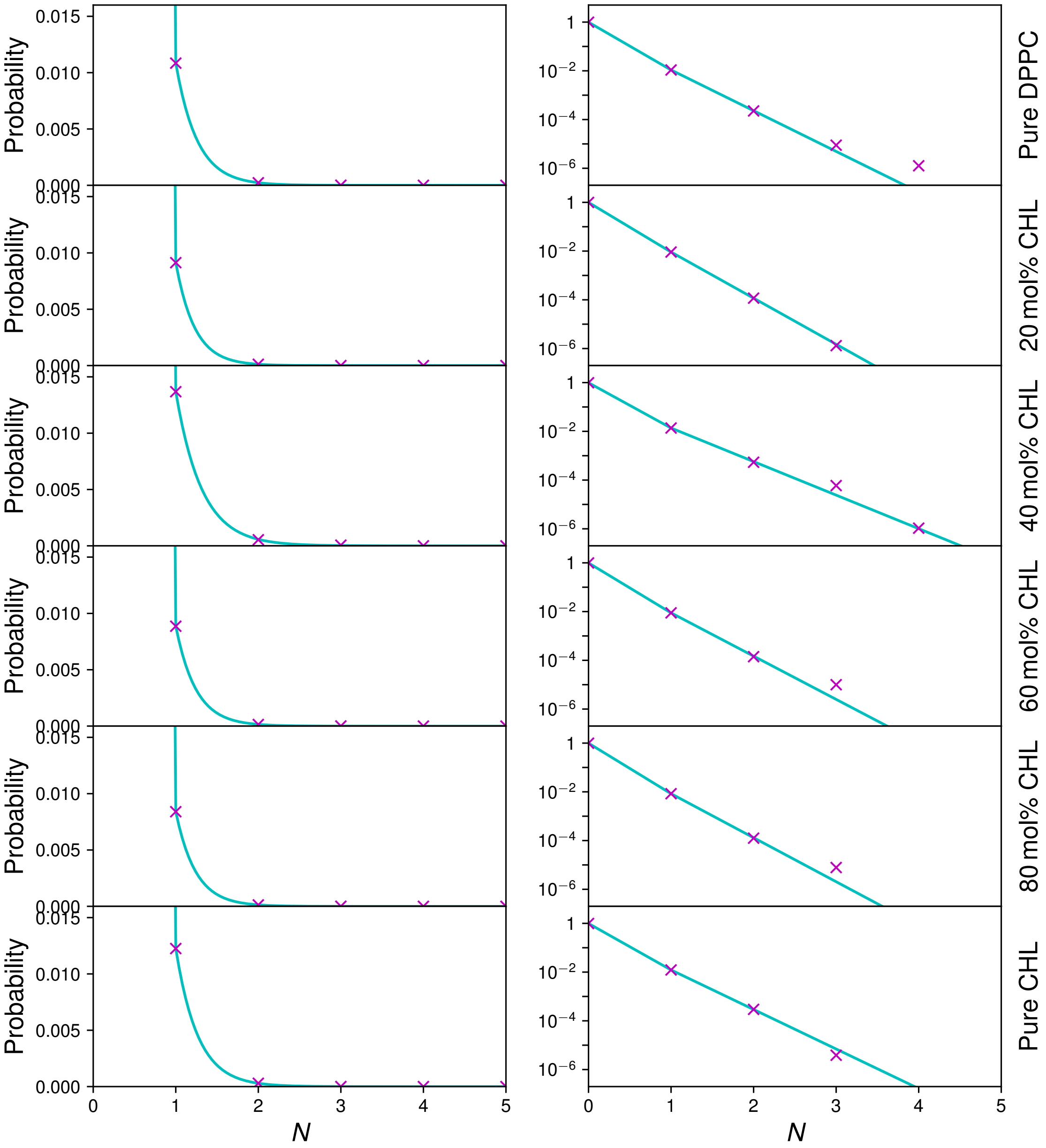


Figure 33: Probability distribution of number of HCs per frame for DPPC-CHL systems. Values from simulations Ω_N are shown as magenta crosses, curves fitted of the form given in equation (54) are shown in cyan, values for A and m are given in table 5. The left hand panel shows the values with a linear scale, while on the right hand side, a semi-log scale is used, with the probability values plotted with a logarithmic scale, this allows us to see more detail for higher values of N. Note that the model was defined differently for $N = 0$ and for $N > 0$ in equation (54), which is why the logarithms of the curves are not linear.

7.5 Discussion

Altogether, these results present evidence that phospholipid bilayers promote ice nucleation to some extent, however there is evidence to suggest that they are not especially potent INAs. They have also further elucidated the interplay between structural and chemical factors. In particular, pure DPPC, CHL and DMPI bilayers proved to have slightly more of an impact on ice nucleation than mixed bilayers of the same constituents. The condensing effect which CHL has on DPPC bilayers has a big impact on the accessibility of hydrogen bonding sites and this has been postulated as an explanation for why mixed DPPC-CHL bilayers have a lesser effect on ice nucleation than pure DPPC bilayers, despite CHL crystals and monolayers being known to be highly potent INAs. It is perhaps curious that no such potency was discovered for CHL bilayers. The LPS systems studied also displayed some ability to promote ice nucleation, however again their potency as INAs do not appear to be great.

These results present a step towards understanding the factors which make a surface a good or bad ice nucleating agent. The availability and accessibility of hydrogen bonding sites have been shown to correlate well with the cluster analysis. The structure of bilayers, but particularly the structure of these accessible hydrogen bonding sites is a key factor in the ice nucleating potential of a surface via “templating”. The orientational order parameter also provides some insight into the behavior of water molecules at the surface of the bilayer, with the trends seen particularly interesting in the DPPC-CHL and LPS systems.

Chapter 8

Synthesis of DPPC-CHL Vesicles

In addition to the computational study detailed in the previous chapters, a small amount of experimental work was undertaken. Although reduced in scale, partially due to the lack of lab access during the COVID pandemic, there are some conclusions which could be drawn from these experiments — chiefly to verify the setup of the simulation systems in terms of their scale. This chapter also builds foundations for possible future experimental work, in particular the concept that experimental work involving vesicles should be comparable with two-dimensional bilayer simulations.

8.1 Dynamic light scattering

Dynamic light scattering (DLS) is an analytical technique, wherein a polarised laser beam is shone through an aqueous sample. The intensity of scattered light at a fixed scattering angle θ is detected and this intensity is used to approximate the size of particles within the sample solution.[214] The cause of this phenomenon is Rayleigh scattering,[215,216] whereby small particles cause visible light to scatter. If the particles are sufficiently small in comparison to the wavelength of the laser used (typically the expected requirement is $d < \lambda/10$, where d is the particle diameter and λ is the wavelength of the laser) then it can be assumed that the light will be scattered isotropically in all directions.

According to the Rayleigh approximation, the intensity of scattered light, I, is related to the particle diameter and the wavelength by the expressions $I \propto d^6$ and $I \propto \lambda^{-4}$. The proportional relationship to d^6 means that measuring light scattered by a polydisperse solution is somewhat challenging, and therefore DLS works best for monodisperse solutions. The quartic inverse relationship with the wavelength means that a greater intensity of light is scattered when using a smaller wavelength, thus there is a trade off between having a wavelength which is sufficiently longer than the particle diameters but is not so long as to reduce the intensity of scattered light by too great an extent.[217]

The principle behind DLS is the measurement of Brownian motion, which can be related to the size of the particles in a solution. The larger a particle is, the slower its Brownian motion; with the velocity of this motion being defined by the translational diffusion coefficient, D. If this value is known, along with the viscosity of the solution, the size of a particle may be approximated, assuming that the particle is spherical, via the Stokes-Einstein equation:

$$d(H) = \frac{k_B T}{3\pi\eta D}, \tag{55}$$

where $d(H)$ is the hydrodynamic diameter, k_B is the Boltzmann constant, T is the temperature and η is the viscosity of the solution.[217]

If a laser is shone through a cuvette containing particles in solution, the Rayleigh scattering of light will form a speckle pattern which fluctuates constantly due to Brownian motion. In DLS, a correlator is used to assess the sizes of the particles. The principle of a correlator is to numerically compare the signals at different times, with perfect correlation (i.e. identical signals) taking the value of 1.0 and no correlation taking the value of 0.0. If δt is a very small time interval, typically on the scale of nanoseconds or microseconds, the signals at some initial time t and $t + \delta t$ will be very similar, but as we increase the time difference, $t + 2\delta t$, $t + 3\delta t$, $\cdots$, the signal will become less and less similar to the signal at time t, until eventually (as $t \to \infty$) there will be no correlation whatsoever. If the particles in a solution are large, the Brownian motion will be slow and therefore the correlation will decay slowly. Conversely, when the particles are small, the Brownian motion will be rapid and the correlation between signals will rapidly vanish. The shape of the correlogram (i.e. the graph of correlation over time) also contains information about the mono or polydispersity of the particles, with a more monodisperse sample giving a steeper line of decay.[217]

For a monodisperse sample, the correlation function $G(\tau)$, where τ is the time difference, can be given by the expression:

$$G(\tau) = \alpha + \beta \exp[-2\Gamma\tau], \tag{56}$$

where α is the baseline of the correlation function (i.e. $G(\tau) \to A$ as $\tau \to \infty$) and β is the intercept of the correlation function (i.e. $G(0) = \beta$). The value of Γ is defined thus:

$$\Gamma = Dq^2, \tag{57}$$

where D is the translational diffusion coefficient and

$$q = \frac{4\pi n}{\lambda} \sin(\theta/2), \tag{58}$$

where n is the refractive index of the dispersant, λ is the wavelength of the laser and θ is the scattering angle, where the detector is positioned. For polydisperse samples, the respective exponential decays from the different particle sizes must be summed. By combination of equations (55) and (56), the hydrodynamic diameter can be approximated, with the polydispersity index also possible to approximate from the correlation function.[217]

8.2 Microlitre nucleation by immersed particles

The microlitre by immersed particle (µl-NIPI) instrument as described in detail by Whale *et al.*[62] consists of a hydrophobic surface on top of a cold stage, with an enclosure in which humidity may be controlled by means of a dry gas flow. Around forty droplets of volume 1 µl are deposited on the surface at ambient conditions. The chamber is then closed, and dry zero-grade nitrogen pumped through to control prevent frost growth. The cold stage is then programmed to perform a 1 °C/ min temperature ramp from 20 to −35 °C. A camera placed above the stage is used to record the progress at a rate of 1 frame per second, with the fraction of droplets frozen over temperature being the parameter of interest. This is a naturally scalable process, as any number multiple runs can be combined into one, provided that the conditions are sufficiently controlled. As nucleation is a stochastic process, in theory, the larger the number of droplets, the closer to the correct result should be obtained. The limit here is of course the time taken to run one experiment — approximately one hour for around forty droplets.

From the distribution of the proportion of water droplets frozen over temperature, one can calculate the differential nucleus concentration $k(T)$ at temperature T as follows:[218]

$$k(T) = -\frac{1}{X\Delta T} \cdot \ln\left(1 - \frac{\Delta N}{N(T)}\right), \tag{59}$$

where ΔT is the step in temperature between frames, $N(T)$ is the number of unfrozen droplets remaining at temperature T, ΔN is the number of nucleation events between temperature T and temperature $T + \Delta T$ and X is a normalisation constant to some unit size of droplet. In this case, as we assume the droplets to all be of equal size 1 µl, we may simply use $X = 1$. The differential nucleus concentration must be calculated for both the samples in question, and a control of the solution in which the samples are immersed. Ideally any solutes in addition tho the sample of interest should have no effect on ice nucleation. From these two measurements, $k(T)$ for the control can be subtracted from $k(T)$ for the sample to determine what effect the immersed particles have on ice nucleation.

8.3 Experimental methods

DPPC and CHL were obtained from Sigma-Aldrich. DPPC and CHL were dissolved in 2 ml of 3:1 chloroform-methanol mixture in round-bottomed flasks. Exact amounts of each lipid to produce vesicles ranging from 0–80 mol% CHL are given in table 6, along with the exact molar concentrations of CHL, attained using the molecular masses of 734.0 g mol^{-1} and 386.7 g mol^{-1} for DPPC and CHL, respectively.[219,220] For the sake of brevity, the systems will be referred to throughout this chapter by the target molar concentration of CHL rather than the exact values. Each solution was dried in a rotary evaporator at 200 rpm for one hour. The resulting lipid film was re-dissolved in an appropriate (see table 6) volume of 25 mM sodium phosphate (NaPi) buffer solution, pH 6.5, to a lipid concentration of 3.3 mg ml^{-1}. The solutions were then sonicated for two minutes without heating. Finally, the solutions were frozen at −20 °C, thawed and then sonicated for another two minutes. This final step was repeated four times in total before transferring the vesicle solutions into falcon tubes for storage at −20 °C.

Table 6: Mass of DPPC and CHL and volume of NaPi buffer used for vesicle synthesis.

Target mol% CHL	**0**	**5**	**10**	**15**	**20**	**25**	**30**	**35**	**40**	**80**
Mass DPPC [mg]	24.9	43.6	20.2	20.0	17.7	18.3	17.0	16.1	16.2	6.5
Mass CHL [mg]	0	2.0	1.2	2.1	2.3	3.7	4.5	4.8	6.1	13.7
Volume NaPi [ml]	7.55	13.8	6.48	6.70	6.06	6.67	6.52	6.33	6.76	6.12
Exact mol% CHL	**0.0**	**8.0**	**10.1**	**16.6**	**19.8**	**27.7**	**33.4**	**36.1**	**41.7**	**80.0**

In order to carry out DLS on the vesicles, first they were thawed. Once defrosted, the samples were sonicated for two minutes. Samples were diluted to 0.6 mg ml^{-1} in 25 mM NaPi buffer, pH 6.5. The diameters of the vesicles were approximated via DLS, using a Zetasizer Nano-series instrument (Malvern Instruments, UK) with 1 cm path-length UV-transparent disposable cuvettes. A sample of each solution was processed six times for 300 s at 20 °C and a further six times at 50 °C. Intensity and number distributions were exported, using the Malvern Zetasizer software.

Microdrop freezing assays were attempted with the samples, unfortunately it was found that the buffer solution in which the vesicles were suspended itself acts as an INA, see figure 34. Therefore this experiment was inconclusive. It would be possible as a future work to attempt to synthesise similar vesicles in de-ionised water, which does not freeze until a very low temperature (∼232 K).[1]

8.4 Results

Results from the DLS experiments were somewhat varied. In some systems, namely 0, 5, 30 and 80 mol% CHL, very similar distributions were obtained for all runs and at both temperatures. In the case of the 10, 15 and 20 mol% CHL systems, consistent distributions were obtained for each of the six runs at both 20 and 50 K, however there was a discrepancy in the hydrodynamic diameters calculated when comparing the runs at the two temperatures. The remaining systems (25, 35 and 40 mol% CHL) showed more uncertainty in the analysis of the DLS results, with inconsistent hydrodynamic diameters calculated from the different runs — manifesting as two separate peaks in the number distribution or, in the case of 35 mol% CHL at 50 K, a broader distribution of multiple peaks. Number distributions from DLS for the ten systems with the values averaged over six runs and with outliers removed are shown in figure 35. DLS as a technique works best on monodisperse solutions and it is fairly probable that the messiness in some of these DLS results is due to a high degree of polydispersity. For the further study of such vesicles it might be wise to reduce this polydispersity, e.g. by extruding through polycarbonate membranes to create more uniform vesicle populations.[221]

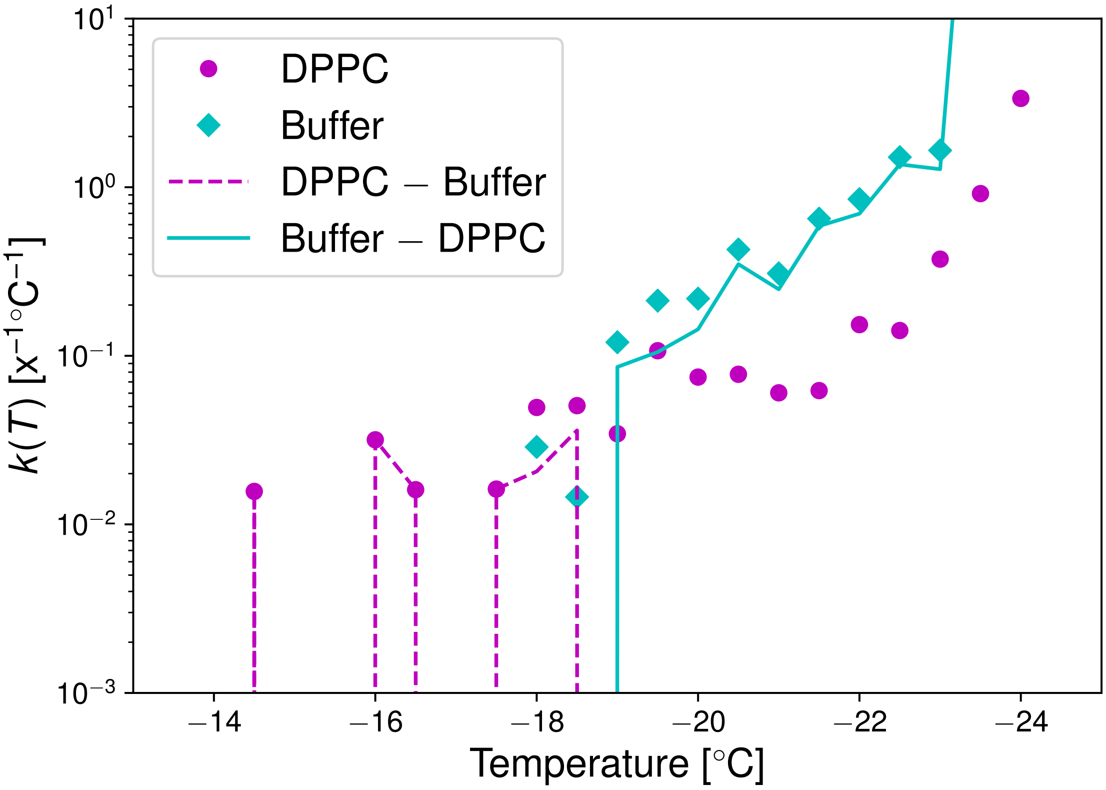


Figure 34: Differential nucleus concentrations $k(T)$ (see equation (59)) for DPPC vesicles and buffer solution (magenta circles and cyan diamonds) difference between $k(T)$ values shown with magenta dashed line (DPPC-buffer) and cyan solid line (buffer-DPPC). A higher $k(T)$ indicates a larger number of nucleation events at temperature T.

Despite the suspected polydispersity of the samples, in all cases it was clear that vesicles formed with diameters of around 40–600 nm. When we consider the dimensions of the simulation boxes, as detailed in table 1, we note that the x and y dimensions ranged from 3.31–4.20 nm. If we now consider a fairly extreme case where we choose one of the smaller vesicles, with $d = 40\,\text{nm}$, and we consider a 4×4 nm section of the surface, we would be interested to know how much curvature that section would have. As we are assuming a perfectly spherical shape, we can simplify it by considering only two dimensions, i.e. considering a circle of diameter 4 nm, and a section of the edge with length 4 nm. Clearly the circumference of this circle is 40π nm, and thus the central angle of the segment is:

$$\theta = 2\pi \cdot \frac{4\,\text{nm}}{40\pi\,\text{nm}} = 0.2^{\text{c}}, \tag{60}$$

with angles expressed in radians. From here we can easily obtain that the difference in height between the edges and the centre of the 4×4 nm region would be:

$$\delta z = 20\,\text{nm}(1 - \cos\theta) \approx 0.4\,\text{nm}. \tag{61}$$

Such a slight curvature in a region of this size, even in the case of the smallest of the

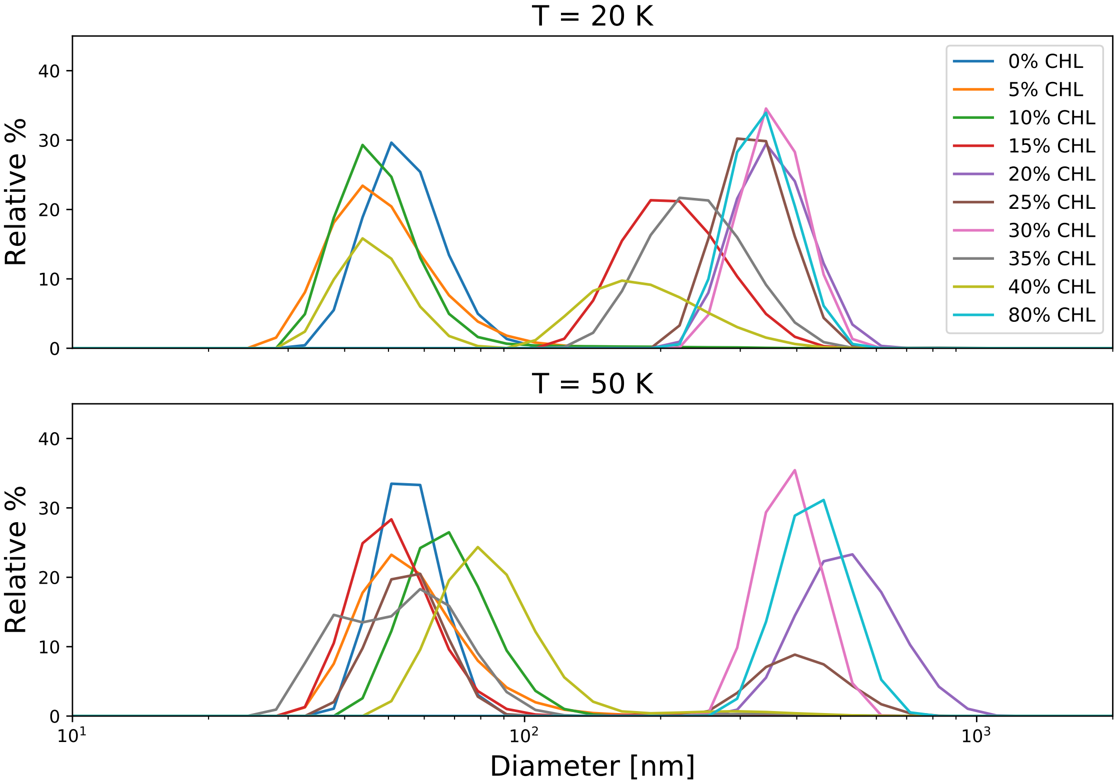


Figure 35: Number distributions from DLS for the ten systems. Values averaged over six runs, with outliers removed.

vesicles would suggest that vesicles such as the ones created by this method are directly comparable to the bilayers that have been simulated, and that vesicles such as these could indeed provide experimental insight to complement future theoretical research.

8.5 Discussion

Altogether the experimental work detailed in this chapter introduces some methods which can be used to complement molecular simulations of lipid bilayers, specifically with a focus on their potential to act as ice nucleating agents. In particular, the method described in section 8.3 synthesises vesicles which are of a large enough diameter that the are comparable to planar simulation systems. An experimental method for assessing such vesicles' ability to promote (or inhibit) ice nucleation was described in section 8.2. Unfortunately, the vesicles were synthesised in a buffer solution which itself promoted ice nucleation, so these experiments were unsuccessful. In future work it would be preferable to synthesise vesicles in de-ionised water, whereupon the µl-NIPI technique should yield better results.

Chapter 9

Permeation of Drug Molecules Through Phospholipid Bilayers

9.1 Naproxen and felodipine

Naproxen is a non-steroidal anti-inflammatory agent, used to treat inflammatory diseases such as osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, tendinitis, bursitis and acute gout. It is also used for pain relief, particularly for patients suffering moderate migraines, and for treatment of primary dysmenorrhea and for reduction of fevers. It is typically administered orally.[222] Felodipine is a calcium-channel blocking agent, used for management of hypertension. It is also administered orally.[223]

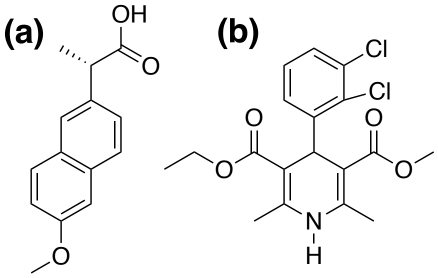


Figure 36: The chemical structures of **(a)** naproxen and **(b)** felodipine.

A key issue in the delivery of small drug molecules such as naproxen and felodipine is their ability to permeate across cellular membranes, specifically the gastrointestinal membrane.[224,225] In fact a very large number, 25–40 %, of approved drugs suffer from low membrane permeability.[225–228] These two drug molecules were chosen as suitable candidates for permeation studies and with the view of comparing to experimental results.

9.2 Enhanced sampling for MD simulations

While MD simulations can be excellent for the study of equilibrium dynamics, the extent of timescales which are accessible are limited by computing resources. For the systems studied in this thesis, which are on the order of 100,000 atoms, simulations longer than a few microseconds are unfeasible without coarse-graining, which would clearly have a negative effect on our ability to study water-water and water-bilayer dynamics, due to the reduction in resolution.[*] Many natural phenomena occur on a timescale far greater than such timescales, so called "rare events", rendering MD unsuitable for their study. One method for overcoming this issue is to apply some form of bias to the simulation — to accelerate the rate of occurrence of such rare events.

9.2.1 Collective variables

Many enhanced sampling techniques work by applying some form of bias to a simulation, in order to force it to explore the full free energy landscape of interest. Naturally, even an extremely small molecular simulation has an enormous number of coordinates — far too many to be able to fully explore the entire range of possible configurations. Thus it is essential to reduce the dimensionality of the system to a small number of important characteristics. Thus, one must choose a collective variable (CV)[†] or set of CVs which are of as little dimension as possible, but which fully describe the process of interest. Selecting such CVs acts to reduce a very complex simulation box to a single, or a few dimensions.

There are a great many different types of CV which may be used for biased MD. Two examples are the use of order parameters, such as q_6 (see equation (8)), and the potential energy of the system; as used in combination by Trudu *et al.*[233], to study the freezing of a Lennard-Jones fluid, and Quigley and Rodger[234] to study ice nucleation and growth. Another CV which is commonly used, in particular for exploring chemical reaction pathways is the coordination number of important atoms, as used by Ensing *et al.*[235] to study the E2 reaction between fluoroethane and a fluoride ion. The simplest types of CVs are geometric quantities, such as distances, angles and dihedral angles formed between atoms, molecules or larger groups of atoms. It is the geometric CV which shall be used later on in this chapter, with z-distance between two centres of mass parameterising a one-dimensional free-energy landscape.

9.2.2 Metadynamics

Metadynamics (MT) is one such method of accelerating rare events. Employing a CV or set of CVs, the principle of MT is to procedurally apply Gaussian potentials to this reduced dimension space, biasing the simulation away from free-energy regions which have been previously explored. This allows the simulation to overcome free energy barriers and discover distant local minima.[236,237] On subtracting the applied biases, this results in a single free energy curve (or surface). The MT method has been explained without great

---

[*] In fact, this loss in detail makes it impossible to parameterise the force field in such a way that reproduces all properties faithfully.[229–231]

[†] A collective variable (CV) is a differentiable function which describes the state of the molecular system, chosen such as to describe the process of some transition of interest.[232]

detail here as while it was attempted, it proved unsuccessful in this use case. Nonetheless, details of the MT simulations are given in section 9.3.3.

9.2.3 Umbrella sampling

Umbrella sampling (US) is another method for biasing molecular simulations. Similarly to MT, US employs a CV, or set of CVs, to reduce the system to one or a few dimensions. In US, an energetic bias is applied at selected points across the range of the CV(s), in order to connect energetically separated regions.[238–240] These bias potentials w are defined to depend solely on the reaction coordinate and are added to the unbiased energy (E^{u}) of the system to create the final biased energy (E^{b}) used in the simulation:

$$E^{\mathrm{b}}(\mathbf{r}) = E^{\mathrm{u}}(\mathbf{r}) + w(\xi), \tag{62}$$

where $\mathbf{r}$ is the state of the system and ξ is the reaction coordinate. According to derivations by Kästner[240], the unbiased free energy $A(\xi)$, for the individual window, is given by:

$$A(\xi) = -\frac{1}{\beta}\ln(P^{\mathrm{b}}(\xi)) - w(\xi) + F, \tag{63}$$

where $\beta = 1/(k_B T)$, with k_B the Boltzmann constant and T the temperature. P^{b} is the biased probability distribution along the reaction coordinate ξ and

$$F = -\frac{1}{\beta}\ln\left\langle \exp\left[-\beta w(\xi)\right] \right\rangle_{\xi} \tag{64}$$

with $\langle\cdot\rangle_{\xi}$ denoting an ensemble average over ξ.

Typically for simulations such as these, a harmonic bias potential is applied of the form:

$$w(\xi) = \frac{k}{2}(\xi - \xi_0)^2, \tag{65}$$

where k is the spring constant and ξ_0 is the centre of the biasing potential.

In practice, a single window for US will result in a simulation which is more restricted along the reaction coordinate than an unbiased simulation. Thus a number of simulations are run with biasing potentials at selected points. When choosing the number of simulation windows for US and the bias potential to employ, it is important to ensure that the entire region of interest within the phase space is sampled. This can be verified by summing the individual histograms:

$$\Omega(\xi_j) = \sum_{i=1}^{S}\Omega_i(\xi_j), \tag{66}$$

where $\Omega_i(\xi_j)$ is the value of the histogram for some bin ξ_j and for the i-th simulation, where S is the total number of simulation windows. If $\Omega(\xi_j) = 0$ for any ξ_j then the the space has not been fully sampled and therefore one must either adjust the biasing potential or add more windows. Such individual histograms Ω_i are shown in the top panels of figures 39 and 41 with the sums, Ω shown in the second panels. Strictly speaking, the count is plotted and not a histogram, however as the bins are of equal width, this is irrelevant. Naturally, these separate runs will only provide reliable information within the region of phase space that is fully explored, and thus the unbiased free energy estimates obtained A_i must be combined.

9.2.4 Weighted histogram analysis method

The most commonly employed method for combining these free energy curves is the weighted histogram analysis method (WHAM),[241,242] which weights the individual free energy estimates in such a way as to minimise the statistical error. Suppose we have S simulations, employing biasing potentials w_i, over the reaction coordinate ξ, with $i = 1, \ldots, S$. Let $P_i^{\mathrm{b}}(\xi)$ be the probability distribution of values over ξ from the i-th biased simulation, then the best approximation for the unbiased probability distribution we can get from this single simulation is:

$$P_i^{\mathrm{u}}(\xi) = \exp\left[\beta\left(w_i(\xi) - f_i\right)\right] P_i^{\mathrm{b}}(\xi), \tag{67}$$

where f_i is the free energy obtained from adding the biasing potential to the reference potential, as in equation (62).

The basic principle behind WHAM is to perform a weighted average of the individual simulations:

$$P(\xi) = C \sum_{i=1}^{S} \rho_i(\xi) P_i^{\mathrm{u}}(\xi), \tag{68}$$

where ρ_i is the weighting for the i-th window and C is a normalisation constant. The weights ρ_i are chosen so as to minimise the statistical error of P, subject to

$$\sum_{i=1}^{S} \rho_i = 1. \tag{69}$$

The following derivation closely follows the work of Kumar *et al.*[241] and Souaille and Roux[242]. Assuming that all the simulations are statistically independent, the weightings ρ_i can be separated from the statistical error:

$$\sigma^2[P(\xi)] = C \sum_{i=1}^{S} \rho_i^2(\xi) \sigma^2[P_i^{\mathrm{u}}(\xi)], \tag{70}$$

where $\sigma^2[P(\xi)]$ and $\sigma^2[P_i^{\mathrm{u}}(\xi)]$ represent the statistical errors in $P(\xi)$ and $P_i^{\mathrm{u}}(\xi)$, respectively. We can choose ρ_i to minimise this error using the method of Lagrange multipliers, subject to the condition (69). This involves solving the set of equations:

$$\frac{\partial}{\partial \rho_i}\left(\sigma^2[P(\xi)] - \lambda\left(\sum_{i=1}^{S} \rho_i - 1\right)\right) = 0 \quad , \quad \forall i = 1, \ldots, S. \tag{71}$$

Here λ is the Lagrange multiplier. The error $\sigma^2[P_i^{\mathrm{u}}(\xi)]$ does not depend on the values of ρ_i, thus equation (71) rearranges to give:

$$\sigma^2[P_i^{\mathrm{u}}(\xi)]^{-1} = \frac{2C}{\lambda}\rho_i(\xi) \quad , \quad \forall i = 1, \ldots, S. \tag{72}$$

Since this is true for all $i = 1, \ldots, S$, we can sum to give:

$$\sum_{j=1}^{S} \sigma^2[P_j^{\mathrm{u}}(\xi)]^{-1} = \sum_{j=1}^{S} \frac{2C}{\lambda}\rho_j(\xi) = \frac{2C}{\lambda} C \sum_{j=1}^{S} \rho_j(\xi) = \frac{2C}{\lambda}. \tag{73}$$

Equation (71) can also be rearranged and combined with equation (73) to give:

$$\rho_i(\xi) = \frac{\lambda}{2C}\sigma^2[P_i^{\mathrm{u}}(\xi)]^{-1} = \frac{\sigma^2[P_i^{\mathrm{u}}(\xi)]^{-1}}{\sum_{j=1}^{S}\sigma^2[P_j^{\mathrm{u}}(\xi)]^{-1}} \quad , \quad \forall i = 1, \ldots, S. \tag{74}$$

From equation (67), we get the following:

$$\sigma^2[P_i^{\mathrm{u}}(\xi)] = \exp\big[2\beta\,(w_i(\xi) - f_i)\big] \cdot \sigma^2[P_i^{\mathrm{b}}(\xi)] \tag{75}$$

and $\sigma^2[P_i^{\mathrm{b}}(\xi)]$ can be expressed as follows:[242,243]

$$\sigma^2[P_i^{\mathrm{b}}(\xi)] = \frac{g_i(\xi)}{n_i \Delta\xi}\overline{P_i^{\mathrm{b}}(\xi)}, \tag{76}$$

where n_i is the length of the i-th simulation, $\Delta\xi$ is the width of the bins used to calculate the histogram, $\overline{P_i^{\mathrm{b}}(\xi)}$ is the probability distribution obtained from a hypothetical infinite-length simulation and $g_i(\xi) = 1 + \tau_i(\xi)/\Delta t$, where $\tau_i(\xi)$ is the correlation time and Δt is the time step. The factors g_i can be assumed to be equal for all simulations under US such as is detailed in this thesis.

By substituting $\overline{P_i^{\mathrm{b}}(\xi)}$ for $P_i^{\mathrm{b}}(\xi)$ and $P(\xi)$ for in equation (67), we obtain an estimate for $\overline{P_i^{\mathrm{b}}(\xi)}$:

$$\overline{P_i^{\mathrm{b}}(\xi)} = \exp\big[-\beta\,(w_i(\xi) - f_i)\big]P(\xi). \tag{77}$$

Now, using equations (74) to (77), we obtain the following:

$$\begin{aligned}
\rho_i(\xi) &= \frac{\sigma^2[P_i^{\mathrm{u}}(\xi)]^{-1}}{\sum_{j=1}^{S}\sigma^2[P_j^{\mathrm{u}}(\xi)]^{-1}} = \frac{\exp\big[-2\beta\,(w_i(\xi) - f_i)\big] \cdot \sigma^2[P_i^{\mathrm{b}}(\xi)]^{-1}}{\sum_{j=1}^{S}\Big(\exp\big[-2\beta\,(w_j(\xi) - f_j)\big] \cdot \sigma^2[P_j^{\mathrm{b}}(\xi)]^{-1}\Big)} \\
&= \frac{g_i(\xi)^{-1}\Delta\xi P(\xi)^{-1} \cdot n_i \exp\big[-\beta\,(w_i(\xi) - f_i)\big]}{\sum_{j=1}^{S}\big(g_j(\xi)^{-1}\Delta\xi P(\xi)^{-1} \cdot n_j \exp\big[-\beta\,(w_j(\xi) - f_j)\big]\big)} \\
{}^{*} &= \frac{n_i \exp\big[-\beta\,(w_i(\xi) - f_i)\big]}{\sum_{j=1}^{S}\big(n_j \exp\big[-\beta\,(w_j(\xi) - f_j)\big]\big)}.
\end{aligned} \tag{78}$$

---

[*] Using the assumption that the g_i are equal for all simulations.

Finally, we substitute equations (67) and (78) into equation (68) to give

$$\begin{aligned} P(\xi) &= C\sum_{i=1}^{S}\rho_i(\xi)P_i^{\mathrm{u}}(\xi) = C\sum_{i=1}^{S}\frac{n_i\exp\left[-\beta\left(w_i(\xi)-f_i\right)\right]}{\sum_{j=1}^{S}n_j\exp\left[-\beta\left(w_j(\xi)-f_j\right)\right]}\cdot P_i^{\mathrm{u}}(\xi) \\ &= C\sum_{i=1}^{S}\frac{n_i\exp\left[-\beta\left(w_i(\xi)-f_i\right)\right]}{\sum_{j=1}^{S}n_j\exp\left[-\beta\left(w_j(\xi)-f_j\right)\right]}\cdot\exp\left[\beta\left(w_i(\xi)-f_i\right)\right]P_i^{\mathrm{b}}(\xi) \\ &= C\sum_{i=1}^{S}\frac{n_iP_i^{\mathrm{b}}(\xi)}{\sum_{j=1}^{S}n_j\exp\left[-\beta\left(w_j(\xi)-f_j\right)\right]}. \end{aligned} \tag{79}$$

Equation (79) is the first of what is known as the WHAM equations.

To derive the second WHAM equation, first note that, by definition:

$$\exp\left[-\beta f_k\right] = \int P(\xi)\exp\left[-\beta w_k(\xi)\right]\mathrm{d}\xi, \tag{80}$$

and thus

$$\exp\left[-\beta f_k\right] = C\int\sum_{i=1}^{S}\frac{n_iP_i^{\mathrm{b}}(\xi)\exp\left[-\beta w_k(\xi)\right]\mathrm{d}\xi}{\sum_{j=1}^{S}n_j\exp\left[-\beta\left(w_j(\xi)-f_j\right)\right]}\mathrm{d}\xi, \tag{81}$$

which can be iterated to compute the free energy f_k.

9.3 Simulation details

A Hydro Soy PC* (HSPC, see figure 37(a)) bilayer was first created and equilibrated at 300 K before adding molecules of the drugs naproxen and felodipine (see figure 36). As with the drug molecules, this specific membrane was selected for comparison with experimental results.

9.3.1 Hydro Soy PC bilayer

A pure 1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC, see figure 37(b)) lipid bilayer system was constructed, using CHARMM-GUI,[147–151] with 30 lipids per leaflet (60 per system). The resulting structure and topology files were manually edited, removing two chain carbons from the primary tail, to create an HSPC bilayer. A TIP4P/Ice water layer 30 Å thick was placed either side of the bilayers, using the MD package GROMACS[152–158]

This bilayer water system was energy minimised, equilibrated, elongated by adding vacuum regions and equilibrated with EW3DC as described in section 3.3, using the same force fields as for the previous systems. These initial equilibrations were carried out at 328.15 K. Following equilibration in the liquid crystalline phase,[244] the system was annealed to 300 K and equilibrated at this cooler temperature, again following the procedure from section 3.3. Now at room temperature, the bilayer has entered the gel phase.[244] Once again following the procedure laid out in section 3.3, the system was equilibrated

---

[*] L-α-phosphatidylcholine, hydrogenated (Soy).

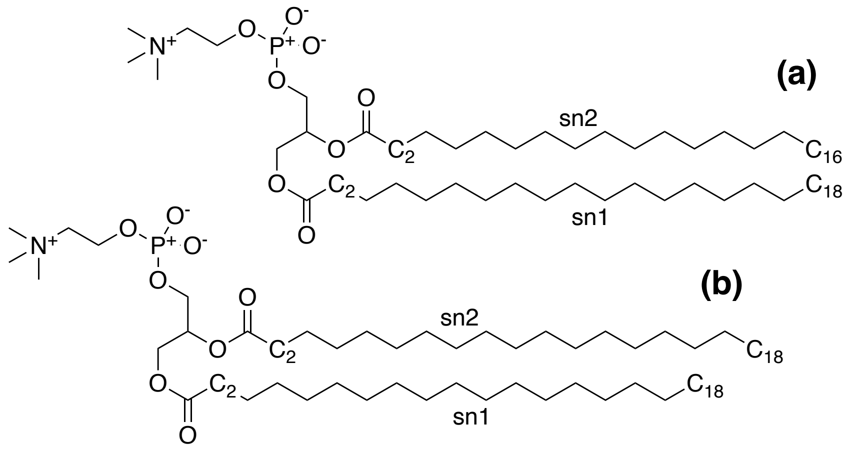


Figure 37: The chemical structures of **(a)** HSPC and **(b)** DSPC.

at 300 K, sampling the NpT ensemble. The in-plane dimensions of the simulation box, once equilibrated and quenched, was 3.89 nm. The systems have square cross section so the x and y dimensions are the same. As with the other simulations, these simulations were carried out in GROMACS using the `v-rescale` thermostat, Berendsen barostat with surface tension coupling (where relevant), Verlet cutoff scheme, plain cutoff for Van der Waals interactions, with `force-switch`.

9.3.2 Adding small drug molecules

The small drug molecules naproxen and felodipine (see figure 36) were each hydrated with TIP4P water, within a 3.2 nm cubic box. In these simulations, naproxen was simulated with the hydrogen from its carboxylic group dissociated. Using three dimensional PBCs in GROMACS,[152–158] the two systems were energy minimised using the SETTLE[166] algorithm to constrain the geometry of the water molecules, and the LINCS[167] algorithm to constrain the geometry of the drug molecules. Following this, equilibration runs under the NVT and subsequently the NpT ensemble were undertaken. The CHARMM36[31,160–164] and TIP4P/Ice[116] force fields were used.

Following equilibration of the drug molecules, the drug molecules were solvated into the water layers of the HSPC system, creating four bilayer-drug systems: one with a single naproxen molecule, one with a single felodipine molecule, one with two naproxen molecules (one either side of the bilayer) and one with two felodipine molecules (again, one either side of the bilayer). The idea with the two molecule systems was to identify whether interactions between two drug molecules had a relevant effect on the free energy landscape, however the final US simulations only used a single molecule. These four systems were equilibrated via a 20 ns NVT run, followed by a 20 ns NpT run. The simulations were run using EW3DC, at a temperature of 300 K, with the same simulation setup as in section 3.3. During both runs, walls were imposed in the z-direction, at 45 ns from either side of the bilayer centre, prohibiting the drug molecule(s) from exiting the water layer, using PLUMED2.[186–188]

9.3.3 Metadynamics

Following the unbiased equilibration, MT runs were initiated, using the NpT ensemble. Biasing and walls, as before, were applied, using PLUMED2.[186–188] The variable used for biasing was the z-component of the centre of mass of the drug molecule, taken relative to the centre of the bilayer, in other words the CV was:

$$\delta z = \mathrm{COM}(\mathrm{drug})_z - \mathrm{COM}(\mathrm{bilayer})_z, \tag{82}$$

where $\mathrm{COM}(G)$ is the centre of mass of atom group G and $\mathbf{r}_z$ is the z component of vector $\mathbf{r}$. For the systems with two drug molecules, only one was biased but the walls were applied to both. Various parameters were tried for the MD setup, ranging from small gaussians, with width 0.1 nm and height $1.0\,\mathrm{kJ\,mol^{-1}}$, to larger gaussians, with width 0.3 nm and height $4.0\,\mathrm{kJ\,mol^{-1}}$.

As discussed in section 9.4, these MT runs were unsuccessful in characterising the free energy landscape. It is possible that a MT run involving a more complex CV, such as one comprised of both the z-component and also some measure of the molecules orientation may yield better results.

9.3.4 Umbrella Sampling

In parallel with the MT runs, US runs were run as a continuation from the equilibrated trajectories. Setting up simulations employing US is more complicated than in the case of MT, as many separate simulations are required, with biases applied at different points along the CV. As with MT, the CV used, δz, was the z-component of the centre of mass of the drug molecule, taken relative to the centre of the bilayer. An initial, continuous simulation was run, with the location of the bias moving by 5 Å every 5 ns; i.e. a 5 ns run with the bias centred at $z = +40$ Å, a 5 ns run with the bias centred at $z = +35$ Å and so on up until $z = -40$ Å. The biases applied were harmonic, with spring constant: $k = 250\,\mathrm{kJ\,mol^{-1}}$ (see equation (65)). The walls at ±45 Å were also applied as before. These first US runs primarily served to generated starting configurations for subsequent runs, with the drug molecules at various z positions across the simulation box.

Following those initial runs, a number of different spring constants were tested in order to find a setup that samples the entire range of the CV (in this case z). This was achieved by increasing the spring constant on successive runs, until the full energy landscape was surveyed, i.e. so that there was sufficient overlap in the histograms from the first panels of figures 39 and 41, which show the distribution of the drug molecules across all the runs. While this was achievable in this case with 1 Å spacing, increasing the spring constant will cause those histograms to narrow, and so it could have been necessary to add additional “umbrellas”. Finally the value $k = 2000\,\mathrm{kJ\,mol^{-1}}$ was settled on, which was strong enough to overcome the free energy barriers within the bilayer, while being sufficiently weak so that there was overlap between simulations using 1 Å spacing. The WHAM code written by Grossfield[245] was used to compute the free energy landscapes with respect to the CV δz. The US simulation windows were run for 10 ns for the regions outside of the bilayer. The regions within the bilayer were run for a total of 30 ns for the felodipine system, after which the free energy surface had converged reasonably well. For the naproxen system, the parallel runs were extended to a total of 50 ns each as the free energy surface had not converged so successfully, however the convergence was still incomplete (see section 9.4).

9.4 Results

Figure 38 shows the results of MT simulations. These simulations were run with Gaussian biases of width 0.1 nm and height 1.0 kJ mol^{-1}, at temperature 300 K and for run lengths of 290–301 ns with 2 fs MD steps and biases applied every 100 steps. The two systems with a single drug molecule seem to have had little success in exploring the free energy landscape.

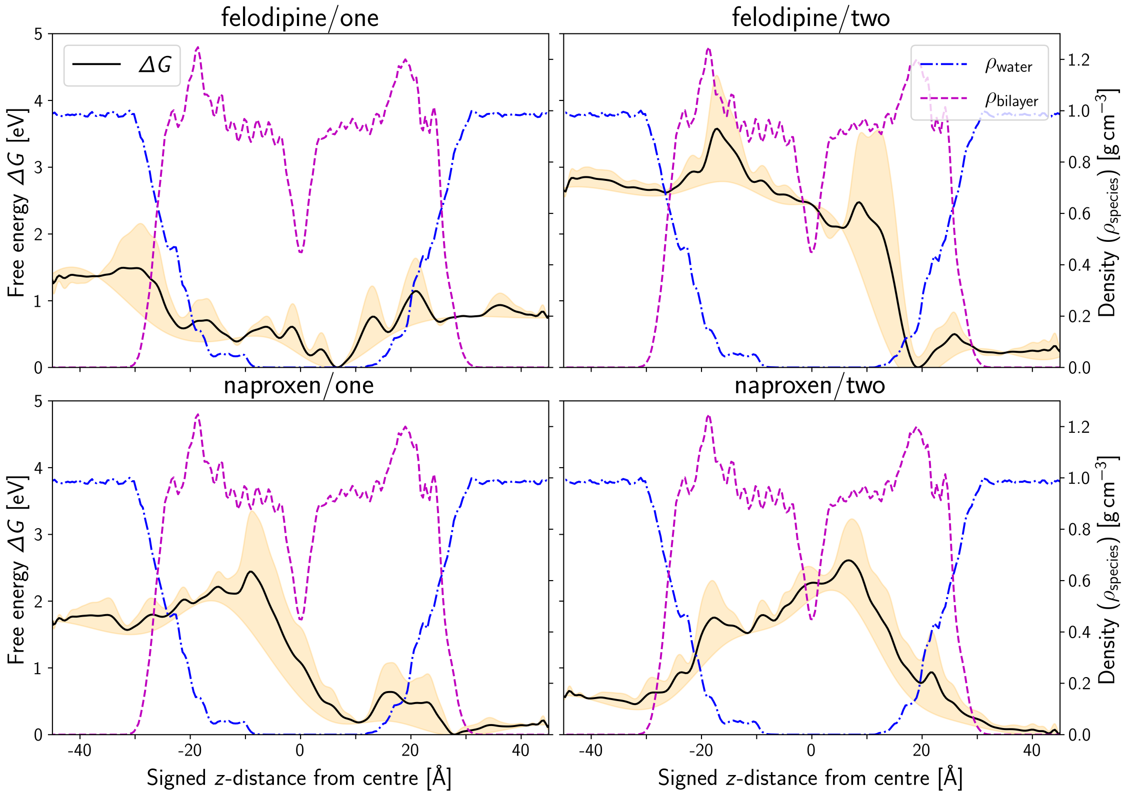


Figure 38: Results from MT of drug-bilayer systems. The top left panel shows the system with one felodipine molecule, the top right with two felodipine molecules, the bottom left with one naproxen molecule and the bottom right with two naproxen molecules. The average water and bilayer density over z are plotted with dot-dashed blue lines and dashed magenta lines respectively. The free energy surface is displayed with a solid black line, with the yellow shaded region showing the error computed by re-weighting the free energy surface and taking the difference.

There is perhaps more success in the two systems with the second unbiased drug molecule, with the naproxen free energy surface looking fairly reasonable and the right hand side of the felodipine system also looking somewhat as expected, however the error regions are still large. It is possible that the issues encountered with the MT attempts are down to the choice of the CV and a CV with more variables could have better explored the free energy landscape, in particular the orientation of the drug molecules could have been incorporated into the biasing which may have yielded better results. With the simplistic CV used, US is rather more suitable since we have a good idea of how we expect the free energy landscape to look, with respect to the z-coordinate of the drug molecules.

Figures 39 and 41 show the results of US for the systems containing felodipine and naproxen, respectively. In each the top panel shows the histograms $\Omega_i(\xi_j)$ of z-distribution of the drug molecule (with respect to the bilayer centre) for each individual simulation, see equation (66). Note that the water layer regions were run for a shorter time, therefore the lower heights for those curves. The bins used for the histograms were of equal 0.2 Å width, from −45 Å to 45 Å. The second panel shows the sum of these histograms, $\Omega(\xi_j)$. Strictly speaking, the count is plotted and not a histogram, however as the bins are of equal width, this is irrelevant. The third panel shows the estimate for the free energy as computed by the WHAM software.[245] For both systems, looking at $\Omega(\xi_j)$, which is shown in the second panels, we note that $\Omega(\xi_j) >> 0$ in every bin ξ_j. This justifies that the choice of 1 Å spacing between simulation windows with the spring constant $k = 2000\,\mathrm{kJ\,mol^{-1}}$ since the entire phase space has been sampled.

With this confirmed, we move our attention to the third panels which show the free energy. In the case of felodipine, we notice that the drug molecule is rather content to sit on the surface of the bilayer, with local free energy minima seen at the edge of the diffusion layers on both sides of the bilayer. Permeation across the hydrophilic headgroup region does require a free energy barrier to be overcome, upon which the global free energy minima are seen amongst the lipid molecules, at around 10 Å from the centre of the bilayer, see figure 40. Thus it is energetically favourable for the molecule felodipine to embed itself within such lipid membranes; this could be of great interest to the field of drug delivery. For the molecule felodipine to cross over the bilayer centre, a rather large free-energy barrier must be overcome, further exacerbated by the depth of the minima on either side. This is the least energetically favourable location for the drug molecule.

Refocussing now to examine the free energy surface for the naproxen molecule as it traverses the bilayer, we notice a rather less symmetrical free energy curve. Nonetheless, we can still identify some characteristic features of the free energy landscape. As with felodipine, naproxen finds local free energy minima at the interface between bilayer and water, indicating that naproxen is also likely to sit on the surface of the bilayer. There is then a free energy barrier to overcome for the drug molecule to permeate within the bilayer. Within the bilayer, the free energy landscape is slightly inconclusive and does not appear to have fully converged, however it appears that there could be a minimum in the centre of the bilayer. This is interesting as it is not where we saw the minima in the case

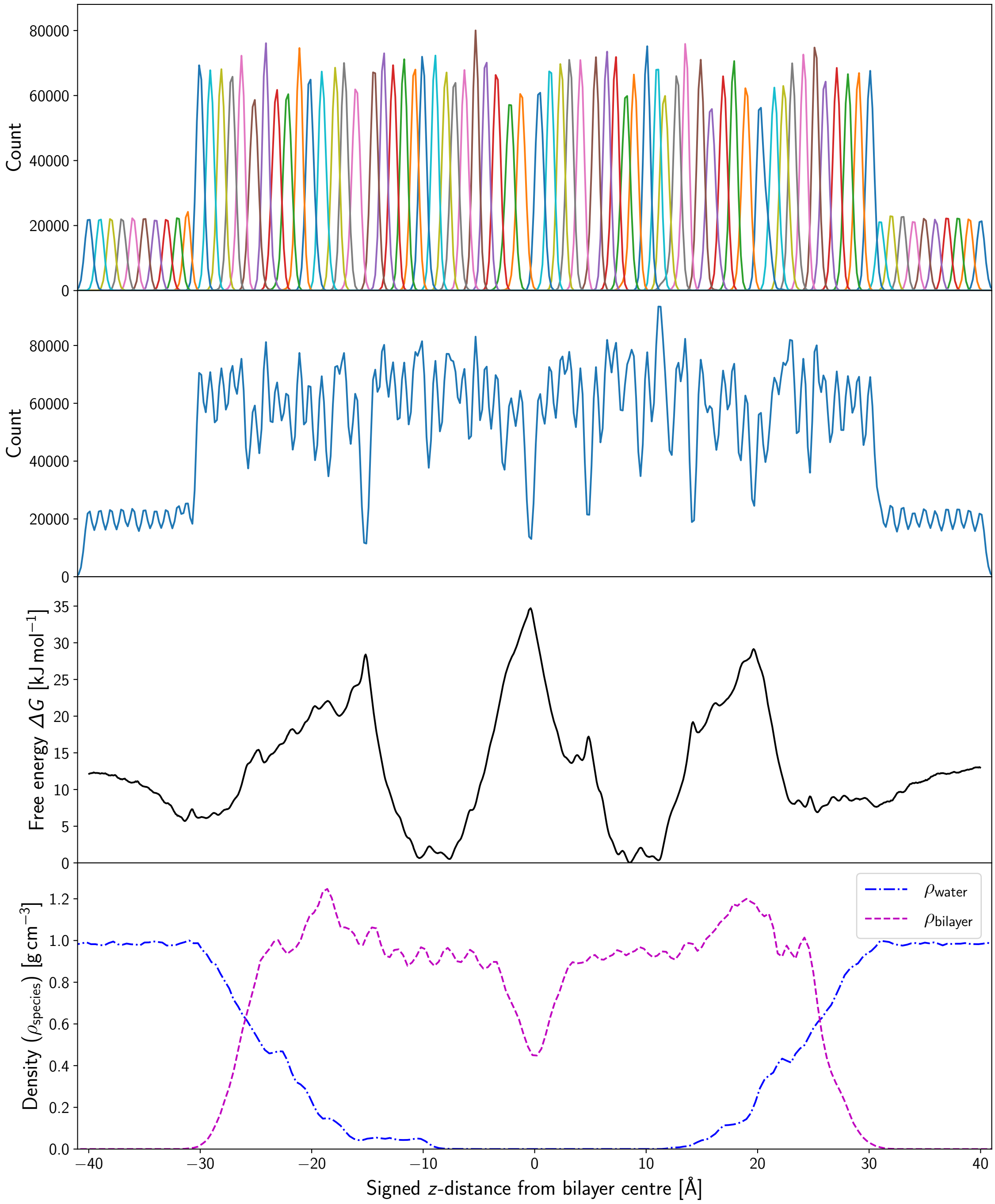


Figure 39: Results from US of felodipine-bilayer system. The top panel shows the histograms $\Omega_i(\xi_j)$ of z-distribution of the felodipine molecule (with respect to the bilayer centre) for each individual simulation, see equation (66). Note that the water layer regions were run for a shorter time, therefore the lower heights for those curves. The bins used for the histograms were of equal 0.2 Å width, from −45 Å to 45 Å. The second panel shows the sum of these histograms, $\Omega(\xi_j)$. Strictly speaking, the count is plotted and not a histogram, however as the bins are of equal width, this is irrelevant. The third panel shows the estimate for the free energy as computed by the WHAM software.[245] The fourth panel shows the density profiles for the bilayer and the water to provide a reference for the other panels. The x-scale is the same for all panels.

of felodipine; it is possible that the more linear molecule naproxen is able to embed itself in the central region of the bilayer, where the bilayer density is reduced, while the slightly less compact felodipine molecule would disrupt the hydrophobic tailgroups substantially to do the same and is instead more likely to embed amongst the tails.

It is also notable that the height of the free energy barriers are lower for naproxen than for felodipine, this suggests that it is easier for naproxen to traverse the bilayer than for felodipine. As previously mentioned, in the context of MT, taking into account the orientation of the drug molecules could have been useful for the sake of biasing — this could also have been incorporated into the CV for US, although there are trade-offs with expanding from a one-dimensional CV to one of two or three dimensions.

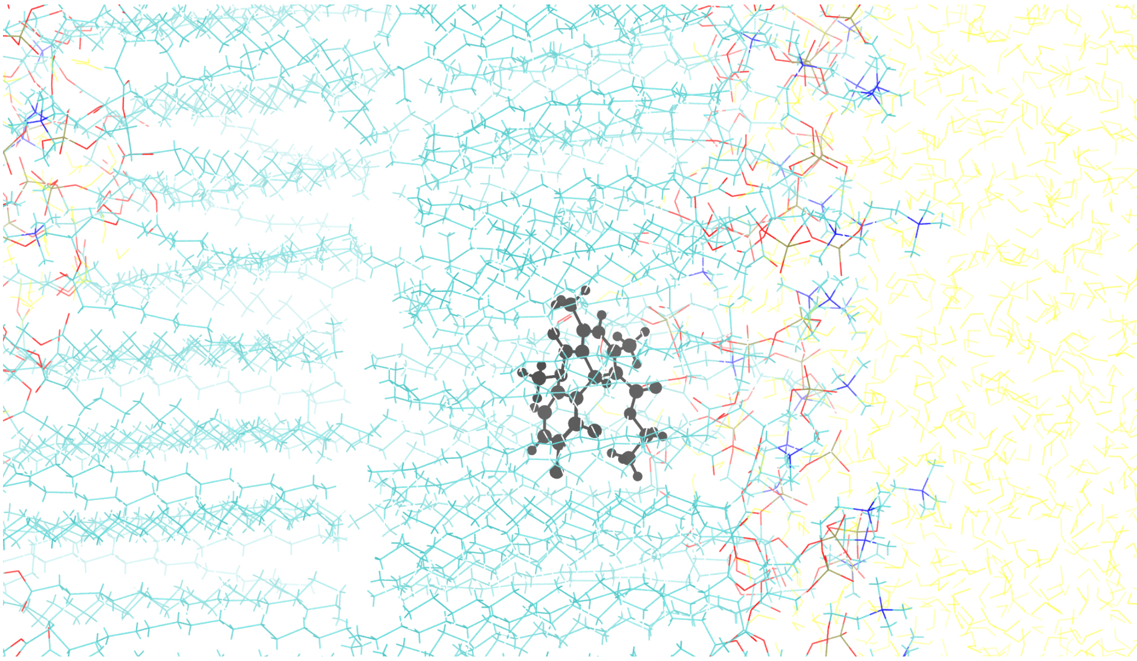


Figure 40: Visualisation of the free energy minimum for felodipine within the bilayer. Felodipine is shown all in black. DPPC is drawn as lines with carbon, oxygen, phosphorus nitrogen and hydrogen atoms coloured in cyan, red, yellow, blue and white, respectively. Water is displayed as yellow lines.

9.5 Discussion

Altogether the US simulations have provided some insight into the free energy landscape for these two molecules as they permeate across HSPC bilayers. The final free energy surface calculated for felodipine appears to have converged reasonably well and shows deep free energy minima around the base of the HSPC headgroups. For felodipine, the region within the bilayer does not appear to have converged as conclusively, however the minima seen for felodipine do not appear to exist.

This observation is in good accordance with experimental work,[246] where felodipine and naproxen loaded liposomes were prepared with HSPC and CHL. They found that the

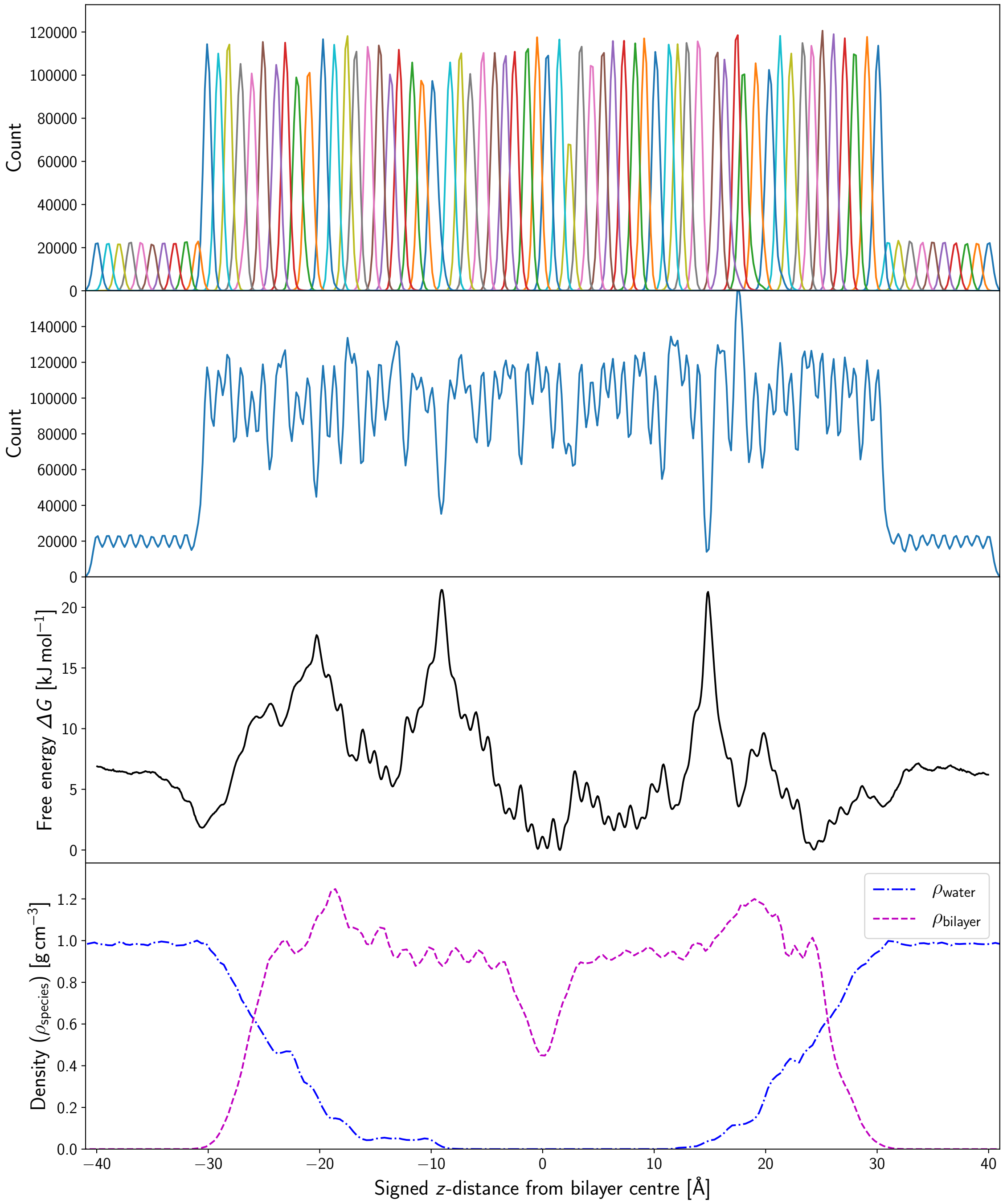


Figure 41: Results from US of naproxen-bilayer system. The top panel shows the histograms $\Omega_i(\xi_j)$ of z-distribution of the naproxen molecule (with respect to the bilayer centre) for each individual simulation, see equation (66). Note that the water layer regions were run for a shorter time, therefore the lower heights for those curves. The bins used for the histograms were of equal 0.2 Å width, from −45 Å to 45 Å. The second panel shows the sum of these histograms, $\Omega(\xi_j)$. Strictly speaking, the count is plotted and not a histogram, however as the bins are of equal width, this is irrelevant. The third panel shows the estimate for the free energy as computed by the WHAM software.[245] The fourth panel shows the density profiles for the bilayer and the water to provide a reference for the other panels. The x-scale is the same for all panels. This system does not appear to have fully converged.

liposomes containing naproxen formed larger particles than with felodipine, postulated to be caused by the fact that naproxen locates on the surface of the liposomal bilayer, whereas felodipine locates at the minima within the bilayer. Another comparison which can be drawn with the experimental results is that Cullen *et al.*[246] found a much higher amount of drug release over 24 hours for naproxen loaded liposomes than for those containing felodipine. This matches the observation from the US simulations, where the free energy barriers for permeation appear to me significantly lower for naproxen than for felodipine.

A possible way to improve these simulations would be through the use of a CV, incorporating the orientation of the drug molecules. Under this CV, MT may have yielded better results which could perhaps better elucidate the free energy landscape of the naproxen system in particular. It would also be possible to use such a CV with US simulations, although this would of course require a much higher number of "umbrellas".

Conclusions

The core question of this thesis was the following. Do lipid bilayers act as biological ice nucleating agents and if so to what extent? This is a very important question regarding our understanding of the mechanisms of biological ice nucleation, for example during cryopreservation processes. A wealth of analysis has been run for MD simulations of three classes of lipid bilayers: DPPC-CHL, DPPC-DMPI and the three asymmetric LPS membranes. Through comparison to published results regarding CHL crystals[21] and self-assembled CHL monolayers,[24] it appears that while such bilayers can and do promote ice nucleation, any such effect is far weaker than that of the most potent INAs.

Furthermore, by comparing the different bilayers studied herein, it was shown that pure DPPC, CHL and DMPI bilayers acted as slightly more potent INAs than the mixed membranes. This can be somewhat elucidated by considering the interplay between the structural and chemical properties of the bilayers. The number of hydrogen bonding sites for example is a chemical property which helps to promote ice nucleation, although the total number is less important than the accessibility of the bonding sites — a structural quality. In particular, with mixed DPPC-CHL bilayers, the CHL molecules tend to be deeply embedded into the bilayer, causing the hydroxy group, which promotes ice nucleation so actively in crystals[21] and monolayers,[24] to be inaccessible to water and therefore unable to promote ice nucleation. In addition to the computational simulations of these phospholipid bilayers, DPPC-CHL vesicles were synthesised. This confirmed the validity of the size of simulation box used and could be further extended to experimentally assess the ice nucleating ability of such vesicles, for example via µl-NIPI.

The LPS membranes also showed some potential as ice nucleating agents, with both the Re and Ra mutants facilitating the emergence of ice nuclei within the extended diffusion layers, wherein we see pockets of fully coordinated water. This may be surprising due to the low average water density in these regions, although as mentioned much of the water is in fact fully coordinated. In addition the sugar residues on the Re and Ra mutants do contain a large number of hydrogen bonding sites —potentially playing an active role in facilitating this ice nucleation. The Lipid A system with no sugars coating does not appear to have much of an ice nucleating effect —seemingly much less than any of the phospholipid membranes studied.

Some of the analysis protocols discussed herein are novel. In particular, the normalisation ϑ of the orientational order parameter θ (see section 4.4.2) from which vastly improved resolution was obtained of the distribution of orientations in a geometrically sound fashion. A method for combining order parameters was also discussed (see section 4.5) which is easy to understand from a human perspective. This is a simpler version of a dimensionality reduction compared to more complex methods which are well documented.[184,185] Many computational algorithms are also described in chapter 6 and much new code has been written to perform these calculations.

The overall narrative herein suggests that while cellular membranes can promote ice nucleation to some extent, it is probable that it is other entities within biological systems which cause the onset of ice nucleation at the relatively mild supercooling levels seen during cryopreservation. A wide array of different order parameters and properties of both the bilayers and water molecules have been explored, building a large body of data regarding the physical properties of such bilayers and interfacial water layers under supercooled conditions, as well as the interactions between bilayers and water molecules. Altogether, this thesis presents a contribution towards a greater understanding of heterogeneous ice nucleation in biological matter as well as consideration towards the structural properties of various supercooled bilayers.

Separately to the study of ice nucleation, permeation of the small drug molecules felodipine and naproxen across phospholipid membranes has been investigated, utilising the enhanced sampling techniques: MT and US. Using US, a degree of information about the free energy landscape as the two molecules traverse the bilayer was found. Both molecules were found to have free energy minima on the surface of the bilayer, suggesting they may be likely to adhere to the surface. Naproxen had a free energy minimum in the centre, between the two leaflets, while the minima for felodipine within the bilayer were found at around 10 Å from the centre. I have suggested that it may be the geometry of the molecules, in addition to their chemical properties which govern whether they are likely to embed between the leaflets or amongst the tails. The theoretical free energy landscape calculated was consistent with experimental data from collaborators; although the comparison has not been discussed within this thesis. As an avenue for further research, simulations could be run to assess the free energy landscapes for permeation of other drug molecules through bilayers using the same protocols as detailed in chapter 9.

Bibliography

[1] P. Gallo, K. Amann-Winkel, C. A. Angell, M. A. Anisimov, F. Caupin, C. Chakravarty, E. Lascaris, T. Loerting, A. Z. Panagiotopoulos, J. Russo, J. A. Sellberg, H. E. Stanley, H. Tanaka, C. Vega, L. Xu and L. G. M. Pettersson, *Chem. Rev.*, 2016, **116**, 7463–7500.

[2] D. A. Knopf, P. A. Alpert and B. Wang, *ACS Earth Sp. Chem.*, 2018, **2**, 168–202.

[3] H. C. Price, K. J. Baustian, J. B. McQuaid, A. Blyth, K. N. Bower, T. Choularton, R. J. Cotton, Z. Cui, P. R. Field, M. Gallagher, R. Hawker, A. Merrington, A. Miltenberger, R. R. Neely, S. T. Parker, P. D. Rosenberg, J. W. Taylor, J. Trembath, J. Vergara-Temprado, T. F. Whale, T. W. Wilson, G. Young and B. J. Murray, *J. Geophys. Res. Atmos.*, 2018, **123**, 2175–2193.

[4] A. Bogdan, *J. Phys. Chem. A*, 2018, **122**, 7777–7781.

[5] A. Sanchez-Marroquin, O. Arnalds, K. J. Baustian-Dorsi, J. Browse, P. Dagsson-Waldhauserova, A. D. Harrison, E. C. Maters, K. J. Pringle, J. Vergara-Temprado, I. T. Burke, J. B. McQuaid, K. S. Carslaw and B. J. Murray, *Sci. Adv.*, 2020, **6**, 8137–8161.

[6] E. S. Thomson, D. Weber, H. G. Bingemer, J. Tuomi, M. Ebert and J. B. Pettersson, *Sci. Rep.*, 2018, **8**, 1–9.

[7] R. D. Hercamp, *SAE Trans.*, 1999, **108**, 234–242.

[8] L. Eickhoff, K. Dreischmeier, A. Zipori, V. Sirotinskaya, C. Adar, N. Reicher, I. Braslavsky, Y. Rudich and T. Koop, *J. Phys. Chem. Lett.*, 2019, **10**, 966–972.

[9] T. H. Jang, S. C. Park, J. H. Yang, J. Y. Kim, J. H. Seok, S. Park, C. W. Choi, S. R. Lee and J. Han, *Integr. Med. Res.*, 2017, **6**, 12–18.

[10] D. E. Pegg, *Methods Mol. Biol.*, 2007, **368**, 39–57.

[11] K. W. Yong, W. K. Z. Wan Safwani, F. Xu, W. A. B. Wan Abas, J. R. Choi and B. Pingguan-Murphy, *Biopreserv. Biobank.*, 2015, **13**, 231–239.

[12] A. A. Mandawala, S. C. Harvey, T. K. Roy and K. E. Fowler, *Theriogenology*, 2016, **86**, 1637–1644.

[13] L. E. McGann, H. Yang and M. Walterson, *Cryobiology*, 1988, **25**, 178–185.

[14] D. Gao and A. K. Critser, *ILAR J.*, 2000, **41**, 187–196.

[15] A. Soni and G. N. Patey, *J. Phys. Chem. C*, 2021, **125**, 26927–26941.

[16] Q. Xu, H. Wang, J. Wu and Z. Zhang, *Cryst. Growth Des.*, 2021, **21**, 4354–4361.

[17] M. A. Wilson and A. Pohorille, *J. Am. Chem. Soc.*, 1994, **116**, 1490–1501.

[18] P. B. Moore, C. F. Lopez and M. L. Klein, *Biophys. J.*, 2001, **81**, 2484–2494.

[19] M. Stępniewski, A. Bunker, M. Pasenkiewicz-Gierula, M. Karttunen and T. Róg, *J. Phys. Chem. B*, 2010, **114**, 11784–11792.

[20] M. Pasenkiewicz-Gierula, K. Baczynski, M. Markiewicz and K. Murzyn, *Biochim. Biophys. Acta - Biomembr.*, 2016, **1858**, 2305–2321.

[21] G. C. Sosso, T. F. Whale, M. A. Holden, P. Pedevilla, B. J. Murray and A. Michaelides, *Chem. Sci.*, 2018, **9**, 8077–8088.

[22] E. Mocé, E. Blanch, C. Tomás and J. Graham, *Reprod. Domest. Anim.*, 2010, **45**, 57–66.

[23] I. Massie, C. Selden, H. Hodgson and B. Fuller, *Tissue Eng. Part C. Methods*, 2011, **17**, 765–774.

[24] G. C. Sosso, P. Sudera, A. T. Backes, T. F. Whale, J. Fröhlich-Nowoisky, M. Bonn, A. Michaelides and E. H. Backus, *Chem. Sci.*, 2022, **13**, 5014–5026.

[25] T. J. McIntosh, *Biochim. Biophys. Acta - Biomembr.*, 1978, **513**, 43–58.

[26] N. Kučerka, S. Tristram-Nagle and J. F. Nagle, *Biophys. J.*, 2006, **90**, L83.

[27] M. Belička, Y. Gerelli, N. Kučerka and G. Fragneto, *Soft Matter*, 2015, **11**, 6275–6283.

[28] C. Hofsäß, E. Lindahl and O. Edholm, *Biophys. J.*, 2003, **84**, 2192–2206.

[29] S. Leekumjorn and A. K. Sum, *Biochim. Biophys. Acta - Biomembr.*, 2007, **1768**, 354–365.

[30] T. Róg, M. Pasenkiewicz-Gierula, I. Vattulainen and M. Karttunen, *Biochim. Biophys. Acta - Biomembr.*, 2009, **1788**, 97–121.

[31] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell and R. W. Pastor, *J. Phys. Chem. B*, 2010, **114**, 7830–7843.

[32] T. B. Stanishneva-Konovalova and O. S. Sokolova, *Comput. Theor. Chem.*, 2015, **1058**, 61–66.

[33] Y. Wang, P. Gkeka, J. E. Fuchs, K. R. Liedl and Z. Cournia, *Biochim. Biophys. Acta - Biomembr.*, 2016, **1858**, 2846–2857.

[34] G. M. Cooper, *The Cell: A Molecular Approach*, Sinauer Associates, Sunderland, MA, USA, 2nd edn., 2000.

[35] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts and P. Walter, *Molecular Biology of the Cell*, Garland Science, New York, NY, USA, 6th edn., 2015.

[36] P. Yeagle, *The Membranes of Cells*, Academic Press, Amsterdam, Netherlands, 3rd edn., 2016.

[37] S. J. Cox, Z. Raza, S. M. Kathmann, B. Slater and A. Michaelides, *Faraday Discuss.*, 2013, **167**, 389–403.

[38] Y. Bi, R. Cabriolu and T. Li, *J. Phys. Chem. C*, 2016, **120**, 1507–1514.

[39] J. D. Bernal and R. H. Fowler, *J. Chem. Phys.*, 1933, **1**, 515–548.

[40] L. Pauling, *J. Am. Chem. Soc.*, 1935, **57**, 2680–2684.

[41] V. F. Petrenko and R. W. Whitworth, *Physics of Ice*, Oxford University Press, Oxford, United Kingdom, 1st edn., 1999.

[42] G. Tammann, *Ann. Phys.*, 1900, **307**, 1–31.

[43] P. W. Bridgman, *Proceedings of the American Academy of Arts and Sciences*, 1912, **47**, 441–558.

[44] L. S. Whatley and A. Van Valkenburg, in *Advances in High Pressure Research (Vol. 1)*, ed. R. S. Bradley, Academic Press, London, United Kingdom, 1966, ch. High pressure optics, pp. 327–371.

[45] K. Yamamoto, *Jpn. J. Appl. Phys.*, 1980, **19**, 1841–1845.

[46] I. M. Chou and H. T. Haselton, *Rev. High Press. Sci. Technol. No Kagaku To Gijutsu*, 1998, **7**, 1132–1134.

[47] C. G. Salzmann and B. J. Murray, *Nat. Mater.*, 2020, **19**, 13066.

[48] G. C. Sosso, J. Chen, S. J. Cox, M. Fitzner, P. Pedevilla, A. Zen and A. Michaelides, *Chem. Rev.*, 2016, **116**, 7078–7116.

[49] A. Määttänen, H. Vehkamäki, A. Lauri, S. Merikallio, J. Kauhanen, H. Savijärvi and M. Kulmala, *J. Geophys. Res. E Planets*, 2005, **110**, 1–12.

[50] R. P. Sear, *J. Phys. Condens. Matter*, 2007, **19**, 033101.

[51] V. G. Baidakov and A. O. Tipeev, *J. Chem. Phys.*, 2012, **136**, 74510.

[52] J. D. Hoffman, *J. Chem. Phys.*, 1958, **29**, 1192–1193.

[53] W. Pan, A. B. Kolomeisky and P. G. Vekilov, *J. Chem. Phys.*, 2005, **122**, 174905.

[54] T. Kovács and H. K. Christenson, *Faraday Discuss.*, 2012, **159**, 123–138.

[55] J. D. Atkinson, B. J. Murray and D. O'Sullivan, *J. Phys. Chem. A*, 2016, **120**, 6513–6520.

[56] E. Sanz, C. Vega, J. R. Espinosa, R. Caballero-Bernal, J. L. Abascal and C. Valeriani, *J. Am. Chem. Soc.*, 2013, **135**, 15008–15017.

[57] Z. A. Kanji, L. A. Ladino, H. Wex, Y. Boose, M. Burkert-Kohn, D. J. Cziczo and M. Krämer, *Meteor. Mon.*, 2017, **58**, 1.1–1.33.

[58] A. Reinhardt and J. P. Doye, *J. Chem. Phys.*, 2014, **141**, 84501.

[59] S. Duane, A. D. Kennedy, B. J. Pendleton and D. Roweth, *Phys. Lett. B*, 1987, **195**, 216–222.

[60] M. Fitzner, G. C. Sosso, S. J. Cox and A. Michaelides, *J. Am. Chem. Soc.*, 2015, **137**, 13658–13669.

[61] G. Vali, *J. Atmos. Sci.*, 1971, **28**, 402–409.

[62] T. F. Whale, B. J. Murray, D. O'Sullivan, T. W. Wilson, N. S. Umo, K. J. Baustian, J. D. Atkinson, D. A. Workneh and G. J. Morris, *Atmos. Meas. Tech.*, 2015, **8**, 2437–2447.

[63] P. Mazur, *J. Gen. Physiol.*, 1963, **47**, 347–369.

[64] P. Mazur, *Cell Biophys.*, 1990, **17**, 53–92.

[65] K. Muldrew and L. E. McGann, *Biophys. J.*, 1994, **66**, 532–541.

[66] J. E. Lovelock, *Proc. R. Soc. Lond. B. Biol. Sci.*, 1957, **147**, 427–433.

[67] A. M. Karow and W. R. Webb, *Cryobiology*, 1965, **2**, 99–108.

[68] C. Polge, A. U. Smith and A. S. Parkes, *Nature*, 1949, **164**, 666.

[69] D. E. Pegg, *Semin. Reprod. Med.*, 2002, **20**, 5–13.

[70] Z. Hubálek, *Cryobiology*, 2003, **46**, 205–229.

[71] J. E. Lovelock and M. W. Bishop, *Nature*, 1959, **183**, 1394–1395.

[72] S. J. Slichter, M. Jones, J. Ransom, I. Gettinger, M. K. Jones, T. Christoffel, E. Pellham, S. L. Bailey, J. Corson and D. Bolgiano, *Transfus. Med. Rev.*, 2014, **28**, 212–225.

[73] M. M. Harding, L. G. Ward and A. D. Haymet, *Eur. J. Biochem.*, 1999, **264**, 653–665.

[74] M. K. Chattopadhyay, *Resonance*, 2007, **12**, 25–30.

[75] M. I. Gibson, *Polym. Chem.*, 2010, **1**, 1141–1152.

[76] D. E. Mitchell, A. E. R Fayter, R. C. Deller, M. Hasan, J. Gutierrez-Marcos and M. I. Gibson, *Mater. Horizons*, 2019, **6**, 364.

[77] A. Trounson and L. Mohr, *Nature*, 1983, **305**, 707–709.

[78] J. Konc, K. Kanyó, R. Kriston, B. Somoski and S. Cseh, *Biomed Res. Int.*, 2014, **2014**, 307268.

[79] R. Riggs, J. Mayer, D. Dowling-Lacey, T. F. Chi, E. Jones and S. Oehninger, *Fertil. Steril.*, 2010, **93**, 109–115.

[80] H. Rozati, T. Handley and C. Jayasena, *J. Clin. Med.*, 2017, **6**, 89.

[81] R. G. Bunge and J. K. Sherman, *Nature*, 1953, **172**, 767–768.

[82] J. K. Sherman, *Fertil. Steril.*, 1963, **14**, 49–64.

[83] W. H. Perloff, E. Steinberger and J. K. Sherman, *Fertil. Steril.*, 1964, **15**, 501–504.

[84] A. Z. Szell, R. C. Bierbaum, W. B. Hazelrigg and R. J. Chetkowski, *J. Assist. Reprod. Genet.*, 2013, **30**, 743–744.

[85] I. Ullah, R. B. Subbarao and G. J. Rho, *Biosci. Rep.*, 2015, **35**, 191.

[86] C. Miyagi-Shiohira, K. Kurima, N. Kobayashi, I. Saitoh, M. Watanabe, Y. Noguchi, M. Matsushita and H. Noguchi, *Cell Med.*, 2015, **8**, 3.

[87] K. W. Yong, B. Pingguan-Murphy, F. Xu, W. A. B. Wan Abas, J. R. Choi, S. Z. Omar, M. A. N. Azmi, K. H. Chua and W. K. Z. Wan Safwani, *Sci. Rep.*, 2015, **5**, 9596.

[88] F. Wang, T. Yasuhara, T. Shingo, M. Kameda, N. Tajiri, W. J. Yuan, A. Kondo, T. Kadota, T. Baba, J. T. Tayra, Y. Kikuchi, Y. Miyoshi and I. Date, *BMC Neurosci.*, 2010, **11**, 1–9.

[89] M. Nakhaeifard, M. H. G. Kashani, I. Goudarzi and A. Rezaei, *Cell J.*, 2018, **20**, 348–354.

[90] L. Danielyan, S. Beer-Hammer, A. Stolzing, R. Schäfer, G. Siegel, C. Fabian, P. Kahle, T. Biedermann, A. Lourhmati, M. Buadze, A. Novakovic, B. Proksch, C. H. Gleiter, W. H. Frey and M. Schwab, *Cell Transplant.*, 2014, **23**, 123–139.

[91] T. Ma, K. Gong, Q. Ao, Y. Yan, B. Song, H. Huang, X. Zhang and Y. Gong, *Cell Transplant.*, 2013, **22**, 113–126.

[92] H. Yang, H. Yang, Z. Xie, L. Wei and J. Bi, *PLoS One*, 2013, **8**, e69129.

[93] M. Nakano, K. Kubota, E. Kobayashi, T. S. Chikenji, Y. Saito, N. Konari and M. Fujimiya, *Sci. Rep.*, 2020, **10**, 1–15.

[94] K. R. Prabakar, J. Domínguez-Bendala, R. Damaris Molano, A. Pileggi, S. Villate, C. Ricordi and L. Inverardi, *Cell Transplant.*, 2012, **21**, 1321–1339.

[95] I. O. Unsal, Z. Ginis, F. A. Pinarli, A. Albayrak, E. Cakal, M. Sahin and T. Delibasi, *Stem Cell Rev. Reports*, 2015, **11**, 526–532.

[96] J. Liu, Q. Hu, Z. Wang, C. Xu, X. Wang, G. Gong, A. Mansoor, J. Lee, M. Hou, L. Zeng, J. R. Zhang, M. Jerosch-Herold, T. Guo, R. J. Bache and J. Zhang, *Am. J. Physiol. - Hear. Circ. Physiol.*, 2004, **287**, 501–511.

[97] P. J. Psaltis, A. C. Zannettino, S. G. Worthley and S. Gronthos, *Stem Cells*, 2008, **26**, 2201–2210.

[98] S. Bahsoun, K. Coopman and E. C. Akam, *J. Transl. Med.*, 2019, **17**, 397.

[99] S. Guven and U. Demirci, *Nanomedicine*, 2012, **7**, 1787–1789.

[100] B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, 1957, **27**, 1208–1209.

[101] M. Karplus and J. A. McCammon, *Nat. Struct. Biol.*, 2002, **9**, 646–652.

[102] S. A. Hollingsworth and R. O. Dror, *Neuron*, 2018, **99**, 1129–1143.

[103] B. Guillot, *J. Mol. Liq.*, 2002, **101**, 219–260.

[104] S. Izadi, R. Anandakrishnan and A. V. Onufriev, *J. Phys. Chem. Lett.*, 2014, **5**, 3863–3871.

[105] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, Intermol. Forces Proc. Fourteenth Jerusalem Symp. Quantum Chem. Biochem. Held Jerusalem, Isr. April 13–16, 1981, 1981, pp. 331–342.

[106] S. H. Lee and J. C. Rasaiah, *J. Chem. Phys.*, 1994, **101**, 6964–6974.

[107] S. Esteban-Martín, H. Jelger Risselada, J. Salgado and S. J. Marrink, *J. Am. Chem. Soc.*, 2009, **131**, 15194–15202.

[108] V. Molinero and E. B. Moore, *J. Phys. Chem. B*, 2009, **113**, 4008–4016.

[109] M. K. Coe, R. Evans and N. B. Wilding, *J. Chem. Phys.*, 2022, **156**, 154505.

[110] J. C. Johnston, N. Kastelowitz and V. Molinero, *J. Chem. Phys.*, 2010, **133**, 154516.

[111] E. B. Moore and V. Molinero, *Nature*, 2011, **479**, 506–508.

[112] H. Chan, M. J. Cherukara, B. Narayanan, T. D. Loeffler, C. Benmore, S. K. Gray and S. K. Sankaranarayanan, *Nat. Commun.*, 2019, **10**, 379.

[113] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.

[114] H. W. Horn, W. C. Swope, J. W. Pitera, J. D. Madura, T. J. Dick, G. L. Hura and T. Head-Gordon, *J. Chem. Phys.*, 2004, **120**, 9665–9678.

[115] J. L. Abascal and C. Vega, *J. Chem. Phys.*, 2005, **123**, 234505.

[116] J. L. Abascal, E. Sanz, R. G. Fernández and C. Vega, *J. Chem. Phys.*, 2005, **122**, 234511.

[117] D. R. Nutt and J. C. Smith, *J. Chem. Theory Comput.*, 2007, **3**, 1550–1560.

[118] U. S. Midya and S. Bandyopadhyay, *J. Phys. Chem. B*, 2014, **118**, 4743–4752.

[119] K. Thurmer and S. Nie, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 11757–11762.

[120] G. C. Sosso, G. A. Tribello, A. Zen, P. Pedevilla and A. Michaelides, *J. Chem. Phys.*, 2016, **145**, 211927.

[121] S. J. Buxton, D. Quigley and S. Habershon, *J. Chem. Phys.*, 2019, **151**, 144503.

[122] S. Fleischer and G. Brierley, *Biochem. Biophys. Res. Commun.*, 1961, **5**, 367–372.

[123] D. Chapman and S. A. Penkett, *Nature*, 1966, **211**, 1304–1305.

[124] E. Oldfield and D. Chapman, *Biochem. Biophys. Res. Commun.*, 1971, **43**, 610–616.

[125] A. M. Smondyrev and M. L. Berkowitz, *Biophys. J.*, 1999, **77**, 2075–2089.

[126] M. R. Vist and J. H. Davis, *Biochemistry*, 1990, **29**, 451–464.

[127] T. H. Huang, C. W. Lee, S. K. Das Gupta, R. G. Griffin and A. Blume, *Biochemistry*, 1993, **32**, 13277–13287.

[128] M. Shaghaghi, A. Keyvanloo, Z. Huang, F. C. Szoka and J. L. Thewalt, *Langmuir*, 2017, **33**, 14405–14413.

[129] M. Vega, L. Lurio, J. Lal, E. A. Karapetrova and E. R. Gaillard, *Phys. Chem. Chem. Phys.*, 2020, **22**, 19089–19099.

[130] J. P. Bradshaw, R. J. Bushby, C. C. Giles and M. R. Saunders, *Biochemistry*, 1999, **38**, 8393–8401.

[131] B. Payrastre, K. Missy, S. Giuriato, S. Bodin, M. Plantavid and M. P. Gratacap, *Cell. Signal.*, 2001, **13**, 377–387.

[132] W.-C. Hung, M.-T. Lee, F.-Y. Chen and H. W. Huang, *Biophys. J.*, 2007, **92**, 3960–3967.

[133] F. Leeb and L. Maibaum, *Biophys. J.*, 2018, **115**, 2179–2188.

[134] T. J. Silhavy, D. Kahne and S. Walker, *Cold Spring Harb. Perspect. Biol.*, 2010, **2**, a000414.

[135] M. Huber, C. Kalis, S. Keck, Z. Jiang, P. Georgel, X. Du, L. Shamel, S. Sovath, S. Mudd, B. Beutler, C. Galanos and M. A. Freudenberg, *Eur. J. Immunol.*, 2006, **36**, 701–711.

[136] F. Cochet and F. Peri, *Int. J. Mol. Sci.*, 2017, **18**, 2318.

[137] P. C. Hsu, B. M. Bruininks, D. Jefferies, P. Cesar Telles de Souza, J. Lee, D. S. Patel, S. J. Marrink, Y. Qi, S. Khalid and W. Im, *J. Comput. Chem.*, 2017, **38**, 2354–2363.

[138] R. Shrivastava and S. S. Chng, *J. Biol. Chem.*, 2019, **294**, 14175–14184.

[139] B. Bertani and N. Ruiz, *EcoSal Plus*, 2018, **8**, 10.1128/ecosalplus.ESP–0001–2018.

[140] J. D. Nickels and J. Katsaras, in *Membrane Hydration*, ed. E. A. Disalvo, Springer International Publishing, Switzerland, 2015, ch. Water and Lipid Bilayers, pp. 45–68.

[141] K. Åman, E. Lindahl, O. Edholm, P. Håkansson and P. O. Westlund, *Biophys. J.*, 2003, **84**, 102–115.

[142] R. A. Böckmann, A. Hac, T. Heimburg and H. Grubmüller, *Biophys. J.*, 2003, **85**, 1647–1655.

[143] S. J. McKinnon, S. L. Whittenburg and B. Brooks, *J. Phys. Chem.*, 1992, **96**, 10497–10506.

[144] D. Ehre, E. Lavert, M. Lahav and I. Lubomirsky, *Science (80-.).*, 2010, **327**, 672–675.

[145] A. Belitzky, E. Mishuk, D. Ehre, M. Lahav and I. Lubomirsky, *J. Phys. Chem. Lett.*, 2016, **7**, 43–46.

[146] S. Curland, E. Meirzadeh, H. Cohen, D. Ehre, J. Maier, M. Lahav and I. Lubomirsky, *Angew. Chemie - Int. Ed.*, 2018, **57**, 7076–7079.

[147] S. Jo, T. Kim, V. G. Iyer and W. Im, *J. Comput. Chem.*, 2008, **29**, 1859–1865.

[148] B. R. Brooks, C. L. Brooks, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch, A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York and M. Karplus, *J. Comput. Chem.*, 2009, **30**, 1545–1614.

[149] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, C. L. Brooks, A. D. MacKerell, J. B. Klauda and W. Im, *J. Chem. Theory Comput.*, 2016, **12**, 405–413.

[150] E. L. Wu, X. Cheng, S. Jo, H. Rui, K. C. Song, E. M. Dávila-Contreras, Y. Qi, J. Lee, V. Monje-Galvan, R. M. Venable, J. B. Klauda and W. Im, *J. Comput. Chem.*, 2014, **35**, 1997–2004.

[151] S. Jo, J. B. Lim, J. B. Klauda and W. Im, *Biophys. J.*, 2009, **97**, 50–58.

[152] H. J. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.

[153] E. Lindahl, B. Hess and D. van der Spoel, *J. Mol. Model.*, 2001, **7**, 306–317.

[154] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.

[155] B. Hess, C. Kutzner, D. Van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, 2008, **4**, 435–447.

[156] S. Pronk, S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. Van Der Spoel, B. Hess and E. Lindahl, *Bioinformatics*, 2013, **29**, 845–854.

[157] S. Páll, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, Solving Softw. Challenges Exascale, 2015, pp. 3–27.

[158] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindah, *SoftwareX*, 2015, **1-2**, 19–25.

[159] P. C. Hsu, F. Samsudin, J. Shearer and S. Khalid, *J. Phys. Chem. Lett.*, 2017, **8**, 5513–5518.

[160] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, *J. Comput. Chem.*, 2010, **31**, 671–690.

[161] K. Vanommeslaeghe and A. D. MacKerell, *J. Chem. Inf. Model.*, 2012, **52**, 3144–3154.

[162] K. Vanommeslaeghe, E. P. Raman and A. D. MacKerell, *J. Chem. Inf. Model.*, 2012, **52**, 3155–3168.

[163] W. Yu, X. He, K. Vanommeslaeghe and A. D. MacKerell, *J. Comput. Chem.*, 2012, **33**, 2451–2468.

[164] I. Soteras Gutiérrez, F. Y. Lin, K. Vanommeslaeghe, J. A. Lemkul, K. A. Armacost, C. L. Brooks and A. D. MacKerell, *Bioorganic Med. Chem.*, 2016, **24**, 4812–4825.

[165] J. B. Lim, B. Rogaski and J. B. Klauda, *J. Phys. Chem. B*, 2012, **116**, 203–210.

[166] S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 952–962.

[167] B. Hess, H. Bekker, H. J. Berendsen and J. G. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.

[168] D. Bostick and M. L. Berkowitz, *Biophys. J.*, 2003, **85**, 97–107.

[169] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Academic Press, San Diego, CA, USA, 2nd edn., 2002.

[170] J. R. Espinosa, C. Vega and E. Sanz, *J. Phys. Chem. C*, 2016, **120**, 8068–8075.

[171] J. Seelig, *Q. Rev. Biophys.*, 1977, **10(3)**, 353–418.

[172] J. Seelig and N. Waespe-Šarĉević, *Biochemistry*, 1978, **17**, 3310–3315.

[173] L. S. Vermeer, B. L. De Groot, V. Réat, A. Milon and J. Czaplicki, *Eur. Biophys. J.*, 2007, **36**, 919–931.

[174] T. J. Piggot, J. R. Allison, R. B. Sessions and J. W. Essex, *J. Chem. Theory Comput.*, 2017, **13**, 5683–5696.

[175] P. J. Steinhardt, D. R. Nelson and M. Ronchetti, *Phys. Rev. B*, 1983, **28**, 784–805.

[176] W. Lechner and C. Dellago, *J. Chem. Phys.*, 2008, **129**, 114707.

[177] T. Li, D. Donadio, G. Russo and G. Galli, *Phys. Chem. Chem. Phys.*, 2011, **13**, 19807–19813.

[178] E. B. Moore, E. De La Llave, K. Welke, D. A. Scherlis and V. Molinero, *Phys. Chem. Chem. Phys.*, 2010, **12**, 4124–4134.

[179] L. A. Báez and P. Clancy, *Ann. N. Y. Acad. Sci.*, 1994, **715**, 177–186.

[180] P. M. Rodger, T. R. Forester and W. Smith, *Fluid Phase Equilib.*, 1996, **116**, 326–332.

[181] S. Parui and B. Jana, *J. Phys. Chem. B*, 2019, **123**, 811–824.

[182] K. Tay and F. Bresme, *J. Mater. Chem.*, 2006, **16**, 1956–1962.

[183] B. Glatz and S. Sarupria, *Langmuir*, 2018, **34**, 1190–1198.

[184] I. S. Dhillon and S. Sra, Proceedings of the 18th International Conference on Neural Information Processing Systems, Cambridge, MA, USA, 2005, p. 283–290.

[185] C. Cortes and V. Vapnik, *Mach. Learning*, 1995, **20**, 273–297.

[186] The PLUMED Consortium, *Nat. Methods*, 2019, **16**, 670–673.

[187] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

[188] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.

[189] G. C. Sosso, T. Li, D. Donadio, G. A. Tribello and A. Michaelides, *J. Phys. Chem. Lett.*, 2016, **7**, 2350–2355.

[190] L. Lupi, A. Hudait, B. Peters, M. Grünwald, R. Gotchy Mullen, A. H. Nguyen and V. Molinero, *Nature*, 2017, **551**, 218–222.

[191] M. Fitzner, G. C. Sosso, S. J. Cox and A. Michaelides, *Proc. Natl. Acad. Sci.*, 2019, **116**, 2009–2014.

[192] M. Fitzner, G. C. Sosso, F. Pietrucci, S. Pipolo and A. Michaelides, *Nat. Commun.*, 2017, **8**, 2257.

[193] M. Fitzner, P. Pedevilla and A. Michaelides, *Nat. Commun.*, 2020, **11**, 4777.

[194] S. A. Pandit, S. Vasudevan, S. W. Chiu, R. J. Mashl, E. Jakobsson and H. L. Scott, *Biophys. J.*, 2004, **87**, 1092–1100.

[195] Y. Zhang, A. Lervik, J. Seddon and F. Bresme, *Chem. Phys. Lipids*, 2015, **185**, 88–98.

[196] F. Giberti, M. Salvalaglio, M. Mazzotti and M. Parrinello, *Chem. Eng. Sci.*, 2015, **121**, 51–59.

[197] F. Giberti, M. Salvalaglio, M. Parrinello and E. Zurich, *IUCrJ*, 2015, **2 (Pt2)**, 256–266.

[198] S. Buchoux, *Bioinformatics*, 2017, **33**, 133–134.

[199] T. L. Malkin, B. J. Murray, A. V. Brukhno, J. Anwar and C. G. Salzmann, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 1041–1045.

[200] Y. Bi and T. Li, *J. Phys. Chem. B*, 2014, **118**, 13324–13332.

[201] S. Le Roux and P. Jund, *Comput. Mater. Sci.*, 2010, **49**, 70–83.

[202] Sosso Group, *HIN/clathrates*, `https://github.com/gcsosso/HIN/tree/clathrates/`.

[203] G. B. Arfken, H. J. Weber and F. E. Harris, *Mathematical Methods for Physicists*, Academic Press, Waltham, MA, USA, 7th edn., 2012.

[204] P. Gasparotto, M. Fitzner, S. J. Cox, G. C. Sosso and A. Michaelides, *Nanoscale*, 2022, **14**, 4254–4262.

[205] L. R. Maki, E. L. Galyan, M.-M. Chang-Chien and D. R. Caldwell, *Appl. Microbiol.*, 1974, **28**, 456.

[206] G. G. de Araujo, F. Rodrigues, F. L. T. Gonçalves and D. Galante, *Sci. Rep.*, 2019, **9**, 7768.

[207] B. G. Pummer, H. Bauer, J. Bernardi, S. Bleicher and H. Grothe, *Atmos. Chem. Phys.*, 2012, **12**, 2541–2550.

[208] K. Dreischmeier, C. Budke, L. Wiehemeier, T. Kottke and T. Koop, *Sci. Reports 2017 71*, 2017, **7**, 1–13.

[209] E. Gute and J. P. Abbatt, *Atmos. Environ.*, 2020, **231**, 117488.

[210] E. Gute, R. O. David, Z. A. Kanji and J. P. D. Abbatt, *ACS Earth Sp. Chem.*, 2020, **4**, 2312–2319.

[211] J. Monroe, M. Barry, A. Destefano, P. A. Gokturk, S. Jiao, D. Robinson-Brown, T. Webber, E. J. Crumlin, S. Han and M. S. Shell, *Annu. Rev. Chem. Biomol. Eng.*, 2020, **11**, 523–557.

[212] A. Alizadeh, M. Yamada, R. Li, W. Shang, S. Otta, S. Zhong, L. Ge, A. Dhinojwala, K. R. Conway, V. Bahadur, A. J. Vinciquerra, B. Stephens and M. L. Blohm, *Langmuir*, 2012, **28**, 3180–3186.

[213] J. Chanda, L. Ionov, A. Kirillova and A. Synytska, *Soft Matter*, 2015, **11**, 9126–9134.

[214] B. J. Berne and R. Pecora, *Dynamic Light Scattering With Applications to Chemistry, Biology and Physics*, Dover Publications, Inc., Mineola, NY, USA, 2nd edn., 2000.

[215] Hon. J. W. Strutt, *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, 1871, **41**, 447–454.

[216] Lord Rayleigh F.R.S., *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, 1881, **12**, 81–101.

[217] Malvern Panalytical, *Dynamic Light Scattering: An Introduction in 30 Minutes*, `https://www.malvernpanalytical.com/en/learn/knowledge-center/technical-notes/TN101104DynamicLightScatteringIntroduction`, (accessed August 2022).

[218] G. Vali, *Atmos. Meas. Tech.*, 2019, **12**, 1219–1231.

[219] PubChem, *1,2-Dipalmitoylphosphatidylcholine*, `https://pubchem.ncbi.nlm.nih.gov/compound/1_2-Dipalmitoylphosphatidylcholine`, (accessed August 2022).

[220] PubChem, *Cholesterol*, `https://pubchem.ncbi.nlm.nih.gov/compound/Cholesterol`, (accessed August 2022).

[221] R. L. Brooks and A. M. Dixon, *Biochim. Biophys. Acta - Biomembr.*, 2020, **1862**, 183160.

[222] AHFS DI Monographs, *Naproxen*, `https://www.drugs.com/monograph/naproxen.html`, (accessed May 2022).

[223] AHFS DI Monographs, *Felodipine*, `https://www.drugs.com/monograph/felodipine.html`, (accessed May 2022).

[224] E. G. Lovering and K. B. McRae, *J. Pharm. Sci.*, 1973, **62**, 971–974.

[225] K. Qian, L. Stella, D. S. Jones, G. P. Andrews, H. Du and Y. Tian, *Pharmaceutics*, 2021, **13**, 889.

[226] T. Takagi, C. Ramachandran, M. Bermejo, S. Yamashita, L. X. Yu and G. L. Amidon, *Mol. Pharm.*, 2006, **3**, 631–643.

[227] A. M. Thayer, *Chem. Eng. News*, 2010, **88**, 13–18.

[228] N. J. Babu and A. Nangia, *Cryst. Growth Des.*, 2011, **11**, 2662–2679.

[229] S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469–2473.

[230] G. Tóth, *J. Phys. Condens. Matter*, 2007, **19**, 335222.

[231] K. R. Hadley and C. McCabe, *Mol. Simul.*, 2012, **38**, 671–681.

[232] F. Pietrucci and A. Laio, *J. Chem. Theory Comput.*, 2009, **5**, 2197–2201.

[233] F. Trudu, D. Donadio and M. Parrinello, *Phys. Rev. Lett.*, 2006, **97**, 105701.

[234] D. Quigley and P. M. Rodger, *J. Chem. Phys.*, 2008, **128**, 926.

[235] B. Ensing, A. Laio, F. L. Gervasio, M. Parrinello and M. L. Klein, *J. Am. Chem. Soc.*, 2004, **126**, 9492–9493.

[236] A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 12562–6.

[237] A. Laio and F. L. Gervasio, *Reports Prog. Phys.*, 2008, **71**, 126601.

[238] G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.*, 1974, **28**, 578–581.

[239] G. M. Torrie and J. P. Valleau, *J. Comput. Phys.*, 1977, **23**, 187–199.

[240] J. Kästner, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 932–942.

[241] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, *J. Comput. Chem.*, 1992, **13**, 1011–1021.

[242] M. Souaille and B. Roux, *Comput. Phys. Commun.*, 2001, **135**, 40–57.

[243] H. Müller-Krumbhaar and K. Binder, *J. Stat. Phys.*, 1973, **8**, 1–24.

[244] R. Goldberg, A. Schroeder, Y. Barenholz and J. Klein, *Biophys. J.*, 2011, **100**, 2403–2411.

[245] A. Grossfield, *WHAM: the weighted histogram analysis method, version 2.0.11*, `http://membrane.urmc.rochester.edu/wordpress/?page_id=126`, (accessed March 2022).

[246] S. Cullen, C. M. Miles, W. Gu, G. P. Andrews, Y. Tian and G. C. Sosso, unpublished work.

Appendix A

Equivalence of infinities

In section 4.4.1, it was claimed that for a homogeneous system at equilibrium, specifically a water box, infinite time, infinite space and infinite time *and* space are mathematically equivalent. To demonstrate that this is the case, consider a simulation trajectory Ψ as an $n \times k$ matrix:

$$\Psi = \begin{pmatrix} \mathbf{x_{11}} & \cdots & \mathbf{x_{1k}} \\ \vdots & \ddots & \vdots \\ \mathbf{x_{n1}} & \cdots & \mathbf{x_{nk}} \end{pmatrix}, \tag{83}$$

where n is the number of particles, k is the number of frames and $\mathbf{x_{ij}}$ is a vector containing the parameters for particle i at time j.

Consider two theoretical infinite simulations: Ψ_1 which is infinite in time and Ψ_2 which is infinite in both space and time. Similarly to equation (83), we can write:

$$\Psi_1 = \begin{pmatrix} \mathbf{x_{11}} & \mathbf{x_{12}} & \cdots \\ \mathbf{x_{21}} & \mathbf{x_{22}} & \cdots \\ \mathbf{x_{31}} & \mathbf{x_{32}} & \cdots \\ \mathbf{x_{41}} & \mathbf{x_{42}} & \cdots \end{pmatrix} \quad \text{and} \quad \Psi_2 = \begin{pmatrix} \mathbf{y_{11}} & \mathbf{y_{12}} & \mathbf{y_{13}} & \cdots \\ \mathbf{y_{21}} & \mathbf{y_{22}} & \mathbf{y_{23}} & \cdots \\ \mathbf{y_{31}} & \mathbf{y_{32}} & \mathbf{y_{33}} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}, \tag{84}$$

where $\mathbf{x_{ij}}$ and $\mathbf{y_{ij}}$ are the position vectors for the two simulations. Note that for ease of explanation, simulation Ψ_1 has 4 particles, however the same method extends to any finite N.

To prove that the two trajectories are mathematically equivalent, a bijection must be defined between the sets $\Phi_1 = \{\mathbf{x_{ij}} \mid i \in \{1, 2, 3, 4\},\ j \in \mathbb{N}\}$ and $\Phi_2 = \{\mathbf{y_{ij}} \mid i, j \in \mathbb{N}\}$, where $\mathbb{N}$ is the natural numbers. To define a bijection between these two sets, first note that we can clearly arrange the elements of Φ_1 into an single list:

$$L_1 = \{\mathbf{x_{11}},\ \mathbf{x_{12}},\ \mathbf{x_{13}},\ \mathbf{x_{14}},\ \mathbf{x_{21}},\ \mathbf{x_{22}},\ \mathbf{x_{23}},\ \ldots\}\,. \tag{85}$$

We can do the same with Φ_2 by traversing the matrix along its diagonals, starting on the top row and moving diagonally down and to the left:

$$L_2 = \{\mathbf{y_{11}},\ \mathbf{y_{12}},\ \mathbf{y_{21}},\ \mathbf{y_{13}},\ \mathbf{y_{22}},\ \mathbf{y_{31}},\ \mathbf{y_{14}},\ \ldots\}\,. \tag{86}$$

Finally, mapping the i-th element of L_1 to the i-th element of L_2 defines a bijection $f : \Phi_1 \mapsto \Phi_2$ and thus the two simulations are mathematically equivalent. By the same argument, a finite time, infinite space simulation is mathematically equivalent to a infinite time *and* space simulation, and so the three are all equivalent.

Appendix B

PLUMED input for computing largest icy cluster per frame

The largest icy cluster per frame was computed, as described in section 5.1, with the PLUMED 2.3.0 software,[186–188] as follows. First we set up the parameters for hydrogen bond identification, specifically distance cutoffs of 0.324 nm for O···O and 0.25 nm for O···H and angular cutoff of 0.2π for O−H···O. The indices specify the atomic indices of the water oxygen and hydrogen atoms for the given system.

```
HBOND_COORD SPECIES=6457-17145:4 HYDROGENS=6458-17146:4,6459-17147:4
    RCUTOO=0.324 RCUTOH=0.25 ACUT=0.20pi LABEL=hb
```

Secondly, we select all the oxygen atoms with four hydrogen bonds, in other words, every fully coordinated water molecule — a water molecule that is only partially coordinated cannot be part of a cluster unless it is on the surface.

```
MFILTER_BETWEEN DATA=hb LOWER=3.9 UPPER=4.1 SMEAR=0.0 LABEL=rsumsb
```

Next, we compute the Steinhardt vectors. The `LOCAL_Q6` function is then used to calculate the degree of local order around each atom and we filter only the atoms with `LOCAL_Q6` greater or equal to 0.45 in order to locate all the icy molecules.

```
Q6 SPECIES=rsumsb SWITCH={GAUSSIAN D_0=0.324 R_0=0.00001 D_MAX=0.3241} LABEL=q6
    LOWMEM
LOCAL_Q6 DATA=q6 SWITCH={GAUSSIAN D_0=0.324 R_0=0.00001 D_MAX=0.3241} LABEL=lq6
    LOWMEM
MFILTER_MORE DATA=lq6 SWITCH={GAUSSIAN D_0=0.45 R_0=0.00001 D_MAX=0.4501}
    LABEL=cf
```

A contact matrix is then created based on distance and this is used to compute the clustering, including surface atoms.

```
CONTACT_MATRIX ATOMS=cf WTOL=0.1 SWITCH={GAUSSIAN D_0=0.324 R_0=0.00001
    D_MAX=0.3241} LABEL=mat
```

```
DFSCLUSTERING MATRIX=mat LABEL=cls SERIAL
CLUSTER_WITHSURFACE CLUSTERS=cls RCUT_SURF=0.324 LABEL=scls SERIAL
CLUSTER_NATOMS CLUSTERS=scls CLUSTER=1 LABEL=lambda
```

Finally, the output file is written.

```
OUTPUT_CLUSTER CLUSTERS=scls CLUSTER=1 STRIDE=1 FILE=dfs_surf.dat
FLUSH STRIDE=2000
ENDPLUMED
```

Appendix C

PLUMED input for calculating SMAC parameter

The SMAC parameter, as defined in section 5.3.1, was calculated using the PLUMED 2.3.0 software,[186–188] as follows. First we define the molecular axis of each lipid molecule or tail (see section 7.2 for how these were defined for the different lipids).

```
MOLECULES ...
MOL1=6241,6305,6264
MOL2=6315,6379,6338
MOL3=6389,6453,6412
####### etc #######
LABEL=m3
... MOLECULES
```

We then calculate the SMAC parameter, we are only interested in the packing of the lipid tails, so we use a single kernel with angle 0.

```
SMAC ...
DATA=m3 LOWMEM
KERNEL1={GAUSSIAN CENTER=0 SIGMA=0.580}
SWITCH={RATIONAL R_0=0.5} MEAN MORE_THAN={RATIONAL R_0=0.6} SWITCH_COORD={EXP
    R_0=6}
LABEL=s2
... SMAC
```

Finally, we write the output file.

```
PRINT ARG=s2.* FILE=smac.dat
ENDPLUMED
```