Saponin Penetration and Interactions with Membranes

Sarai Guerrero
The University of Western Ontario

Supervisor
Karttunen, Mikko
The University of Western Ontario

Graduate Program in Chemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Master of Science
© Sarai Guerrero 2019

Follow this and additional works at: https://ir.lib.uwo.ca/etd

Part of the Lipids Commons, Other Chemistry Commons, and the Physical Chemistry Commons

Recommended Citation
https://ir.lib.uwo.ca/etd/6675

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.
Abstract

The saponin QS21 has been used and studied for years as an adjuvant agent to improve vaccines for both humans and animals. However, how the saponin interacts with the membrane and itself remains poorly understood. We studied the QS21-A majority isomer using all-atom classical molecular dynamics simulations in model bilayers composed of cholesterol and either dioleoyl-phosphatidylcholine (DOPC) or dipalmitoyl-phosphatidylcholine (DPPC). As this is one of the few computational studies on QS21-A, we had very few resources to compare our results to. Nevertheless, we were able to gain insight into possible configurations that QS21-A takes after penetration, the effect of the bilayer on QS21-A micelle formation, as well as hydrogen bonding and radius of gyration. We hope that our results will help with the development of better drugs and vaccines, but more work is needed to verify these results and understand both isomers.

Keywords

QS21, molecular dynamics simulations, saponins, bilayer, mechanism of action, cholesterol, DPPC, DOPC, radius of gyration, hydrogen bonding
Summary for Lay Audience

Vaccines are used to protect cells from various ailments with the most common one being viral in nature. To improve these vaccines, compounds called adjuvant agents are used. Saponins are a family of molecules known for creating soap-like foam in water that have been used to fulfil this role. The most studied saponin is a molecule called QS21, which has been used since its discovery. Despite its power, the use of QS21 is limited since we lack a full understanding on how it works, the limited supply, storage issues, and noted toxicity of the molecule. We do not even know if there is a difference between the two types of QS21. Due to this, we employed a computer-based method known as molecule dynamics to study the more common type of QS21 to gain insight to how it works with the bilayer. Using bilayers made from cholesterol and phosphatidylcholines, we looked at the shape and structure a group of QS21 molecules take as well as how the molecule bonded with the bilayer. We learnt about the structure that the QS21 molecules took, how the bilayer possibly changed the interactions QS21 made, and a possible method of how QS21 penetrates the bilayer. We hope that this information will help the development of better drugs and vaccines, but studies with the molecule must continue to gain a deeper understanding of QS21 and to verify these results.
Dedications and Acknowledgments

I would also like to thank the members of the SoftSimu research group for helping me with my questions as I completed my research. Whether it was about GROMACS, Linux, or sharing anger over Cedar’s crashes, they were always there. It was hard going from my analytical experimental background to a theoretical and computational project, but they tried to make the transition as smooth as possible. Trying not to forget anyone, there was Hendrik, Vaibhav, David/Angel. Cecilia, Alireza, David Jr./David Colin, Shahin, Rafael, Conrard, Mahmood, and Elham (I hope I didn’t forget anyone). Extra gratitude to Shahin who helped with running the raw data through the programs while I got ahead in writing, and Hendrik who helped me through the GROMACS tutorials when they didn’t make sense or were missing key steps. Finally, I would like to thank Mikko for supervising me during my Master’s degree, and Dr. Tomasz Rog for creating three of the systems I worked with.

I would like to acknowledge everyone who had the unique blessing of being dragged into listening to the issues and problems that came up during my time. Their names are too numerous to count, but they all listened to me while I was working through possible solutions and explanations. A double thank you to Robert Harbottle who was enlisted for this duty and was never able to escape it because he worked on the same floor as me and his work area was right now to the restroom and water fountain.
Table of Contents

Abstract ......................................................................................................................... ii
Summary for Lay Audience .......................................................................................... iii
Dedications and Acknowledgments................................................................................ iv
Table of Contents........................................................................................................... v
List of Tables ................................................................................................................ viii
List of Figures ............................................................................................................... ix
Abbreviations ............................................................................................................... xvii

Chapter 1 Introduction .................................................................................................. 1
  1.1 The Cell.................................................................................................................. 1
  1.2 The Cellular Membrane ....................................................................................... 2
    1.2.1 Membrane Models ....................................................................................... 2
    1.2.2 Lipids ........................................................................................................... 5
    1.2.3 Cholesterol .................................................................................................. 7
  1.3 The Use of Vaccines .............................................................................................. 8
  1.4 Saponins ................................................................................................................ 9
    1.4.1 QS21’s Background ..................................................................................... 10
    1.4.2 Currenting Suggestions for QS21’s Mechanism of Action............................ 13
  1.5 Computational Studies of QS21 ............................................................................. 14
  1.6 Thesis Structure .................................................................................................... 14
  1.7 References ............................................................................................................ 15

Chapter 2 The Methods Used ....................................................................................... 21
  2.1 Molecular Dynamic Simulations .......................................................................... 21
    2.1.1 The Methods Types of Molecular Dynamic Simulations ............................... 21
    2.1.2 Limitations of Molecular Dynamic Simulations ........................................... 22
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.2</td>
<td>Integrators</td>
<td>23</td>
</tr>
<tr>
<td>2.3</td>
<td>Constraints</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>Periodic Boundary Conditions</td>
<td>25</td>
</tr>
<tr>
<td>2.5</td>
<td>Potentials, Force Fields, and Water Models</td>
<td>26</td>
</tr>
<tr>
<td>2.6</td>
<td>Dealing with Long-Range Coulomb Interactions</td>
<td>31</td>
</tr>
<tr>
<td>2.6.1</td>
<td>The Problem with Truncating the Coulomb Potential</td>
<td>31</td>
</tr>
<tr>
<td>2.6.2</td>
<td>Ewald Summation and Particle-Mesh Ewald</td>
<td>33</td>
</tr>
<tr>
<td>2.7</td>
<td>Ensembles and How to Make Them</td>
<td>36</td>
</tr>
<tr>
<td>2.8</td>
<td>References</td>
<td>38</td>
</tr>
<tr>
<td>3.1</td>
<td>System Information and Creation</td>
<td>40</td>
</tr>
<tr>
<td>3.1.1</td>
<td>Simulation Parameters</td>
<td>40</td>
</tr>
<tr>
<td>3.1.2</td>
<td>Creation Methods and Created Systems</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Data Analysis</td>
<td>42</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Visual or Qualitative Inspection</td>
<td>42</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Area Per Lipid, Membrane Area, and Membrane Thickness</td>
<td>42</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Lipid Order Parameter</td>
<td>44</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Hydrogen Bonding</td>
<td>45</td>
</tr>
<tr>
<td>3.2.5</td>
<td>Radius of Gyration</td>
<td>46</td>
</tr>
<tr>
<td>3.3</td>
<td>References</td>
<td>47</td>
</tr>
<tr>
<td>4.1</td>
<td>The Data Used</td>
<td>50</td>
</tr>
<tr>
<td>4.2</td>
<td>Membrane Properties</td>
<td>51</td>
</tr>
<tr>
<td>4.3</td>
<td>Reduction Effect and Membrane-QS21-A Interactions</td>
<td>58</td>
</tr>
<tr>
<td>4.4</td>
<td>QS21-A Properties</td>
<td>67</td>
</tr>
</tbody>
</table>
Chapter 5 Concluding Remarks and Future Work ........................................... 73

5.1 Gathered Conclusions .................................................................................. 73

5.2 Future Work ................................................................................................. 74

5.2.1 Testing the Effect of QS21 and Its Isomers ............................................ 74

5.2.2 Testing the Effect of Cholesterol ............................................................ 75

5.2.3 Mixed Lipid Bilayer .................................................................................. 75

5.2.4 Free Energy Difference .......................................................................... 76

5.3 References .................................................................................................... 76

Appendix A: Verlet and Leapfrog Integrator .................................................... 77

Appendix B: Ewald Summation and Particle-Mesh Ewald ............................... 79

Appendix C: Berendsen and Nosé-Hoover Thermostat ................................... 86

Appendix D: Berendsen and Parrinello-Rahman Barostat ............................... 88

Appendix E: Additional Figures from Chapter 4 .............................................. 91

Curriculum Vitae ............................................................................................... 111
List of Tables

Table 3.1: Number of Atoms and Molecules in the Studied Systems .........................42

Table 3.2: Dimensional Information for Systems After Pre-Equilibration ..................42

Table 4.1: Different Membrane Properties for the Systems ...................................57

Table 4.2: Summation of the Number of H-bonds Between QS21-A and the Bilayer .......61

Table 4.3: $R_g$ and H-bond Average for the Systems .............................................69
List of Figures

Figure 1.1: A comparison of the two general categories of cells eukaryote and prokaryotes. The difference is based on the presence of membrane-bound organelles. This image was released to the public domain. Source: https://web.archive.org/web/20130507094245/http://www.ncbi.nlm.nih.gov/About/primer/genetics_cell.html ................................................................. 1

Figure 1.2: The fluid mosaic model of the cellular membrane. While lipids do make up a large faction of the membrane, other molecules are also needed for it to be functional such as cholesterol and various proteins. This image was released to the public domain. Source: wikipedia.org/wiki/Cell_membrane ................................................................. 3

Figure 1.3: A timeline of how the membrane model has developed over time. The reference numbers match the numbering found in Section 1.7 ................................................................. 4

Figure 1.4: Two of the three membrane lipid families, phosphoglyceride/glycerophospholipids and sphingolipids, that compose the membrane with some examples. The phosphoglyceride examples include phosphotidylethanolamine (PE), phosphotidylcholine (PC), phosphotidylserine (PS), phosphatidyglycerol (PG), and phosphatidylinositol (PI). The example of sphingolipids is sphingomyelin (SM). These lipids tend to have a charged or polar hydrophilic headgroup while the tails are non-polar and hydrophobic. This amphiphilic nature leads to the headgroups face the aqueous environment while the tails remain protected in the hydrophobic membrane core. Cholesterol can be found in Figure 1.6. ................................................................. 5

Figure 1.5: Different structures that lipids can take in aqueous environments. The white bends are the hydrophilic heads while the yellow strings are the hydrophobic tails. Due to the amphiphilic nature of these lipids, the formation of these structures is spontaneous and based on the lipid concentration found in the environment. Since the headgroups and the tails prefer different environments, the end structure maximizes water-headgroup interactions while minimizing water-tail interactions. There is a general trend that more cylindrical lipids will form bilayers. This image was released to the public domain with the placement of the bilayer altered to adjacent of micelle. Source: wikipedia.org/wiki/Lipid ................................................................. 6
Figure 1.6: Cholesterol shown using two different representations. A type of steroid alcohol or sterol, (A) is the 2D chemical structure while (B) shows a side view of a 3D model. The presence of the one flat side called the α-face and the “rough” side called the β-face is essential for cholesterol to fulfill its role in the cell membrane.

Figure 1.7: The molecular structure of QS21 with the terminal sugar in the tetrasaccharide chain region defining the isomer with (A) D-apiose defining the majority QS21-Api isomer, and (B) D-xylose defining the minority QS21-Xyl isomer. These isomers exist in a 2:1 ratio in the Quillaja Saponaria Molina tree. The molecule is colour-coded to identify the different components of the saponin. Starting from the left; D-xlose is yellow, D-glucuronic acid is pink, D-galactose is purple, triterpene quillaic acid is cyan, D-fucose is green, L-rhamnose is mauve, D-apiose is red, 3,5-dihydroxy-6-methyl-octanoic acid is orange, and L-arabinose is grey. This colour code is used throughout the thesis.

Figure 2.1: A comparison of the different types of simulations discussed in the order of (A) ab initio, (B) classical MD, and (C) coarse-grained MD. The ab initio method accounts for the electronic variables, so the electrons are shown for each atom. The classical method does not consider these variables, so the bonds are shown as lines. Coarse-grained methods make pseudo-atoms from combining multiple atoms together.

Figure 2.2: A visualization of how the Leapfrog integrator works to find the positions and velocities for a simulation. The positions, shown in purple, are found at every whole-time unit. The velocities, shown in dashed blue, are found at every half time unit. The positions also require the use of the velocities to find them.

Figure 2.3: An illustration of how periodic boundary conditions creates an infinite system by repeating the main box with its purple atoms in all directions. When atoms in the main box with the purple atoms move, this motion is reflected in all replicated boxes with its orange atoms with the same force and direction.

Figure 2.4: The four different components of the total potential shown graphical on example molecules. (A) shows bond stretching between two atoms and angle bending between three
atoms. (B) shows the torsion between four atoms and the electrostatic interactions between the non-bonded atoms.

Figure 2.5: A model of how neighbours are found in molecular dynamic simulations for non-bonded interactions. With periodic bounty conditions in place, a $r_{cut-off}$ is used that is smaller than half the box length. All the atoms within this distance are used to calculate interactions. This leads to $O(N^2)$ computation cost, but it is possible to reduce this value by using something like Verlet neighbour listing. In this method, a layer of extra atoms from $r_{cut-off}$ to $r_{Verlet}$ is stored in a Verlet neighbour list. By storing these atoms to call upon them later on, the computation cost is reduced to $O(N\log N)$.

Figure 2.6: A comparison different types of force fields in the order of (A) all-atom, (B) united-atom, and (C) coarse-grained. All-atom force fields account for every atom in a simulation, but at the cost of increased computational demand. United-atom force fields combine some atoms (typically methyl and methylene hydrogens with the corresponding carbons) together as to minimize the amount of detail lost while benefiting from the decreased computational cost. Coarse-grained force fields combine many atoms together and makes pseudo-atoms that have lost the finer details of the other force fields but is able to take on larger simulations.

Figure 2.7: A comparison of three types of water models to (A) a water molecule with electrons. The water models are named based on the number of “sites” and degrees of freedom. A 3-site model (B) with two hydrogens and one oxygen atom. A 4-site model (C) add a dummy atom that represent the lone pairs’ negative charge. A 5-site model (D) also use a dummy atom, but there is a second one so that both lone pairs are accounted for.

Figure 2.8: Electrostatic or dielectric screening of two charges with and without a solvent with a large dielectric constant. When $q_1$ and $q_2$ are within the distance $r_{seen}$ (A), the two are able to “see” the charge from the other. By pulling these charges to $r_{hidden}$ (B), they are “hidden” from each other and long-range electrostatic interactions will not occur. By using a solvent with a large dielectric constant $\varepsilon$ (C), $q_1$ and $q_2$ will only see $1/\varepsilon$ of the charge they normally would at $r_{seen}$. This is the underlying theory behind how truncation would still be usable for electrostatic interactions.
Figure 2.9: An example of the correlation shells for the different states of matter as (A) to (C) with their radial distribution functions graphed in (D) at zero temperature. Each state of matter will produce a unique peak pattern due to the amount of diffusion that each undergo. (A) Solids have a repeating lattice structure and will have (A) a repeating pattern in the peaks shown as the blue line. (B) Liquids are more diffuse and will have the peaks die off shown as a dashed red line. (C) Gases suffer the most diffusion and will normally have only one peak, shown as a dotted green line.

Figure 2.10: A mock NaCl unit cell which has the length $L$ and a volume of $L^3$ with three sides shown. Overall, NaCl is neutral but is composed of cations and anions. The electrostatic potential of ionic crystals such as this led to problems as, since direct summation of the Coulomb interactions led to conditional convergence only. Ewald solved this paradox by developing a method known as the Ewald summation shown in Ref. [31].

Figure 2.11: How the point charges, shown in blue, horizontal lines, of a system are found. These are assigned screening charges of equal but opposite magnitude, shown as orange peaks. Compensating charges, shown as purple peaks, are equal but opposite magnitude to the screening charges. When these factors are added together, the screening and compensating charges cancel out and the point charges remain.

Figure 2.12: The idea behind how the Particle-Mesh Ewald speeds up the Ewald summation. The charges or particles of the system are assigned to a grid or mesh and Poisson’s equation for this grid is solved. This leads to the potentials required to solve for the forces needed. After this, the forces are interpolated back to the charges or particles.

Figure 2.13: The three common types of ensembles/conditions used represented with beakers (A) microcanonical or NVE, (B) canonical or NVT, and (C) isobaric-isothermal or NPT or Gibbs. Thermostats, which are needed for constant temperature, are shown as a heat bath. Barostats, needed for constant pressure, are shown as a piston. The number of particles is kept constant in all of them with (A) and (B) also keeping the volume constant. (C) trades in constant volume for constant pressure.
Figure 3.1: The PC lipid DPPC as seen in the FATSLiM software. After treating the lipids as beads, FATSLiM will find the area per lipid using Voronoi cells, membrane area by area summation, and membrane thickness by measuring the distance between two reference points. 44

Figure 3.2: Solving the angle $\theta$ between the bilayer normal $\mathbf{v}_{\text{normal}}$ and carbon-hydrogen vector $\mathbf{v}_{\text{C-H}}$ on a DPPC tail. This value is used to find the lipid order parameter and thus the lipid order of the lipid tails. ........................................................................................................................................44

Figure 3.3: How the carbons were numbered by GROMACS for analysis. ..................................................45

Figure 3.4: The criterion for hydrogen bonding between the hydrogen atom, shown in yellow, of a donor, shown in red, and an acceptor, shown in purple. The distance $r_{HB}$ and the angle $\alpha_{HB}$ were selected to be 0.35 nm and 30° respectively. The $r_{HB}$ corresponds to the first minimum of the RDF of the TIP3P water model........................................................................................................46

Figure 4.1: Images of (A) 1-DPPC at 900 ns and (B) 50-DPPC at 700 ns without water. ..........51

Figure 4.2: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 1-DPPC. ........................................................................................................................................52

Figure 4.3: Images of (A)1-DOPC at 900 ns, (B) 5-DPPC-SI at 700 ns, (C) 10-DOPC-SI at 700ns, and (D) 50-DOPC at 700 ns all with water removed.................................................................53

Figure 4.4: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 1-DOPC...........................................................................................................................................55

Figure 4.5: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 10-DOPC-SI. ........................................................................................................................................56

Figure 4.6: The terminal D-Api of QS21-Api at (A) 180 ns, (B) 190 ns, and (C) 200 ns as it penetrated the bilayer into the Chol-lipid backbone area in the 50-DOPC system. (D) shows that D-Api had penetrated the bilayer to reach the Chol and form H-bonds with terminal sugars being shown using their GROMACS names. .........................................................................................................................58
Figure 4.7: The effect of reducing the number of QS21-A molecules from 50 to 10 on (A) the area per lipid, and (B) the membrane thickness. After reduction, shown as a green arrow, the area per lipid increased while thickness deceased by a noticeable factor to values that were very close to what is in Table 4.1.

Figure 4.8: The shape of the penetrating QS21-A molecule in 10-DOPC-SI where (A) is a sideview with the bilayer shown with DOPC in iceblue lines, Chol in black lines, nitrogens are blue atoms, and phosphorus are tan atoms, (B) is the sideview without the bilayer, and (C) is the bird’s eye view without the bilayer.

Figure 4.9: The H-bonds between QS21-A and DOPC for (A) 1-DOPC and (B) 5-DOPC-SI over time.

Figure 4.10: The H-bonds between QS21-A and DOPC for (A) 10-DOPC-SI and (B) 50-DOPC over time.

Figure 4.11: The H-bonds between QS21-A and Chol for (A) 5-DOPC-SI, (B) 10-DOPC-SI, (C) 50-DOPC over time.

Figure 4.12: The QS21-A micelle of the 50-DPPC system using a bird’s eye view clearly showing the micelle is not spherical and possibly needing more time to reach its equilibrium shape.

Figure 4.13: The different QS21-A micelles of (A) 10-REF, (B) 50-REF, (C) 5-DOPC-SI, (D) 10-DOPC-SI, and (E) 50-DOPC. The purple atoms are the K⁺ counter-cations used in the systems. (B) and (E) indicate the systems with 50 QS21-A molecules needed more time to reach their equilibrium shape.

Figure 4.14: A comparison of the number of QS21-A to QS21-A H-bonds over the number of QS21-A molecules in the system in the presence and absence of a bilayer. Since the lines are overlapping with minute differences, the graph implies that bilayers do not affect the QS21-A to QS21-A H-bonding.

Figure 4.15: A comparison of the average number of QS21-A to water H-bonds over the number of QS21-A molecules in the system in the presence and absence of a bilayer. While there is a
difference in the values, this could have been caused by the random placement of the QS21-A molecules in the systems.

Figure E.1: Images showing the 1-DPPC and 50-DPPC systems after pre-equilibration and at 900 ns and 700 ns respectively.

Figure E.2: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 50-DPPC.

Figure E.3: Images showing 1-DOPC, 50-DOPC, 5-DOPC-SI, and 10-DOPC-SI systems after pre-equilibration and at 900 ns and 700 ns respectively.

Figure E.4: (A) the area per lipid values for (B) membrane thickness, and (C) lipid order parameter for 5-DOPC-SI.

Figure E.5: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 50-DOPC.

Figure E.6: The H-bonds between QS21-A and (A) the lipids and (B) the Chol of the 50-DPPC system over time.

Figure E.7: The penetrated QS21-A molecule (A) at the start of penetration 200 ns, (B) at 700 ns in the 10-DOPC-SI system, and (C) with nitrogen, blue atoms, phosphorus, tan atoms, lipid backbone oxygens, gold atoms, and Chol headgroup oxygen, red atoms, shown to illustrate the location of the molecule.

Figure E.8: Images showing the 10-REF and 50-REF systems after pre-equilibration and at 800 ns and 900 ns respectively. The end point image of 50-REF has periodic conditions on for a better image of the micelle.

Figure E.9: The $R_g$ graphs of the micelles in (A) 10-REF and (B) 50-REF.

Figure E.10: The $R_g$ graphs of the single QS21-A molecules in (A) 1-DOPC and (B) 1-DPPC. This was done to observe how a single QS21-A changed over time.

Figure E.11: The $R_g$ graphs of the micelles in (A) 5-DOPC-SI and (B) 10-DOPC-SI.
Figure E.12: The \( R_g \) graphs of the micelles in (A) 50-DOPC and (B) 50-DPPC. ........................102

Figure E.13: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 1-DOPC system. ..................................................................................................................103

Figure E.14: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 5-DOPC system. ..................................................................................................................104

Figure E.15: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 10-DOPC-SI system. ..................................................................................................................105

Figure E.16: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-DOPC system. ..................................................................................................................106

Figure E.17: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 1-DPPC system. ..................................................................................................................107

Figure E.18: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-DPPC system. ..................................................................................................................108

Figure E.19: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 10-REF system. ..................................................................................................................109

Figure E.20: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-REF system. ..................................................................................................................110
Abbreviations

APL – area per lipid
D-Glc – D-glucose or dexter-glucose
D-Gal – D-galactose or dexter-galactose
D-GlcA – D-glucuronic acid or dexter-glucuronic acid
D-Fuc – D-fucose or dexter-fucose
D-Xyl – D-xylose or dexter-xylose
L-Rha – L-rhamnose or laevus-rhamnose
L-Ara – L-arabinose or laevus-arabinose
D-Api – D-apiose or dexter-apiose
DMOS – 3,5-dihydroxyl-6-methyl octanoic acid
QS21-A – QS21-Api or QS21’s D-apiose isomer
QS21-X – QS21-Xyl or QS21’s D-xylose isomer
MOA – mechanism of action
Chol - cholesterol
PC – phosphatidylcholines
DPPC – dipalmitoyl- phosphatidylcholine
DOPC – dioleoyl- phosphatidylcholine
K+ – potassium cation
MD – molecular dynamic
GROMACS – Groningen machine for chemical simulations
LINCS – linear constraint solver
PBC – periodic boundary conditions
RDF – radial distribution function
PME – particle-mesh Ewald
NVE – microcanonical ensemble
NVT – canonical ensemble
NPT – Gibbs ensemble
OPLS-AA – optimized potentials for liquid simulations all-atom
TIP3P – transferable intermolecular potential with 3 points
VMD – visual molecular dynamics
FATSLiM – fast analysis toolbox for simulations of lipid membranes
H-bond – hydrogen bond
COM – centre of mass
Chapter 1
Introduction

1.1 The Cell

The cell is a collection of cytoplasm, enzymes, amino acids, organelles, and other components that act as the source of structure and function for life on Earth and are divided into eukaryotes and prokaryotes [1]. Shown in Figure 1.1, prokaryotes have their DNA floating around with the rest of the components while eukaryotes have their DNA within an organelle called the nucleus.

The cellular membrane serves as an important barrier that separates the cell’s contents from the outside. Its extraordinary properties of flexibility, self-sealing, and selective permeability permits for cells to function as they do [2, 3]. Their flexibility lets the cell change shape if needed while self-sealing prevents fatal leaks while allowing fusing. Their permeability lets nutrients enter the cell while stopping unnecessary molecules.

Figure 1.1: A comparison of the two general categories of cells eukaryote and prokaryotes. The difference is based on the presence of membrane-bound organelles. This image was released to the public domain. Source: https://web.archive.org/web/20130507094245/http://www.ncbi.nlm.nih.gov/About/primer/genetics_cell.html.
1.2 The Cellular Membrane

1.2.1 Membrane Models

The structure, or the model, of the membrane has come together thanks to the work of many scientists. The first step was taken by R. Hooke, of Hooke’s Law fame, working with cork tissue back in 1665 [4,5]. He noticed that the tissue had small compartments within it, and named these compartments cells which has stuck since. Next was the discovery of the membrane or plasma membrane itself by K. Nägeli in 1855 [6]. At this time Nägeli was working with plant cells when he noticed the presence of a semipermeable layer around them. He called this layer the plasma membrane and published his findings with his student C. Cramer [6].

The years 1895 to 1902 saw a great leap in our collective understanding of the membrane when C. Overton, hailed as laying the foundation of membrane science, worked on measuring the membrane permeability of several different compounds [7,8]. He was studying this problem at a time when scientists were still debating whether a physical barrier existed around cells, and his work showed that the cell was capable of allowing some molecules inside while others were rejected. From his work, Overton also developed the suspicion that lipids made up the membrane. E. Gorter and F. Grendel applied Langmuir’s method of calculating the area per lipid (APL) in 1925 to the lipid extracts they calculated from erythrocyte (blood cell) membranes [9]. They used a Langmuir trough and calculated that the lipids were wrapped around the cell in a layer that was two molecules thick which they called the lipid bilayer. Unfortunately, R. Bar et al. found that this was a mistake in 1966 [9, 10]. The group found that the amount that Gorter and Grendel extracted was only enough to cover the cell fully in a monolayer or cover half as a bilayer. Gorter and Grendel also made several other mistakes during their experiment, but these cancelled out and allowed them to reach the correct conclusion. The year 1966 was also the point when membrane models were starting to reach the point of understanding how the membrane was constructed [11]. The scientific community understood that the lipid bilayer was needed and that charges were located on the hydrophilic side. One model was by J. Singer who believed that the lipid bilayer was also composed of proteins that could either pass through the full bilayer or half of it. This was the foundation of the fluid mosaic model of 1972 by J. Singer and G. Nicolson [12].
Figure 1.2: The fluid mosaic model of the cellular membrane. While lipids do make up a large faction of the membrane, other molecules are also needed for it to be functional such as cholesterol and various proteins. This image was released to the public domain. Source: wikipedia.org/wiki/Cell_membrane.

The illustration in Figure 1.2 demonstrates how the fluid mosaic model took into account the area that proteins and cholesterol took up in the membrane and noted that the membrane was a fluid-like phase. The presence of cholesterol and proteins explained why Gorter and Grendel were only able to extract enough lipids to cover half the cell in a bilayer. Singer and Nicolson also made a distinction between the types of proteins found in the membrane. The first type are the peripheral proteins which are associated only to one side of the membrane. The second type are the integral proteins that are anchored in the membrane and pass through the whole width. The fluid mosaic model maximized the hydrophobic and hydrophilic interactions in the membrane using thermodynamic considerations and made it a useful tool in explaining many phenomena observed in membranes [11].

After the fluid mosaic model, the next major step came in 1997 when K. Simons and E. Ikonen published their hypothesis that the membrane contains lipid rafts [13]. These rafts are composed of cholesterol and mostly sphingolipids with some phosphoglycerides or glycerophospholipids. Rafts are thought to help with intracellular signaling and host select proteins. The last point of membrane model history we will cover is from 2003 by G. Vereb et al. when they proposed the dynamically structured mosaic model [14]. They noted that the old model could not account for certain features that were being discovered through experiments. To address this, they stated that the membrane has non-random co-distributed patterns of proteins.
that are supported to form small clusters at the molecular level and large clusters at the sub-micrometer level. This history is summarized in Figure 1.3.

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>1665</td>
<td>The Cell</td>
</tr>
<tr>
<td></td>
<td>- Hooke found small compartments in cork tissue he was working with and called these compartments cells [4,5]</td>
</tr>
<tr>
<td>1855</td>
<td>The Plasma Membrane</td>
</tr>
<tr>
<td></td>
<td>- Nägeli discovered a semi-permeable layer around plant cells and called it the plasma membrane [6]</td>
</tr>
<tr>
<td>1895 - 1902</td>
<td>Overton’s Membrane Permeability Studies</td>
</tr>
<tr>
<td></td>
<td>- Overton measured the membrane permeability of different compounds and developed an early theory about the membrane structure [7, 8]</td>
</tr>
<tr>
<td>1925</td>
<td>The Lipid Bilayer</td>
</tr>
<tr>
<td></td>
<td>- Gorter and Grendel extracted lipids from the membrane and concluded that these molecules covered the cell in a double molecule layer they called the lipid bilayer [9]</td>
</tr>
<tr>
<td>1966</td>
<td>Correcting Gorter and Grendel and Early Models</td>
</tr>
<tr>
<td></td>
<td>- Bar et al. found that there was only enough lipids extracted to form a monolayer or a bilayer layer half the size reported by Gorter and Grendel [9, 10]</td>
</tr>
<tr>
<td></td>
<td>- Early membrane models were starting to piece together how the membrane worked [11]</td>
</tr>
<tr>
<td>1972</td>
<td>The Fluid Mosaic Model</td>
</tr>
<tr>
<td></td>
<td>- Singer and Nicolson published their model that described the membrane as a fluid-like structure composed of lipids, integral proteins, and peripheral proteins [12]</td>
</tr>
<tr>
<td>1997</td>
<td>Lipid Rafts</td>
</tr>
<tr>
<td></td>
<td>- Simons and Ikonen published their theory that cholesterol, sphingolipids, and phosphoglycerides/glycerophospholipids form functional domains in the membrane with greater lateral ordering [13]</td>
</tr>
<tr>
<td>2003</td>
<td>The Dynamically Structured Mosaic Model</td>
</tr>
<tr>
<td></td>
<td>- Vereb et al. released a modified model that gives a more realistic membrane model called the dynamically structure structured mosaic model [14]</td>
</tr>
</tbody>
</table>

**Figure 1.3:** A timeline of how the membrane model has developed over time. The reference numbers match the numbering found in Section 1.7.

Given the different components of a membrane, we will only cover the pieces that are used directly in this thesis. First, we will examine the lipids before going into cholesterol in a different section.
1.2.2 Lipids

Figure 1.4: Two of the three membrane lipid families, phosphoglyceride/glycerophospholipids and sphingolipids, that compose the membrane with some examples. The phosphoglyceride examples include phosphatidylethanolamine (PE), phosphatidylcholine (PC), phosphatidylserine (PS), phosphatidylglycerol (PG), and phosphatidylinositol (PI). The example of sphingolipids is sphingomyelin (SM). These lipids tend to have a charged or polar hydrophilic headgroup while the tails are non-polar and hydrophobic. This amphiphilic nature leads to the headgroups face the aqueous environment while the tails remain protected in the hydrophobic membrane core. Cholesterol can be found in Figure 1.6.

While there are many sub-types of molecules that fall into the family of lipids, the ones found in the membrane are generally divided into three types [15]. There are the phosphoglycerides or glycerophospholipids, sphingolipids, and cholesterol. For the rest of this thesis we will be using the term phospholipids for the phosphoglycerides or glycerophospholipids family for ease. The phospholipid and sphingolipid types of membrane lipids, shown in Figure 1.4, have the same general structure of a charged or polar hydrophilic headgroup and one or two non-polar hydrophobic tails that can be found in other lipids. These tails are typically referred to as sn1 and sn2 with sn1 being the preferentially saturated tail and
sn2 the preferentially unsaturated tail [16]. This also means that most cell membrane lipids are amphiphilic, or have both a hydrophilic and hydrophobic part, in nature. This results in the spontaneous formation of different structures when placed in an aqueous environment to ensure that the tails remain in a protected core as shown in Figure 1.5. The resulting structure will maximize interactions between the aqueous environment and the headgroups while it minimizes interactions between the environment and the hydrophobic tails. It will also depend on the lipid concentration and the shape of the lipid itself.

Figure 1.5: Different structures that lipids can take in aqueous environments. The white bends are the hydrophilic heads while the yellow strings are the hydrophobic tails. Due to the amphiphilic nature of these lipids, the formation of these structures is spontaneous and based on the lipid concentration found in the environment. Since the headgroups and the tails prefer different environments, the end structure maximizes water-headgroup interactions while minimizing water-tail interactions. There is a general trend that more cylindrical lipids will form bilayers. This image was released to the public domain with the placement of the bilayer altered to adjacent of micelle. Source: wikipedia.org/wiki/Lipid.

Another point to address with membrane lipids is that phospholipids are the most abundant class with most being derivatives of glycerol 3-phosphate. These tails are connected to the phosphate group through ester linkages and can have a variety of lengths and units of unsaturation. The other class, the sphingolipids, are derivatives of the amino alcohol group with a hydrocarbon chain sphingosine. This chain acts as one of the tails while the other is a fatty acid chain connected via an amide bond. Generally, the more cylindrical the lipid the more likely it will form a bilayer [17].
1.2.3 Cholesterol

Cholesterol is one of the major constituents of animal cell membranes [18]. It can be found in the membranes at concentrations of 20-70% depending on the membrane’s location [19, 20] and it has an effect on many proteins hosted in the membrane from receptors to ion channels [21, 22]. For a more thorough picture on this molecule, we review some of the structure, effects, and requirements that cholesterol has. All three of these points play a role in how cholesterol interacts with membranes.

Shown in Figure 1.6 (A), cholesterol is a steroid alcohol with a rigid steroid ring connected to a hydroxyl group and a short hydrocarbon chain. Since this is a type of lipid, the hydroxyl group is often referred to as the headgroup and the hydrocarbon chain as the tail. From Figure 1.6 (B) we see that the 3D structure of cholesterol is rather interesting. As the two methyl substituents are located only on one side, cholesterol ends up with an asymmetrical structure. The side without the methyl groups is smooth and is called the α-face. The side with the methyl groups is “rough” and is called the β-face. This naming convention is also carried over to the substituents as they also carry the α and β naming scheme [18].

Cholesterol has been found to have an effect on membrane structure: when a membrane is in the fluid state, the physiologically important state, cholesterol promotes ordering and rigidity in the membrane. However, if a membrane is in the gel state then cholesterol promotes disordering and fluidity [23]. Thus, cholesterol has the ability to regulate membrane fluidity and phase behavior [23-27]. Cholesterol also has an effect on the mechanistic properties of the membrane. With cholesterol present, membranes typically become mechanically stronger [28-31] and more resistant against permeation of small molecules [32-38].

There is also the matter of how strong these effects are, and that depends on the lipids that compose the membrane itself. Listing the main factors, the tail length [39-41], headgroup structure [42-48], backbone structure [49], and units of unsaturation [39, 40, 50-52] all will determine the effect cholesterol. As membranes are composed of many types of lipids, the ratio of lipids in the general area around the cholesterol is another factor.
Figure 1.6: Cholesterol shown using two different representations. A type of steroid alcohol or sterol, (A) is the 2D chemical structure while (B) shows a side view of a 3D model. The presence of the one flat side called the α-face and the “rough” side called the β-face is essential for cholesterol to fulfill its role in the cell membrane.

1.3 The Use of Vaccines

The membrane can protect the insides of the cell from the outside environment, but it cannot protect it from everything. One of these threats are called viruses and they are deceivingly simple. Viruses are the simplest organism with a composition of just a maximum of three parts; a protein coat called the capsid, their genetic material in the capsid, and a possible viral envelope that covers the capsid [53]. Yet their reproduction cycle is vicious as viruses infiltrate other cells and get the native mechanisms to assemble new viruses. These new viruses will then kill the cell after completion [53]. Another issue is that viruses are continuously evolving and adapting to changes in the environment, requiring the defenses against them to follow the same pattern. This is best illustrated with the influenza virus that changes so quickly that a new influenza vaccine needs to be developed twice in a single year [54]. Vaccines are a concoction of biological agents including either active or inactive pathogens that stimulate the adaptive immune response [55].
Vaccination is an effective and inexpensive method of preventing viral infections and improving the health of the human population. Active vaccines use live but weakened pathogens, but they can also cause disease depending on the nature of the pathogen [56]. Inactive vaccines have dead pathogens or some component, like the antigens, that is responsible for the infection/their pathological effects and are safer than active vaccines, but suffer from poor immunogenic effects. In other words, the ability of inactive vaccines to cause an immune response after injection is weak and requires some other agent to boost the effect. This is especially true with subunit or antigen vaccines and we must use an adjuvant agent.

An adjuvant agent is a molecule or a set of molecules that will modify the effect(s) of another agent and in subunit vaccines this will either increase the efficacy of the vaccine or the number of antibodies produced [57, 58]. However, the effect is more complex as the adjuvant agent also defines the nature and the potency of the subunit vaccine itself [59]. It is therefore completely possible to develop two different vaccines by just changing which adjuvant is used in the formulation. Adjuvants are divided into classes based on which type of immune response they trigger after injection; Th1 agents and Th2 agents. Th1 agents will cause a Th1 or cell-mediated proinflammatory response which is responsible for destroying intracellular parasites and causing autoimmune responses otherwise known as tissue damage and cell death. This type of immunity appeared through evolution long before Th2 or humoral anti-inflammatory immunity did and thus finding adjuvants that trigger it is easy and are often derived from pathogens themselves. A Th2 response follows a Th1 response to counter-balance the effects and repairs the damage from the latter. Too much or too strong of a Th2 response leads to allergic responses so a balance is needed for a truly healthy response. However, since Th2 appeared later the number of adjuvant agents that elicit only this response is small and might not work for all patients equally [59]. Even rarer than this is an adjuvant agent that, while labelled as either Th1 or Th2, is capable of eliciting both responses with strong potency.

1.4 Saponins

Saponins are surface-active glycosides that are found in plants and some marine organism [60]. They are divided into two classes based on their aglycone or sapogenin backbone [61, 62]. Triterpene saponins have a thirty carbon aglycone, and steroid saponins have a twenty-seven
carbon aglycone [63, 64]. These classes are also divided further into sub-classes. Steroid saponins are divided into just the two sub-classes, spirostane and furostane. Triterpene saponins are more widespread than steroid saponins, and triterpene saponins are divided into more sub-classes but two examples are lupine and oleanane [61]. All saponins have hydrophilic sugar moieties linked to the aglycone through glycosidic linkages [65, 66] or ester linkages [64]. Common sugar moieties among saponins are D-glucose (D-Glc), D-galactose (D-Gal), D-glucuronic acid (D-GlcA), D-fucose (D-Fuc), D-xylose (D-Xyl), L-rhamnose (L-Rha), and L-arabinose (L-Ara) [61]. The native purpose of a saponin is to act as the pseudo-immune response after being released from the vacuole when plants are infected or wounded [67]. Saponin composition, or which sugars are connected to the aglycone, will determine the bioactivity of the molecule [61, 68]. All saponins share a few properties such as how they produce soap-like foam when mixed in an aqueous solution and they have an amphiphilic nature. Other properties include their ability to perform emulsification [69], hemolytic capacities on erythrocytes [62, 70], and medical attributes such as antimicrobial, antiparasitic, antifungal, or anticancer properties [62]. It should be noted that, predictably, not all saponins will work as adjuvant agents and nor are the effects equal. Quillaja saponins are noted as inducing faster and stronger primary immune responses than saponins from other sources [71].

1.4.1 QS21’s Background

The most hydrophobic of all the Quillaja Saponaria Molina saponins [72], QS21 is named after the host tree and the identifying reverse-phase high performance liquid chromatography (RP-HPLC) peak location [72, 73]. A single molecule weighs 1990 Da with a molecular formula of C_92H_148O_46 [74] and it stands as one of the most studied saponins. It is categorized as a Th1 adjuvant agent, but QS21 is one of the rare adjuvants that produces both Th1 and Th2 immune responses and also stimulates cytotoxic T lymphocyte production [73, 75]. Its usage began in the 1970’s when K. Dalsgaard used the unpurified Molina extract called Quil A in human and animal vaccines [76, 77]. Over time, the components of Quil A were purified and tested individually until QS21’s potency was discovered [72, 73]. Since then, QS21 has been found beneficial in other applications such as the treatment of cancer [78-81].

As a bidesmosidic triterpene saponin, QS21 has two chains connected to the aglycone triterpene quillaic acid with a sugar grouping at the “top” [73]. The quillaic acid is the location of
an aldehyde and a hydroxyl group, where the importance of the former will be explained later. The chains are sugars bonded using the normal linkage types noted in Section 1.4. A list of QS21’s components is D-Xyl, D-Gal, D-GlcA, D-Fuc, L-Rha, L-Ara, D-apiose (D-Api), and 3,5-dihydroxy-6-methyl octanoic acid (DMOS) [72, 73]. The structure is shown in Figure 1.7 in a colour-coded fashion. This code is used throughout the thesis in any QS21 image. The D-GlcA carries a negative charge on the ester at physiological pH, making the entire molecule have an overall negative charge. We found that most QS21 articles use a common nomenclature to separate the parts of QS21 into four domains. The first is the branched trisaccharide domain composed of D-Xyl, D-GlcA, and D-Gal. The second is the triterpene domain which only accounts for the triterpene aglycone. The linear tetrasaccharide domain has D-Fuc, L-Rha, D-Xyl, and D-Api. Lastly, the acyl chain domain has two copies of DMOS and the L-Ara. However, QS21 does have limitations that cause problems.

When working with QS21 there are some limitations that need to be considered beforehand. First QS21 suffers from a limited supply, and thus an increased cost, which is caused by four different sources. One source is that QS21 refers to a 2:1 mixture of the two QS21 isomers QS21-Api (QS21-A) and QS21-Xyl (QS21-X) [82]. The difference is based on whether the linear tetrasaccharide domain ends with a D-Api or a D-Xyl respectively. In this thesis, we will use QS21 when addressing both isomers and the isomer names when addressing the isomers. The second source is that the purification process for QS21 from Quil A is difficult and results in a low-yielding end result [82]. The third source is that the QS21 level in Molina trees, even those growing in the same environment, will not be the same and QS21 will be a rather small fraction even before purification through this number is also changing [83]. Fourthly, the growth and collection of Molina tree bark is rightfully monitored and limited to prevent the destruction of the species.
Figure 1.7: The molecular structure of QS21 with the terminal sugar in the tetrasaccharide chain region defining the isomer with (A) D-apiose defining the majority QS21-Api isomer, and (B) D-xylose defining the minority QS21-Xyl isomer. These isomers exist in a 2:1 ratio in the Quillaja Saponaria Molina tree. The molecule is colour-coded to identify the different components of the saponin. Starting from the left; D-xlose is yellow, D-glucuronic acid is pink, D-galactose is purple, triterpene quillaic acid is cyan, D-fucose is green, L-rhamnose is mauve, D-apiose is red, 3,5-dihydroxy-6-methyl-octanoic acid is orange, and L-arabinose is grey. This colour code is used throughout the thesis.

The next limitation is that QS21 is a toxic molecule. When QS21 is used as an adjuvant agent in a vaccine or injection, it will cause swelling and redness at the injection site [77]. Another side-effect is that the patient may suffer from influenza-like symptoms afterwards [77]. Removing the acyl chain from QS21 could, in theory at least, decrease toxicity but this path is impossible due to the following limitation of QS21: should the acyl chain be removed, a process that occurs spontaneously via ester linkage hydrolysis in aqueous environments, QS21 will lose its adjuvant abilities [84]. This loss of adjuvant ability will also happen if the quillaic acid is inhibited or altered [73], but the hydrolysis also creates hemolytic byproducts. This complicates the storing and vaccine formation using QS21.
Lastly, and most importantly, the usage and research with QS21 is limited by the fact the mechanism of action (MOA) behind how QS21 interacts is still poorly understood [79]. It remains unknown if there are differences between the MOA of QS21-A and QS21-X. Several studies, such as the structure-activity relationship studies completed by the D. Gin group have tried to understand the MOA [85]. The Gin group was able to develop a total synthesis method of both QS21-A [86, 87] and QS21-X [88, 89] which does help with the limited supply of both isomers. They were able to develop several variants, but it is unknown whether the variants even have the same MOA as QS21 [90].

### 1.4.2 Currenting Suggesstions for QS21’s Mechanism of Action

While there have a few ideas on how the MOA of the QS21 isomers, D. Marciani published a review during the completion of this thesis that details the current information concerning the MOA of saponins with a focus on QS21 in 2018 [90]. After researching the past suggestions and the ones presented in this review, we will review only the suggestions presented in the 2018 article as we have found it to reflect current experimental research and understanding of the molecule. Nevertheless, we recommend reading Ref [90] and its reference materials for a more in-depth investigation in the MOA ideas themselves.

The first suggestion is that, like other Quillaja saponin adjuvants, QS21 has two MOAs instead of one as Quillaja saponin adjuvants interact differently with cytotoxic T lymphocyte and dendritic cells [91]. Starting with the former, the aldehyde on the quillaic acid is vital as inhibiting or altering this moiety will leave QS21 adjuvant inactive [73]. It is believed that this inactivity is from QS21 being unable to interact with the receptors on the surface of the cytotoxic T lymphocyte. This interaction stimulates Th1 bias T cell activation via the formation of a Schiff base with a free amino group on the T cell receptor [92, 93]. For the dendritic cells, the MOA suggestion here relies on the amphiphilic nature of the molecule and the acyl chain. We noted in Section 1.4.1 that the acyl chain is important to maintain adjuvant activity, and it was found later that this limitation could be evidence that QS21’s natural membrane interactions are key for the dendritic cell MOA [75]. There is more proof for this in work on QS21-like saponins which showed that they allowed antigens to escape endosome-lysosomes after endocytosis via membrane alterations. This allowed for the antigens to be processed further into the products necessary for the immune response [94, 95].
1.5 Computational Studies of QS21

Due to the limitations of QS21, increasing our understanding by using other means that avoid them is advantageous. Computational methods such as molecular dynamics simulations (see Chapter 2) move the molecules to the digital realm where the physical limitations of QS21 can be controlled. If we wish to investigate ways to prevent acyl chain removal, we can allow for the molecule to undergo the reaction. If we want to study how QS21 interacts with the bilayer, we can stop bond breakage from occurring. Other fields have used computational methods to investigate a range of phenomena from the electrostatic double layer by accounting for the electronic structure [96] to how proteins interact [97]. These results will always be compared to experimental methods to ensure the results make sense, but they offer a starting point or verification that will push experimental work towards an answer. Interestingly, the number of computational works in the literature collection for QS21 is scarce. The work that is out there is a combination of experimental and computational methods where the computational methods verified the experimental results.

Work from 2016 by W. Walkowicz et al. found that QS21-A flexibility was a necessity for adjuvant activity, through no information on QS21-X was given [98] and C. Pedebos et al. discovered some information about the structure and conformation of the QS21 micelle [99]. Yet how QS21, or either of its isomers, interacts with a membrane has not been studied in detail using experimental or computational means. The work in this thesis is one of the few that look at QS21-A bilayer interactions using computational methods. We hope that our work will help with drug design and understanding of QS21’s MOA.

1.6 Thesis Structure

In this thesis, we are using atomistic molecular dynamics to investigate how QS21-A interacts with one of three model lipid bilayers. In these models, we have cholesterol (Chol) in the bilayer along with one of two phosphatidylcholines (PC); dipalmitoyl-phosphatidylcholine (DPPC) or dioleoyl-phosphatidylcholine (DOPC). These bilayers would be in a system with some amount of QS21 and potassium counter-cation (K⁺) to investigate QS21-A interactions with the bilayer and micelle formation.
We have organized the thesis into five chapters with the references for each are found at the end of the chapter. Chapter 1 starts the thesis by covering the background of membranes and saponin adjuvants. Chapter 2 gives an explanation of the methods used to get our results with derivations found in Appendix A to keep Chapter 2 a smoother read. Chapter 3 reports on how we created the systems and analyzed our data. Chapter 4 discusses the results gathered from the systems. Chapter 5 finishes the thesis by giving a conclusion to our work and possible future works.

### 1.7 References


Chapter 2
The Methods Used

2.1 Molecular Dynamic Simulations

2.1.1 The Methods Types of Molecular Dynamic Simulations

At its core, molecular dynamics (MD) simulations are the systematic solving of Newton’s equation to study systems. This means we must solve the second derivative of the position of atomic nucleus \(i\) in terms of time multiplied by the mass of atomic nucleus \(i\) to get the force on atomic nucleus \(i\) shown as

\[
m_i \frac{\partial^2 r_i}{\partial t^2} = F_i. \tag{2.1}
\]

However, there are different methods of solving this equation depending on the level of detail needed as well as what is being simulated. Starting on the quantum side, we start with \textit{ab initio} MD [1] before moving onto classical MD [2] and then coarse-grained MD [2] both of which have moved away from quantum calculations. How these methods differ is shown graphically in Figure 2.1.

\textit{Ab initio} MD solves accounts for both electronic and nuclear degrees of freedom, allowing us to investigate systems in greater detail. To do this, the Schrödinger equation is used. An advantage to this method is that \textit{ab initio} simulations lead to the most possible details out of all the three methods listed. By accounting for the electronic variables, we are able to answer questions related to “complex interactions” with an example back in Section 1.5. These interactions include events such as bond formation and breakage which cannot be accounted for in the other methods as they trade in this level of detail for longer run times and larger systems.

The second type is classical MD, also called just MD, is the one used in this thesis. To get longer run times and larger systems, we forgo the Schrödinger equation and use the negative derivative of the potential in terms of the position to find the force as

\[
F_i = -\frac{\partial U(r_i)}{\partial r_i}. \tag{2.2}
\]
where \( U(r_i) \) is the total potential in terms of positions. These potentials are divided into bonded and non-bonded potentials that are modelled using different mathematical functions. These potentials are inputted into the simulation using force fields, both of which are described in more detail in Section 2.5.

The third type of MD simulations we will cover is coarse-grained MD. We can solve the largest problems in terms of time length and system size using this method as the individual atoms are replaced with pseudo-atoms. These pseudo-atoms are composed of many smaller atoms grouped together and are treated as one. There is a loss of even more detail, but it is replaced with the power to model systems with millions or more atoms [3-5]. This approach has been applied to large biological membranes, proteins and related systems [6-9].

![Figure 2.1](image)

**Figure 2.1:** A comparison of the different types of simulations discussed in the order of (A) \textit{ab initio}, (B) classical MD, and (C) coarse-grained MD. The \textit{ab initio} method accounts for the electronic variables, so the electrons are shown for each atom. The classical method does not consider these variables, so the bonds are shown as lines. Coarse-grained methods make pseudo-atoms from combining multiple atoms together.

### 2.1.2 Limitations of Molecular Dynamic Simulations

Like any method, be it experimental or computational, MD simulations have inherent limitations we must work within. This also means that there are systems and questions that simply cannot be tested effectively using MD. The questions “can an MD simulation produce realistic results for this system” and “do I, the user, understand MD simulations well enough to understand the results and the processes behind getting them” should always be asked before
starting a project. However, aside from these limitations, there are some other limitations that we will address in a bit more detail.

The first limitation is the force fields themselves. There are dozens of force fields currently in use and all of them have been designed and optimized for different molecules, simulation conditions, and key properties [10]. No one force field can produce results that completely account for all properties of all molecules. Thus, we must select the correct force field for the simulation based on the molecules we are using and studying.

The second limitation is with the algorithms themselves which compose the processes and are used in MD to complete the simulations. While authors do test them before publishing, some errors only come up after they have been used repeatedly under a range of conditions [11,12]. An algorithm can only really be proven reliable after years of public use and there can be room for improvements.

The third limitation is the timestep ($\Delta t$) selection. The $\Delta t$ needs to be smaller than the smallest motion in the system, and it must be reasonable. An unreasonable $\Delta t$ can lead to a system with either unfeasible time lengths and computational demands or a system producing inaccurate results. This is also why constraints, explained in Section 2.3, were developed; to restrict the molecular motions to a point where we can use a reasonable $\Delta t$ value. An incorrectly selected $\Delta t$ will cause problems in the algorithm results and thus artifacts that bias the system [13].

Lastly, the very hardware and software in use can themselves be a limiting factor. Starting with the former, all simulations require a set amount of computational power and space in order to work. This amount is determined by the experimental setup of the simulation, the method of completion, and the data analysis performed. These factors lead into the latter, and into the limitations of the software in itself. Each software is better suited for different needs or even specialized for a small set of actions, so care must be taken to select the best option [14].

2.2 Integrators

When selecting an integrator, it should fulfil the follow conditions; the integrator is symplectic, time reversible, and computationally efficient. If an integrator is symplectic, it means
that the integrator conserves the phase space volume [15,16]. By fulfilling this condition, the integrator will create a realistic system and thus is a key condition for an integrator to fulfil. An integrator needs to be time reversible because MD simulations are based on the Newton’s equations which are time reversible themselves. If an integrator cannot preserve this characteristic trait of the equations, then the integrator is altering their nature. Lastly, if an integrator is computationally efficient, then simulations can be performed with reasonable time lengths without causing overt demand on the hardware. This also plays into an integrator being able to handle long $\Delta t$ values.

With these conditions stated, we must also consider the software that we are using for the simulations. For this thesis, since we are using the Groningen Machine for Chemical Simulations (GROMACS) [17] software we used the Leapfrog integrator [18]. The Leapfrog integrator finds the positions $r$ at time $t + \Delta t$ and the velocities $v$ at time $t + \frac{\Delta t}{2}$. These values are found using the equations

\begin{align}
    v_i(t + \Delta t/2) &\approx v_i(t - \Delta t/2) + \left(\frac{\Delta t}{m}\right)F_i \tag{2.3} \\
r_i(t + \Delta t) &\approx r_i(t) + \Delta t v_i(t + \Delta t/2) \tag{2.4}
\end{align}

This integrator fulfills all of the conditions that we mentioned above and the derivation can be found in Appendix A. As shown, the $r$ and $v$ are not found at the same time and thus another step is needed to compute the energy at any one step. Instead, the two appear to be “jumping” over each other like a pair of children play the game of the same name as shown in Figure 2.2. A natural negative to this integrator is that an additional step is needed to find the energy value. While this does increase the number of operations needed, a positive is that the $v$ is included in the equation for finding $r$.

Figure 2.2: A visualization of how the Leapfrog integrator works to find the positions and velocities for a simulation. The positions, shown in purple, are found at every whole-time unit. The velocities, shown in dashed blue, are found at every half time unit. The positions also require the use of the velocities to find them.
2.3 Constraints

As mentioned in Section 2.1.2, the $\Delta t$ plays an important part in any simulation and we must have it smaller than the fastest motion, bond vibrations. Thus, constraints were developed to assist in reaching feasible $\Delta t$ and time length values [19].

The normal method of representing the bonds using a harmonic potential is what leads to this problem, so constraints “fix” this by keeping the bond length and possibly the angle at a set value. The method of doing so is different depending on which algorithm is being used, but the end results are a larger $\Delta t$, longer simulations, feasible experiments, and some detail loss as an unavoidable side-effect. The most commonly used constraint algorithms are SHAKE [20] and Linear Constraint Solver (LINCS) [21].

2.4 Periodic Boundary Conditions

When we create systems for simulations, these systems are surrounded by invisible walls that define our box size. While this is useful during system creation, during active simulations such walls would cause boundary interactions. These boundary interactions are a side-effect of the finite system size, but they will result in our data giving incorrect results. We want our finite system to behave as if it had “infinite” (or macroscopic) size it would normally have in experimental conditions. To do this, we apply periodic boundary conditions (PBC) [16] as shown in Figure 2.3.

![Figure 2.3](image)

**Figure 2.3:** An illustration of how periodic boundary conditions creates an infinite system by repeating the main box with its purple atoms in all directions. When atoms in the main box with the purple atoms move, this motion is reflected in all replicated boxes with its orange atoms with the same force and direction.
PBC deals with these boundary issues by creating copies of the primary simulation box infinitely in all directions. These replicated boxes will reflect any motion produced in the primary simulation box with the same magnitude. Thus, if an atom from the user-created box “moves” outside the main box, an atom from one of the replicated boxes will take its place. An obvious downside to this is that an infinitely size system means that we have an infinitely large number of interactions to deal with when calculating the potential. Nevertheless, there is a trick or a convention that can be applied in order to deal with these interactions. Explained in more detail in Section 2.5, we apply a cut-off distance to our interactions. Thus, only interactions within the cut-off are considered.

2.5 Potentials, Force Fields, and Water Models

Noted in Section 2.1.1, MD simulations require that the potentials of molecules are inputted at the start. These collections of potentials are called force fields and determine how the simulation is calculated. As such, force fields must account for the total potential $V_{\text{total}}$ by accounting for the potential of the bonded $U_{\text{bonded}}$ and non-bonded $U_{\text{non-bonded}}$ factors. In other words,

$$U_{\text{total}} = U_{\text{bonded}} + U_{\text{non-bonded}}. \quad (2.5)$$

The term $U_{\text{bonded}}$ is composed of the different motions that bonded atoms, or bonded atoms within a set distance, can perform with each other; stretching between two atoms, bending between three atoms, and rotation between four atoms. It must be noted that bond bending is also known as bond angle or angle bending and that bond rotation is more likely addressed as bond torsion or just torsion. It is possible to find examples for each of these cases, but for this thesis we will use the names bond stretching, angle bending, and torsion as these are the most commonly used names and show them in Figure 2.4. Updating Equation 2.5 leads to Equation 2.6, leaving only the non-bonded potential to define,

$$U_{\text{total}} = U_{\text{bond stretching}} + U_{\text{angle bending}} + U_{\text{torsion}} + U_{\text{non-bonded}}. \quad (2.6)$$

The term $U_{\text{non-bonded}}$ is for atoms that are not connected to their neighbours by bonds. Non-bonded interactions are typically divided into short- and long-ranged modelled using the Lennard-Jones and Coulomb potential, respectively. With this, the final form of Equation 2.5 is

$$U_{\text{total}} = U_{\text{bond stretching}} + U_{\text{angle bending}} + U_{\text{torsion}} + U_{\text{Lennard-Jones}} + U_{\text{Coulomb}}. \quad (2.7)$$
Figure 2.4: The four different components of the total potential shown graphical on example molecules. (A) shows bond stretching between two atoms and angle bending between three atoms. (B) shows the torsion between four atoms and the electrostatic interactions between the non-bonded atoms.

We now can rewrite Equation 2.7 with a mathematical function for each component. Rewritten in order, the typical mathematical equation that accounts for $U_{\text{total}}$ is

$$U_{\text{Total}} = \sum_{\text{bonds}} \frac{k_i^b}{2} (l_i - l_i^{\text{ref}})^2 + \sum_{\text{angles}} \frac{k_i^a}{2} (\theta_i - \theta_i^{\text{ref}})^2$$

$$+ \sum_{\text{torsions}} A_n [1 + \cos (n \tau - \varphi_n)] + \sum_{l=1}^{N-1} \sum_{j=l+1}^{N} 4\epsilon_{ij} \left[ \frac{(\sigma_{ij})^{12}}{r_{ij}} - \frac{(\sigma_{ij})^{6}}{r_{ij}^{6}} \right]$$

where we will define each variable below. We would also like to note that these functions take into account the Born-Oppenheimer approximation [2].

Starting with bond stretching, which is described by a harmonic potential and can be based on Hooke’s Law, as shown as Equation 2.8, or (albeit much less commonly) the Morse potential [2, 22]. For the parameters, $k_i^b$ is the spring constant which defines the bond stiffness, $l_i^{\text{ref}}$ is the reference distance at equilibrium, and $l_i$ is the distance value we measure at some time. Whether this potential is modelled using a harmonic or Morse potential depends on what is needed from the simulation. If the simulation needs to model bond breaking, the Morse potential can be applied.
The second and third terms model angle bending and torsion respectively. Angle bending is modelled using a harmonic potential based on Hooke’s Law and with its own spring constant. Torsion is modelled using a periodic function, as we can see from the inclusion of a cosine term. The parameters here are $A_n$ which is the energy barrier related to the rotation, $n$ which controls the periodicity, $\tau$ which stands for rotation angle axis, and $\varphi_n$ which shifts the entire curve along the rotation angle axis. An important part of the torsion term is the use of proper and improper dihedrals. A proper dihedral is just a normal dihedral that can be found in a molecule while an improper dihedral is an extra harmonic restraint that is applied to $\tau$ to keep planar groups planar and stop any possible flipping between mirror images.

This leaves the fourth and fifth for the short-range and long-range electrostatic interactions respectively. Starting with the short-range Lennard-Jones potential, we note that it is divided into two parts: repulsion originating from the Pauli exclusion principle and van der Waals interactions. The depth of the potential well is given by $\epsilon_{ij}$, the interatomic distance where the potential is exactly zero by $\sigma_{ij}$, and the distance between the particles $r_{ij}$. Due to the short-range nature of the Lennard-Jones interactions ($1/r^6$), we can apply a truncation or cut-off distance of $r_{\text{cut-off}}$ which saves computational resources. We cannot, however, truncate the long-range Coulomb potential ($1/r$ behaviour). We will discuss this more in Section 2.6, but we will note that applying a cut-off to the long-range interactions leads to serious artifacts.

The short-range interactions can be computed very efficiently since a cut-off, $r_{\text{cut-off}}$, can be applied. A strict rule for $r_{\text{cut-off}}$ is that it must be smaller than half the box length to avoid double-counting of interactions, but typically much shorter cut-offs are use, 0.9-0.14 nm are typical for modern force fields. This allows only the nearest possible image to be considered for the interaction calculations. We call this the nearest image convention shown graphically in Figure 2.5. Direct calculation scales as $O(N^2)$ but this can be improved. For $N$ atoms, the software must look at $N - 1$ other atoms to calculate interactions and divide the answer by two to prevent double-counting. A way to bring this number down is to use Verlet neighbour lists [16, 23, 24]. This method adds an extra layer of atoms starting from $r_{\text{cut-off}}$ to $r_{\text{Verlet}}$, see Figure 2.5. The atoms found in this region are stored in the Verlet neighbour list and can be called upon...
when solving for interactions. By having this list of extra atoms, the computational cost for neighbour searching is brought down to $O(N)$.

![Diagram of molecular dynamics](image)

**Figure 2.5:** A model of how neighbours are found in molecular dynamic simulations for non-bonded interactions. With periodic boundary conditions in place, a $r_{cut-off}$ is used that is smaller than half the box length. All the atoms within this distance are used to calculate interactions. This leads to $O(N^2)$ computation cost, but it is possible to reduce this value by using something like Verlet neighbour listing. In this method, a layer of extra atoms from $r_{cut-off}$ to $r_{Verlet}$ is stored in a Verlet neighbour list. By storing these atoms to call upon them later on, the computation cost is reduced to $O(N)$.

The force fields themselves are built using data gathered from quantum calculations, experimental results, or a mix of both [2]. They also come in three types that provided different levels of details and information; all-atom, united-atom, and coarse-grained, see Figure 2.6. In coarse-grained MD, several atoms are grouped together to form pseudo-atoms that are used instead [2]. This also means that these force fields have traded in detail for increased system sizes and longer run times. United-atom force fields do use some coarse-graining, but will only group together non-polar hydrogen atoms [2]. In simulations of biological models, this becomes advantageous as $C_xH_y$ groups make up a good deal of the atom groups. By coarse-graining only these groups, we are able to increase simulation times and size without losing all the details. Lastly, we have the all-atom force fields that account for every atom in a simulation [2]. This naturally means that they are the most computationally expensive of the force fields, but will provide the most information at the end.
Figure 2.6: A comparison different types of force fields in the order of (A) all-atom, (B) united-atom, and (C) coarse-grained. All-atom force fields account for every atom in a simulation, but at the cost of increased computational demand. United-atom force fields combine some atoms (typically methyl and methylene hydrogens with the corresponding carbons) together as to minimize the amount of detail lost while benefiting from the decreased computational cost. Coarse-grained force fields combine many atoms together and makes pseudo-atoms that have lost the finer details of the other force fields but is able to take on larger simulations.

With the force field selected, there is still the issue of choosing a water model. Since each force field is designed using different parameters, not all water models are necessarily compatible. The water model that is compatible with most force fields is the 3-site water model TIP3P [25, 26]. This is a simple model with rigid structure and accounts for only the oxygen and hydrogen atoms, and the degrees of freedom generated from them. The rigid structure limits the calculations required to work with these models, decreasing the cost to run the simulation.

Figure 2.7: A comparison of three types of water models to (A) a water molecule with electrons. The water models are named based on the number of “sites” and degrees of freedom. A 3-site model (B) with two hydrogens and one oxygen atom. A 4-site model (C) add a dummy atom that represent the lone pairs’ negative charge. A 5-site model (D) also use a dummy atom, but there is a second one so that both lone pairs are accounted for.
Other types of models are the 4-site and 5-site models as illustrated in Figure 2.7 [27]. Both of these models use dummy atoms to improve certain properties such as molecular structure and electrostatics. In the 4-site models, one dummy atom placed along the bisector or the molecule’s angle. The 5-site models have two dummy atoms that are used to account for the lone pairs on the oxygen atom.

2.6 Dealing with Long-Range Coulomb Interactions

2.6.1 The Problem with Truncating the Coulomb Potential

At this point we have discussed that PBC is used to prevent artifacts, how interactions are calculated in these simulations, and how truncating the Coulomb potential may lead to unwanted artifacts when the system has polar or charged groups. It is also a rather well-known problem as numerous articles have reported these problems [12, 28-31]. This latter point has been a driving force to developing better methods of dealing with the long-range Coulomb interactions, but there was a time where this was not possible: It was only a few decades ago that truncation was a necessary evil to accommodate the memory and power that early computers lacked.

Let us assume that there are two charges within a system called $q_1$ and $q_2$ as shown in Figure 2.8. As long as those two charges are within the distance $r_{\text{seen}}$, $q_1$ and $q_2$ will be able to “see” each other. When we bring these charges farther and farther apart, we reach the distance $r_{\text{hidden}}$ where this should not happen anymore. Now let us say that we have placed these two charges in a solvent with a large dielectric constant $\varepsilon$. Due to this large constant, the two charges would be shielded from each other. At distance $r_{\text{seen}}$, the two charges would only see $\frac{1}{\varepsilon}$ of the charge they would normally see. This is known as electrostatic or dielectric screening [12, 28]. As long as $r_{\text{cut-off}} \geq r_{\text{hidden}}$, then $q_1$ and $q_2$ would not be able to see each other and we would have no problems with the simulation. Since electrostatic interactions are of long-range ($\frac{1}{r}$ behaviour), matters are more complex and such a screening does not work well. As this thesis works with the lipid bilayer, we are going to spend some time going over the most serious effect of truncation in this particular system type as it was discussed by Patra et al. [29, 30].
Figure 2.8: Electrostatic or dielectric screening of two charges with and without a solvent with a large dielectric constant. When $q_1$ and $q_2$ are within the distance $r_{seen}$ (A), the two are able to “see” the charge from the other. By pulling these charges to $r_{hidden}$ (B), they are “hidden” from each other and long-range electrostatic interactions will not occur. By using a solvent with a large dielectric constant $\varepsilon$ (C), $q_1$ and $q_2$ will only see $1/\varepsilon$ of the charge they normally would at $r_{seen}$. This is the underlying theory behind how truncation would still be usable for electrostatic interactions.

We noted in Section 1.2.1 that membranes take on a fluid or liquid state within the human body. One of the most natural methods of verifying the state is using the Radial Distribution Function (RDF) [2]. In a graph such as shown in Figure 2.9, the RDF quantifies how far and how many atoms there are from a center atom. This measurement is very important and powerful as each state of matter produces an unique peak pattern. Since a solid has a repeating lattice structure, the RDF peaks will have a repeating pattern. Liquid and gas suffer from more diffusion, so they will have fewer peaks that die down as the distance from the center atom increases. When working with bilayers in simulations, the RDF should always show a liquid pattern since the bilayer is in the liquid state within the human body. However, when treated with truncation methods, Patra et al. showed that the bilayer displayed the pattern fitting a gel phase bilayer [29, 30]. This meant that the artifacts were extremely serious as the bilayer being modelled was not in the correct phase. Given this, we must treat the Coulomb interactions correctly to model bilayers in the correct phase.
Figure 2.9: An example of the correlation shells for the different states of matter as (A) to (C) with their radial distribution functions graphed in (D) at zero temperature. Each state of matter will produce an unique peak pattern due to the amount of diffusion that each undergo. (A) Solids have a repeating lattice structure and will have (A) a repeating pattern in the peaks shown as the blue line. (B) Liquids are more diffuse and will have the peaks die off shown as a dashed red line. (C) Gases suffer the most diffusion and will normally have only one peak, shown as a dotted green line.

2.6.2 Ewald Summation and Particle-Mesh Ewald

Figure 2.10: A mock NaCl unit cell which has the length $L$ and a volume of $L^3$ with three sides shown. Overall, NaCl is neutral but is composed of cations and anions. The electrostatic potential of ionic crystals such as this led to problems as, since direct summation of the Coulomb interactions led to conditional convergence only. Ewald solved this paradox by developing a method known as the Ewald summation shown in Ref. [31].
Based on the information from Section 2.6.1, how do we account for the long-range Coulomb interactions without causing artifacts? The most common approaches based on Ewald summation developed by P. P. Ewald in 1921 [31]. This summation accounts for the full Coulomb potential and produces answers that are in excellent agreement with experimental data. Mathematical derivation of the summation is provided in Appendix B and here we focus on a review of the conceptual background, see also Figure 2.10. After that, we discuss at the computational cost of the summation. We recommend Refs. [2, 28] for a complete history of the topic.

Ewald summation splits the Coulomb interactions into short-range piece and long-range parts. It does this by applying a trick that arises from using two special functions. These are the error function (erf(\(\beta r\))) and the complementary error function (erfc(\(\beta r\))) functions

\[
\text{Erf}(\beta r) = \frac{2}{\sqrt{\pi}} \int_0^{\beta r} e^{-x^2} dx
\]

\[\text{erfc}(\beta r) = 1 - \text{erf}(\beta r) = \frac{2}{\sqrt{\pi}} \int_{\beta r}^{\infty} e^{-x^2} dx . \]

When these functions are divided by \(r\) and are added together, the result is

\[
\frac{\text{erfc}(\beta r)}{r} + \frac{\text{erf}(\beta r)}{r} = \frac{1}{r} .
\]

Conceptually, Equation 2.11 add together two opposite factors. One of them is a screening charge, or a charge cloud with equal but opposite magnitude to the point charge. The other factor is a compensating charge, or a charge cloud that is equal but opposite in magnitude to the screening charge. This is illustrated in Figure 2.11. The above allows for the summation to be divided into two parts. The first part is done in the real or time space and is just a direct sum that accounts for the short-range portion. The second part accounts for the long-range portion and is done in Fourier space.

34
Figure 2.11: How the point charges, shown in blue, horizontal lines, of a system are found. These are assigned screening charges of equal but opposite magnitude, shown as orange peaks. Compensating charges, shown as purple peaks, are equal but opposite magnitude to the screening charges. When these factors are added together, the screening and compensating charges cancel out and the point charges remain.

In the end, the equation for the Ewald summation takes the form

\[
\frac{1}{2} \sum_{i<j}^N \left( \frac{q_i q_j}{r_{ij}} \text{erfc}(\sqrt{\alpha r_{ij}}) \right) + \frac{1}{2V} \sum_{k \neq 0} \left( \frac{4\pi}{k^2} \right) |\rho(k)|^2 \exp \left[ \frac{-k^2}{4\alpha} \right] - \left( \frac{\alpha}{\pi} \right)^2 \sum_{l=1}^N \rho_l^2 = V_{\text{Coul}}. \quad (2.12)
\]

Starting with the first term, this is the one that works with the screening charge and the short-range portion (see Figure 2.11). The \( q \) stands for the charge and \( r_{ij} \) is the distance between \( q_1 \) and \( q_j \). The second term for the long-range contribution and the solution for Poisson’s equation. Briefly, Poisson’s equation, shown using Gaussian notation,

\[
-\nabla^2 \phi = 4\pi \rho(r). \quad (2.13)
\]

However, by solving for Poisson’s equation in the Fourier space, we now have to remove the self-energy that a charge will face. This is the third term shown which is subtracted from the rest of the summation.

The main drawback of the method was and is the computational cost. Normally the Ewald summation is \( O(N^2) \) but it can be reduced to \( O(N^{3/2}) \). Regardless, the origin of it will always be the summation in the Fourier space. This pushed for the development of methods that could perform the Ewald summation while bringing down the computational cost. Several methods have been developed as discussed, for example, in Refs. [2, 28]. The most common is the so-
called the Particle-Mesh Ewald or PME [32, 33] as illustrated in Figure 2.12. PME is also used in GROMACS.

**Figure 2.12:** The idea behind how the Particle-Mesh Ewald speeds up the Ewald summation. The charges or particles of the system are assigned to a grid or mesh and Poisson’s equation for this grid is solved. This leads to the potentials required to solve for the forces needed. After this, the forces are interpolated back to the charges or particles.

PME decreases the computational cost by completing the summation using Fast Fourier transforms. PME first needs to map the charges or particles to a grid or mesh. Poisson’s equation is then solved on the grid. This will have the potentials and forces being solved for lattice point. The problem now is that these forces need to be interpolated back to the particles. By doing this, PME has a computational cost of $O(N\log N)$ using a Gaussian function and Lagrange interpolation.

### 2.7 Ensembles and How to Make Them

Setting up the conditions for a simulation is analogous to setting up the conditions for an experiment. An important factor to note is that, under normal integration, Newton’s equations will produce results that match what is called the microcanonical ensemble or isolated system [16]. A microcanonical ensemble means that the particle number $N$, volume $V$, and total energy $E$ remain constant. In fact, any MD software or force field must be able to recreate this ensemble to be valid. Most experimental work is, however, not done under microcanonical conditions and thus we must change it. Conceptually, in a MD simulation this can be done the same way that
Experimentalists do so, by adding a heat bath, or adding a pressure bath/piston as illustrated in Figure 2.13.

![Figure 2.13: The three common types of ensembles/conditions used represented with beakers (A) microcanonical or NVE, (B) canonical or NVT, and (C) isobaric-isothermal or NPT or Gibbs.](image)

Thermostats, which are needed for constant temperature, are shown as a heat bath. Barostats, needed for constant pressure, are shown as a piston. The number of particles is kept constant in all of them with (A) and (B) also keeping the volume constant. (C) trades in constant volume for constant pressure.

But what is an ensemble? It can be loosely defined as a very large collection of systems, or replicas, that each model/represent a well-defined thermodynamic state with given constraints such as pressure $P$, temperature $T$, or volume $V$ [2, 16]. Any observable is then defined as an average over the systems in the ensemble. A typical MD simulation, however, does not consist of such collection of replicas. Rather, a MD simulation produces a single time-dependent trajectory. Provided the system has reached equilibrium and that the system is ergodic, then, by the ergodic hypothesis, the ensemble and time averages are equivalent. This justifies the use of MD simulations for modeling equilibrium properties.

To obtain the experimentally relevant canonical (NVT) ensemble or the isothermal system, $N$, $V$, and $T$ must be kept constant for this one. From the MD simulation point of view, we introduce a heat bath, or a thermostat, to the system. Several algorithms have been created to keep the temperature constant including the Andersen thermostat [34], Nosé-Hoover thermostat [2, 35], Berendsen thermostat [36], and velocity rescaling [37]. In the simulations presented in this thesis the Nosé-Hoover thermostat. Appendix C has more information on two thermostats.

To be able to simulate the Gibbs (NPT) ensemble or the isobaric-isothermal system, we need to add a pressure bath, or piston, to the system. This will cause the $V$ to change such that $P$
remains constant. Like thermostats, there are several few algorithms that can do this including the Andersen barostat [34], the Berendsen barostat [36], and the Parrinello-Rahman barostat [38-40]. The Parrinello-Rahman barostat is used as it also allows the user created box to change shape. Due to this property, this barostat is often used to simulate solids since phase changes are possible [38-40]. It should also be noted that the NPT ensemble is necessary to model lipid bilayers as they need to flex in the x/y plane. Appendix D has more information on two barostats.

2.8 References

22) Morse, P. M. Phys. Rev. 1929, 34, 57–64.
Chapter 3
System Creation and Data Analysis

3.1 System Information and Creation

3.1.1 Simulation Parameters

As noted in Chapter 2 Section 2.2, we are using GROMACS [1] to complete the MD simulations but we are specifically using GROMACS 2016.3. The timestep $\Delta t$ was set to 2 fs for the Leapfrog integrator [2] and LINCS [3] was used as the constraint algorithm. PBC [4] were applied to prevent boundary artifacts and allow the use of PME [5] to solve for the Coulomb/long-range electrostatic interactions. The Lennard-Jones/short-range electrostatic interactions were dealt with using the Verlet method [4, 6, 7] with $r_{cut-off}$ of 1.0 nm. The Optimized Potentials for Liquid Simulation All-Atom (OPLS-AA) [8, 9] force field was selected as the force field. Water was modelled using the Transferable Intermolecular Potential with 3 Points (TIP3P) [10, 11] water model since it is compatible with the OPLS-AA force field and a 3-point model is more computationally efficient. Systems with bilayers were ran in the NPT ensemble conditions created using the Nosé-Hoover thermostat [12, 13] and Parrinello-Rahman barostat [14-16] to ensure correct compression on the bilayers for an accurate model of their behaviour. Non-bilayer systems did not require the barostat, so they were left in the NVT ensemble with only the thermostat applied to them. We set the temperature to above the main phase transition temperature at 323K in all systems and the NPT systems had the barostat set to 1.0 bar with a compressibility of $4.5 \times 10^{-5}$ bar$^{-1}$ under semi-isotropic conditions. The semi-isotropic conditions allowed for the $x/y$ plane to change independently of the $z$-direction. All systems were ran for 100 ps in NVT pre-equilibration and 100 ps NPT pre-equilibration before moving into their permanent ensemble for continuous runs. Runs were sent to the computing clusters in batches totaling 100 ns with a snapshot of each system being taken at the 100 ns mark.

3.1.2 Creation Methods and Created Systems

Table 3.1 has the number of atoms and molecules of the systems studied in this thesis, and Table 3.2 has the dimensional information and ensembles used for them. The lack of literature
regarding MD studies of QS21 made comparisons to other studies rather difficult, and we started with three different initial systems. These were the 50 QS21-A systems with 538 PC and 486 Chol with varying amounts of water molecules. The OPLS-AA force field was used for both QS21-A and PC lipids. We used the QS21-A molecule information from these systems to create QS21-A reference systems and insert the molecule into new bilayer systems. The aim was to investigate how a QS21-A micelle forms in the presence of a bilayer. Another point of interest was whether or not QS21-A would interact with the bilayer if other QS21-A molecules are present.

The reference systems were the NVT systems created using GROMACS commands to insert 50 and 10 QS21-A molecules into two different water boxes with an equal number of counter-cations. We investigated how QS21-A interacts without the bilayer present both with itself and in micelle formation. The final micelle structure was also studied to understand how QS21-A interactions work.

We will discuss this more in Chapter 4, but the 50 QS21-A DOPC-Chol system displayed interesting interactions between QS21-A and the bilayer and that required further investigation to gain insight into the penetration mechanism. To investigate this more thoroughly, we took the 50 QS21-A system at 200 ns and created two new systems with 10 QS21-A and 5 QS21-A molecules via random molecule removal.

The next set of systems were created to study the interactions between a single QS21-A and a bilayer. These systems were made smaller to allow for simulations to reach longer time scales in a shorter period of wall-clock time. The bilayers were created with 256 PC and 120 CHOL with one QS21-A and one K⁺. The bilayer was compiled using the CHARMM-GUI [17-23] membrane creator. QS21-A and ion addition was done using GROMACS. While cellular membranes are composed of several types of lipids separated into different regions, investigating the effects of individual lipids is necessary to build a clearer picture of what occurs when lipids are mixed.

In this thesis, we will address the systems based on QS21-A molecules and PC lipid such that 50-DOPC means the DOPC-Chol bilayer system with 50 QS21-A molecules, 50-Ref is the 50 QS21-A system with no bilayer, and 10-DOPC-SI is the 10 QS21-A system built from 50-DOPC
to study the interactions between QS21-A and the bilayer. The systems differ based on the PC lipid species in the bilayer and the number of QS21-A as shown in Table 3.1.

### Table 3.1: Number of Atoms and Molecules in the Studied Systems

<table>
<thead>
<tr>
<th>System Name</th>
<th>Atom Total</th>
<th># of H$_2$O</th>
<th># of PC</th>
<th># of Chol</th>
<th># of QS21-A</th>
<th># of K$^+$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-DPPC</td>
<td>344442</td>
<td>74746</td>
<td>538</td>
<td>486</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>50-DOPC</td>
<td>327110</td>
<td>67534</td>
<td>538</td>
<td>486</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>50-Ref</td>
<td>325340</td>
<td>103680</td>
<td>0</td>
<td>0</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>10-Ref</td>
<td>325340</td>
<td>107333</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>10-DOPC-SI</td>
<td>315670</td>
<td>67534</td>
<td>538</td>
<td>486</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>5-DOPC-SI</td>
<td>314240</td>
<td>67534</td>
<td>538</td>
<td>486</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>1-DPPC</td>
<td>193286</td>
<td>50280</td>
<td>256</td>
<td>120</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1-DOPC</td>
<td>208420</td>
<td>54642</td>
<td>256</td>
<td>120</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

### Table 3.2: Dimensional Information for Systems After Pre-Equilibration

<table>
<thead>
<tr>
<th>System Name</th>
<th>Ensemble</th>
<th>x-axis Length (nm)</th>
<th>y-axis Length (nm)</th>
<th>z-axis Length (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>50-DOPC</td>
<td>NPT</td>
<td>15.13</td>
<td>15.13</td>
<td>13.95</td>
</tr>
<tr>
<td>50-Ref</td>
<td>NVT</td>
<td>13.39</td>
<td>13.39</td>
<td>18.62</td>
</tr>
<tr>
<td>10-Ref</td>
<td>NVT</td>
<td>14.48</td>
<td>14.48</td>
<td>16.02</td>
</tr>
<tr>
<td>10-DOPC-SI</td>
<td>NPT</td>
<td>15.14</td>
<td>15.14</td>
<td>13.5</td>
</tr>
<tr>
<td>5-DOPC-SI</td>
<td>NPT</td>
<td>15.11</td>
<td>15.11</td>
<td>13.35</td>
</tr>
<tr>
<td>1-DPPC</td>
<td>NPT</td>
<td>8.79</td>
<td>8.79</td>
<td>24.95</td>
</tr>
<tr>
<td>1-DOPC</td>
<td>NPT</td>
<td>9.96</td>
<td>9.96</td>
<td>21.16</td>
</tr>
</tbody>
</table>

### 3.2 Data Analysis

#### 3.2.1 Visual or Qualitative Inspection

The software Visual Molecular Dynamics (VMD) [24] was used to generate snapshots of the systems every 100 ns during the equilibration stages. It was also during this time that the systems were investigated visually through the VMD windows. Thanks to the built-in tools, it was also possible to visualize possible hydrogen bonding (H-bonding).

#### 3.2.2 Area Per Lipid, Membrane Area, and Membrane Thickness

For this thesis, we use three common membrane properties to understand QS21-A interactions with our bilayers as the basis. All three are indicators of the membrane phase, and all three are related to each other. The APL is the first membrane property that we analyzed in this
thesis as it is a defining characteristic for each bilayer lipid species. It is an experimentally measurable quantity which force fields and system parameters must match this for them to be valid and useable. The simulation must correctly predict the APL within its error range to produce correct results for structural properties such as membrane area, membrane thickness, and lipid order [25-27]. A continuously decreasing or lower APL value can be reflective of a membrane losing fluidity as the lipids pack closer together, moving closer to the gel phase, or the long-range interactions are dealt with improperly as we touched on in Section 2.6.1. The membrane area is the second property and is the summation of the area across the membrane. The third property is the membrane thickness, an important factor in cellular health. If the membrane is too thick, permeability will decrease and the cell will suffer from starvation and waste mismanagement [28]. This also causes problems for membrane proteins and structures needed for regular function like the cytoskeleton which gives the cell shape and scaffolding. If too thin, permeability will increase and the cell can suffer a leakage. In this thesis there are no membrane proteins. Nevertheless, if the permeability of our bilayers is not correct then any observed penetration could be unrealistic.

All three of these properties were investigated using the Fast Analysis Toolbox for Simulation of Lipid Membranes (FATSLiM) software [29], an opensource program. When analyzing these properties, FATSLiM maps the local environments of the lipids for its analysis. This enables it to take account of local fluctuations in the calculations that could cause noise or errors. For the APL, FATSLiM creates Voronoi cells for the calculation. After locating the neighbouring lipids for a central lipid, the neighbours are projected into a 2D plane where the area closest to the central lipid construct the cell and thus the APL. The membrane area is a summation of these values. Thickness is calculated by measuring the distance between two local averages to account for fluctuations in the membrane [29]. This is all done mapping the individual lipids as a bead for the headgroup and a vector for the direction to simplify the calculations. These processes are shown in Figure 3.1 and we will be using the APL and lipid order to verify the membrane phase.
Figure 3.1: The PC lipid DPPC as seen in the FATSLiM software. After treating the lipids as beads, FATSLiM will find the area per lipid using Voronoi cells, membrane area by area summation, and membrane thickness by measuring the distance between two reference points.

3.2.3 Lipid Order Parameter

The lipid order is typically measured by the quantity called the lipid order parameter $S_{\text{lipid}}$

$$S_{\text{lipid}} = \frac{1}{2} (3 \langle \cos^2 \theta \rangle - 1)$$

(3.1)

where $\theta$ is the angle between the bilayer normal $v_{\text{normal}}$ and the carbon-hydrogen vector $v_{C-H}$ illustrated in Figure 3.2. $S_{\text{lipid}}$ ranges between $-\frac{1}{2}$ and 1 with the latter meaning more ordered.

Figure 3.2: Solving the angle $\theta$ between the bilayer normal $v_{\text{normal}}$ and carbon-hydrogen vector $v_{C-H}$ on a DPPC tail. This value is used to find the lipid order parameter and thus the lipid order of the lipid tails.
More orderly (straighter) lipids are more tightly packed than less orderly or kinked lipids. Such ordering emerges when the bilayer moves towards the gel phase with decreasing temperature [26, 27]. It is related to the deuterium order parameter $S_{CD}$, which can be found using NMR, by

$$S_{CD} = -\frac{S_{\text{lipid}}}{2}.$$  \hspace{1cm} (3.2)

To find the order parameter, we used GROMACS to run the calculation on the $sn1$ and $sn2$ separately and reports the $S_{CD}$. The values were then averaged separately to get the overall value for the tail.

![Diagram of lipid molecules]

**Figure 3.3:** How the carbons were numbered by GROMACS for analysis.

### 3.2.4 Hydrogen Bonding

H-bonding/H-bonds are key in many aspects of life from giving water its many unique, life-properties to holding the DNA helix together [30]. They form when a donor atom and its hydrogen are within distance $r_{HB}$ and angle $\alpha_{HB}$ of an acceptor atom as shown in Figure 3.3. However, the donor and acceptor atoms both must be electronegative and the hydrogen is covalently attached to the donor. Since these bonds are a consequence of electrostatic interactions and since the H-bond is directional, we must define H-bond parameters before calculating them. For $r_{HB}$, we typically use the first minimum of the RDF between the atoms likely to form the bond; in the case of water, the first minimum defines its first hydration shell. We discussed the general shape and the peaks of RDFs in Section 2.6. For this thesis we chose to use the $r_{HB}$ based on $r_{OO}$, or the first minimum, between two water oxygens of the TIP3P water...
model [31, 32]. The angle $\alpha_{HB}$ was chosen as being the recorded value normally used for this analysis as the general $\alpha_{HB}$ value. These values are

$$r_{HB} \leq 0.35 \text{ nm}$$  \hspace{1cm} (3.3)

and

$$\alpha_{HB} \leq 30^\circ$$  \hspace{1cm} (3.4)

which are also the default parameters in the GROMACS 2016.3 software [1] and illustrated in Figure 3.4. The more general parameters were used given the variety of molecules and possible H-bond combinations in our systems as well as to retain the ability to compare the results. This is also one, if not the first time, that the H-bonding of QS21-A is investigated.

![Figure 3.4](image)

**Figure 3.4:** The criterion for hydrogen bonding between the hydrogen atom, shown in yellow, of a donor, shown in red, and an acceptor, shown in purple. The distance $r_{HB}$ and the angle $\alpha_{HB}$ were selected to be 0.35 nm and $30^\circ$ respectively. The $r_{HB}$ corresponds to the first minimum of the RDF of the TIP3P water model.

### 3.2.5 Radius of Gyration

To understand the shape of the QS21-A micelle we used the radius of gyration ($R_g$) which measures the compactness of macromolecules based on the distance the mass is from the COM [33]

$$R_g = \left( \frac{\sum_i ||r_i||^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}}$$  \hspace{1cm} (3.5)

where $m_i$ is the mass of atom $i$, and $r_i$ are the positions of atom $i$ with respect to the COM of the molecule. GROMACS has a built-in command to find this value, and this thesis will be the second study that we know of that measures this property to study the QS21-A micelle [34]. However, we will also report the $R_g$ in around the different axis via summing the radii components orthogonal (perpendicular) to the axis.
\[ R_{g,x} = \left( \frac{\sum_i (r_{ix}^2 + r_{iz}^2) m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \]  

(3.6)

where the \( x, y, \) and \( z \) subscripts indicate the axis.

### 3.3 References


4) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*, 1st edition; Oxford University Press, **1987**.


Chapter 4
Our Results and The Discussion

4.1 The Data Used

For this thesis, we are reporting the results for the latest 100 ns of the simulations at the time of writing. 1-DPPC, 1-DOPC, and 50-Ref have results from a total 900 ns while the rest of the systems have results from 700 ns. The time selection for 50-Ref and 10-Ref was also based on when the QS21-A molecules came together to form a single micelle. Chapter 3 explained that 5-DOPC-SI and 10-DOPC-SI were created from the 50-DOPC system at the 200 ns mark. We include the data from 120-200 ns for the 50-DOPC system and 10-DOPC-SI to explain that event and show the effect of QS21-A reduction on the bilayer.

We report the average, indicated using angular brackets, and standard deviation for the values reported. The average was found by adding the collected values together and dividing the result by the number of numbers. Standard deviation is found by

\[
SD = \sqrt{\frac{\sum (x_i - \langle x \rangle)^2}{(n - 1)}}
\] (4.1)

where \(SD\) is the common method of indicating standard deviation in statistical formulas, \(x_i\) is the observed value \(i\), \(\langle x \rangle\) is the average of the values, and \(n\) is the number of observed values and measures how widely the values are dispersed from the average. We report this as \(x \pm SD\).

This chapter is divided into three sections starting with the results and discussion of the membrane properties. The second section is the results and discussion of the interactions between the bilayer and the QS21-A as well as the effect of QS21-A reduction. The third section is the results and discussion of the interactions between QS21-A and water and itself. In this chapter, the snapshots of the systems such as those in Figure 4.1 have the water molecules hidden. The water molecules were hidden in this and other snapshot images to better show the micelles and how QS21-A was interacting with the bilayer and each other.
4.2 Membrane Properties

Figure 4.1: Images of (A) 1-DPPC at 900 ns and (B) 50-DPPC at 700 ns without water.

Reporting the properties, the APL was divided into component-wise value for Chol ($A_{\text{Chol}}$) and for the lipid ($A_{\text{Lipid}}$). The $S_{\text{Lipid}}$ was analyzed at near the beginning and end of the 100 ns for both tails. The thickness measurement used the phosphate atom from the lipid ($D_{\text{P-P}}$).

Starting with 1-DPPC, Figure 4.2 has the APL, thickness, and $S_{\text{Lipid}}$ for sn1 and sn2 at 800 ns and 895 ns. The $\langle A_{\text{Lipid}} \rangle$ is $0.446 \pm 0.002$ nm$^2$ and $\langle A_{\text{Chol}} \rangle$ is $0.327 \pm 0.004$ nm$^2$ plotted in Figure 4.2 (A). Comparing to literature OPLS-AA with some united-atoms based values at the same temperature [1] these values are respectively lower and higher than what the reported $\langle A_{\text{Lipid}} \rangle$ of $\sim 0.542$ nm$^2$ and $\langle A_{\text{Chol}} \rangle \sim 0.270$ nm$^2$. This could be caused by our systems requiring more time to reach the proper phase, or from the simulations just producing different results from Ref. [1]. The $D_{\text{P-P}}$ shown in Figure 4.2 (B) with $\langle D_{\text{P-P}} \rangle$ of $4.75$ nm $\pm 0.017$ nm was higher than the 3.9 nm experimental to 4.4 nm computational [1, 2], and the $S_{\text{Lipid}}$ has values starting at $\sim 0.34$ for carbon 12 and ending at $\sim 0.28$ for carbon 14 while Ref. [2] shows the NMR producing results of $\sim 0.24$ for carbon 12, reaching a max of $\sim 0.45$ before coming back down to $\sim 0.24$. However, all graphs match visually the trends seen in reported DPPC graphs [1, 2]. We saw the same with 50-DPPC; $\langle A_{\text{Lipid}} \rangle$ $0.425 \pm 0.002$ nm$^2$, $\langle A_{\text{Chol}} \rangle$ $0.356 \pm 0.002$ nm$^2$, $\langle D_{\text{P-P}} \rangle$ is $4.71 \pm 0.011$ nm, and the $S_{\text{Lipid}}$ with graphs matching Figure 4.2 closely. These graphs are in Appendix E given the similarities to 1-DPPC. Given this, while differing methods can lead to differences between results [3-5], this can also likely mean that our membranes needed more time.
Figure 4.2: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 1-DPPC.
Figure 4.3 has the four DOPC systems at their end times and Figure 4.3 (C) shows the end result of the event that led to the creation of 5-DOPC-SI and 10-DOPC-SI. This will be discussed more in Section 4.3, but near the 200 ns point a QS21-A molecule began to penetrate the bilayer in the 50-DOPC system. By 700 ns, however, this QS21-A molecule had been pulled back out and into the micelle in both the 5-DOPC-SI and 50-DOPC systems. Despite the addition of a foreign body into the bilayer, the membrane properties for 10-DOPC-SI closely matched the results of the 5-DOPC-SI and 50-DOPC systems. Unfortunately, as penetration only occurred once we cannot say with accuracy if QS21-A penetration effected membrane properties, or if this is only true in a DOPC-Chol bilayer. We also cannot say if the lack of effect is due to how the QS21-A penetrated the bilayer.

![Figure 4.3](image1)

**Figure 4.3:** Images of (A) 1-DOPC at 900 ns, (B) 5-DPPC-SI at 700 ns, (C) 10-DOPC-SI at 700 ns, and (D) 50-DOPC at 700 ns all with water removed.

5-DOPC-SI, 10-DOPC-SI, and 50-DOPC all had very similar values across their membrane properties we studied being within range from each other. As these systems are related to each other, this was a desirable result. Given this, we will place the graphs for 5-DOPC-SI and 50-
DOPC in Appendix E and show the graphs for 1-DOPC in Figure 4.4 and 10-DOPC-SI in Figure 4.5 given the differences.

1-DOPC had $\langle A_{Lipid} \rangle$ of $0.584 \pm 0.006$ nm$^2$ and $\langle A_{Chol} \rangle$ of $0.396 \pm 0.008$ nm$^2$. 5-DOPC-SI, 10-DOPC-SI, and 50-DOPC had $0.488 \pm 0.003$ nm$^2$, $0.485 \pm 0.004$ nm$^2$, and $0.491 \pm 0.003$ nm$^2$ $\langle A_{Lipid} \rangle$ and $0.395 \pm 0.003$ nm$^2$, $0.397 \pm 0.003$ nm$^2$, and $0.396 \pm 0.003$ nm$^2$ respectively for their $\langle A_{Chol} \rangle$. The $\langle A_{Lipid} \rangle$ for 1-DOPC is higher than the computational 43.2-54.6 nm$^2$ reported (depending on the calculation method was done) while the $\langle A_{Chol} \rangle$ values are within the 23.2-39.3 nm$^2$ [6] while the other DOPC systems are just fine. These results reflect the three systems being in similar phases despite the different concentrations of QS21-A. The APL graphs are shown in Figure 4.4 (A) and Figure 4.5 (A).

The $\langle D_{P-P} \rangle$ was the opposite to the $\langle A_{Lipid} \rangle$ with 1-DOPC being thinner at 4.25 nm $\pm 0.03$ nm and a $\langle D_{P-P} \rangle$ of 4.54 $\pm 0.02$ nm for 5-DOPC-SI and 10-DOPC-SI, and 4.52 $\pm 0.02$ nm for 50-DOPC. For the Chol ratios we were using, the reported values are 4.57-4.64 nm for the 5-50 QS21-A systems and 4.64 nm for 1-DOPC, but since a similar trick was done with the tail carbons as reported for the DPPC references this can be attributed to that. The $D_{P-P}$ graphs are shown in Figure 4.4 (B) and Figure 4.5 (B).

Using the sn1 values, $\langle S_{Lipid} \rangle$ for 1-DOPC at 800 ns is $0.180 \pm 0.052$ and $0.175 \pm 0.051$ at 895 ns. 10-DOPC-SI had $0.237 \pm 0.053$ at 600 ns and $0.238 \pm 0.054$ at 695 ns. Again, these values appear to agree with Ref. [6] simply because of the graph shape and trends since Ref. [6] does not prove averages as did Ref. [7]. The graphs for the $S_{Lipid}$ are shown in Figure 4.4 (C) and Figure 4.5 (C).
Figure 4.4: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 1-DOPC.
Figure 4.5: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 10-DOPC-SI.
In summary, the DPPC systems had different values for their membrane properties compared to literature values [1, 2] while the DOPC systems were in agreement [6, 7]. For the former, this may have been caused by these systems needing more walk-clock time to run or by simulations producing slightly different results to experimental results and computational results produced by other methods [3-5]. The 5-DOPC-SI, 10-DOPC-SI, and 50-DOPC had values within standard deviation of each other, a desirable result since the former two were built from the latter. We also saw thatQS21-A penetration, seen in 10-DOPC-SI, may not affect membrane properties but more tests are needed to verify this result. For this penetration event, if these results were translated into an experiment then some other measure has to be used to verify if QS21-A is within the bilayer.

Table 4.1 collects the average and standard deviation values for the systems with the x being used instead of the starting number for time since the single QS21-A systems had a different start and end times compared to the other ones.

<table>
<thead>
<tr>
<th>Table 4.1: Different Membrane Properties for the Systems</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>$A_{Lipid}$</strong></td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>$A_{Chol}$</strong></td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>$D_{P-P}$</strong></td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>$S_{Lipid}$</strong></td>
</tr>
<tr>
<td><strong>sn1 x00 ns</strong></td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>sn1 x95 ns</strong></td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>sn2 x00 ns</strong></td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
<tr>
<td><strong>sn2 x95 ns</strong></td>
</tr>
<tr>
<td>Std Dev</td>
</tr>
</tbody>
</table>
4.3 Reduction Effect and Membrane-QS21-A Interactions

The penetration event of 50-DOPC lead to a QS21-A remaining in the bilayer in the 10-DOPC-SI system. As noted, the penetration event happened near the 200 ns mark of the 50-DOPC system. We observed this by visual inspection using VMD and snapshots of the event are provided Figures 4.6 (A) to (C). To verify this, we investigated H-bonding between D-API and Chol as explained in Chapter 3, Section 3.2.4 with the graph shown in in Figure 4.6 (D).

![Figure 4.6](image)

**Figure 4.6:** The terminal D-API of QS21-API at (A) 180 ns, (B) 190 ns, and (C) 200 ns as it penetrated the bilayer into the Chol-lipid backbone area in the 50-DOPC system. (D) shows that D-API had penetrated the bilayer to reach the Chol and form H-bonds with terminal sugars being shown using their GROMACS names.
After reduction of the QS21-A molecules was performed to create the 5-DOPC-SI and 10-DOPC-SI systems, we analyzed the APL and $D_{p-p}$ of the 10-DOPC-SI for 45 ns afterwards to see if reduction made any difference. The effects were immediately noticeable on the graphs.

**Figure 4.7:** The effect of reducing the number of QS21-A molecules from 50 to 10 on (A) the area per lipid, and (B) the membrane thickness. After reduction, shown as a green arrow, the area per lipid increased while thickness deceased by a noticeable factor to values that were very close to what is in Table 4.1.
Seen in Figure 4.7, the APL increased for both DOPC and Chol while $D_{p-p}$ decreased. Numerically, the difference is harder to visualize. The 50-DOPC part had a $\langle A_{\text{Lipid}} \rangle$ of 0.481 ± 0.003 nm², a $\langle A_{\text{Chol}} \rangle$ of 0.392 ± 0.002 nm², and a $\langle D_{p-p} \rangle$ of 4.56 ± 0.014 nm. The 10-DOPC-SI side had a $\langle A_{\text{Lipid}} \rangle$ of 0.493 ± 0.006 nm², a $\langle A_{\text{Chol}} \rangle$ of 0.396 ± 0.005 nm², and a $\langle D_{p-p} \rangle$ of 4.52 ± 0.020 nm. In the same order, the difference between these values are 0.012 nm², 0.005 nm², and 0.043 nm which, while small, the changes are still outside the value ranges. An interesting observation in these systems is, by the 700 ns marker shown in Table 4.1, these values were within the error range of each other, implying that the effects of reduction are only temporary. This reduction test would need to be repeated to verify these results, but this is not a new observation. For computational studies, the removal of any molecule from a system will naturally lead to a temporary change in measured parameters, but since we could not find any recorded measure of QS21-A reduction on a system with a bilayer in the literature we still find it interesting to note the change in a graphical manner.

It was also previewed in Figure 4.6 (D), our H-bond analysis looked at how the QS21-A interacted with the lipids and the Chol of the bilayer and graphed over time. Based on the visually inspection of the systems, we found that the terminal sugars appeared to interact with the bilayer the most aside from when the QS21-A penetrated the bilayer. For a rough verification of this observation, we gathered the total number of interactions that took place during the 100 ns analysis for QS21-A and for the terminal sugars and found the percent value that belonged to the former. We also plotted the H-bonds between the four terminal sugars and the bilayer on the same graph as the overall QS21-A H-bonds. The summations are shown in Table 4.2 and, while a rough calculation to verify our visual results, it does confirm our visual observations that the terminal sugars make up most of the H-bonds that occur in the systems. 10-DOPC-SI was an outlier as a QS21-A molecule had penetrated the bilayer, allowing a larger range of components to interact with the bilayer.
Table 4.2: Summation of the Number of H-bonds Between QS21-A and the Bilayer

<table>
<thead>
<tr>
<th></th>
<th>1-DOPC</th>
<th>5-DOPC-SI</th>
<th>10-DOPC-SI</th>
<th>50-DOPC</th>
<th>1-DPPC</th>
<th>50-DPPC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lipid Total</td>
<td>16</td>
<td>176</td>
<td>339</td>
<td>238</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>D-Api/API</td>
<td>5</td>
<td>27</td>
<td>96</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L-Ara/ARA</td>
<td>0</td>
<td>3</td>
<td>90</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D-Gal/GAL</td>
<td>9</td>
<td>59</td>
<td>0</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D-Xyl/XY1</td>
<td>0</td>
<td>71</td>
<td>35</td>
<td>177</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Terminal Sugar Total (#)</td>
<td>14</td>
<td>160</td>
<td>221</td>
<td>225</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Terminal Sugar Total (%)</td>
<td>87.5</td>
<td>90.9</td>
<td>65.2</td>
<td>94.5</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Chol Total</td>
<td>0</td>
<td>30</td>
<td>187</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D-Api/API</td>
<td>0</td>
<td>1</td>
<td>22</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L-Ara/ARA</td>
<td>0</td>
<td>0</td>
<td>19</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D-Gal/GAL</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D-Xyl/XY1</td>
<td>0</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Terminal Sugar Total (#)</td>
<td>0</td>
<td>30</td>
<td>41</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Terminal Sugar Total (%)</td>
<td>0</td>
<td>100</td>
<td>21.9</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Next, the DPPC systems are of no interest in this regard with the graphs for 50-DPPC placed in Appendix E. 1-DPPC saw H-bonds between QS21-A and the bilayer while 50-DPPC had D-Xyl form a temporary H-bond with the DPPC headgroups only once in the entire 100 ns. The lack of H-bonds in the 1-DPPC system may be an indication that single QS21-A molecules have negligible interactions with membranes and that in order to interact with a membrane, QS21-A molecules have to first form an aggregate in solution. The single QS21-A systems were created with ~6 nm of water on either side of the bilayer to ensure proper hydration of the lipids and space the QS21-A to rotate and move. However, this also appeared to give the QS21-A enough room to avoid the bilayer throughout the 100 ns. This was likely just a random chance as the in the 1-DOPC system the QS21-A did make its way to the bilayer. Nevertheless, the lack of bilayer interactions in 1-DPPC is advantageous as we can use this QS21-A as a reference system of a single QS21-A in water which we will use in Section 4.4.

The DOPC systems showed more interactions than expected since 10-DOPC-SI had the penetration event, and the micelles in these systems stayed in closer proximity to the bilayer. We will note that, without more tests, we cannot say with accuracy if what we saw was the definite
method of penetration or even interaction with the bilayer. Our speculations will be based on what we saw in our systems only and as a means to explain what we saw.

In general, and as expected, more QS21-A molecules in the system meant more opportunities for creating H-bonds with the bilayer. 1-DOPC only had H-bonds form between QS21-A and the lipid headgroups. The penetration event of 10-DOPC-SI is shown at 700 ns in Figure 4.8 which also shows the shape the QS21-A took after penetration. Over time, we saw QS21-A reach this shape by wiggling through the DOPC lipids. The triterpene and branched trisaccharide domains remain on the exterior while the tetrasaccharide and acyl chain domains have entered the bilayer. The hydrophobic DMOS have migrated into hydrophobic center while the hydrophilic L-Ara has moved back towards the bilayer-water interface. All the sugars of the tetrasaccharide domain have worked their way around the PC backbones and Chol headgroups. Overall, the penetrated molecule visually had a twisted corkscrew appearance. Remembering Table 4.2, this shape did explain why the terminal sugars did not account for as many of the H-bonds as we saw in the other systems.

Along with Table 4.2, we can also use the data from Figures 4.9 to 4.11 to make an educated guess on the location of the terminal sugars within the micelle. For the terminal sugars to compose the majority of the QS21-bilayer H-bonds, these sugars trend to be located on the outside of the micelle or within close proximity of the outside. We also saw in Figure 4.11 that penetration did not continue in the 5-DOPC-SI and 50-DOPC systems. The 5-DOPC-SI system had the QS21-A molecule pulled out and into the micelle earlier in project, but there have been periods where there were temporary points of penetration before the QS21-A was pulled back into the bilayer. For 50-DOPC, the QS21-A remained within the bilayer until ~615 ns mark when it was pulled out and into the micelle of that system. Given this, the shape that the QS21-A took in the 10-DOPC-SI system was a contributing factor in the molecule remaining within the bilayer.

We also had questions on why the DOPC-Chol bilayer was the only one that saw penetration. Based on our results, the DOPC-Chol may have been the only one penetrated as the double bonds increase the APL, lipid packing, and effects the surface tension and the packing of the lipids. All of these may have made penetrating this bilayer easier than the DPPC-Chol bilayers. More images can be found in Appendix E.
Figure 4.8: The shape of the penetrating QS21-A molecule in 10-DOPC-SI where (A) is a sideview with the bilayer shown with DOPC in ice-blue lines, Chol in black lines, nitrogen are blue atoms, and phosphorus are tan atoms, (B) is the sideview without the bilayer, and (C) is the bird’s eye view without the bilayer.
Figure 4.9: The H-bonds between QS21-A and DOPC for (A) 1-DOPC and (B) 5-DOPC-SI over time.
Figure 4.10: The H-bonds between QS21-A and DOPC for (A) 10-DOPC-SI and (B) 50-DOPC over time.
Figure 4.11: The H-bonds between QS21-A and Chol for (A) 5-DOPC-SI, (B) 10-DOPC-SI, (C) 50-DOPC over time.
In summary, we saw that reducing the number of QS21-A molecules of 50-DOPC to create the 10-DOPC-SI and 5-DOPC-SI led to a temporary change in the membrane properties. While this result was expected, with the lack of QS21-A computational works in the published literature, it was interesting to be able to observe this change. We also saw that the terminal sugars are likely found mostly on the outside of the micelle since the majority of the H-bonds between QS21-A and the bilayer are with the terminal sugars. The only system that did not have this trend was the 10-DOPC-SI system, the same one with the penetration event. In this case, the penetrated QS21-A molecule has the quilliac acid and trisaccharide domain located on the outside while the triterpene and tetrasaccharide domains have moved into the bilayer in a twisted corkscrew formation.

### 4.4 QS21-A Properties

When investigating the properties relating to QS21-A only, we note that we did find one article related to the shape of the QS21-A micelle shown as Ref 12. Visually, we found that the micelles were not spherical in any of the systems. The micelle of 50-DPPC system, shown in Figure 4.12, is not spherical. Figure 4.13 shows the same with the other 50 QS21-A micelles.

![Figure 4.12:](image)

**Figure 4.12:** The QS21-A micelle of the 50-DPPC system using a bird’s eye view clearly showing the micelle is not spherical and possibly needing more time to reach its equilibrium shape.
Figure 4.13: The different QS21-A micelles of (A) 10-REF, (B) 50-REF, (C) 5-DOPC-SI, (D) 10-DOPC-SI, and (E) 50-DOPC. The purple atoms are the K⁺ counter-cations used in the systems. (B) and (E) indicate the systems with 50 QS21-A molecules needed more time to reach their equilibrium shape.

Comparing these results to the ones from Ref. [8], it does appear that QS21-A should form a single micelle, but the micelle is irregular or tubular shape. Our more irregular shapes for the 50 QS21-A systems shown in Figure 4.13 and in Table 4.3 do match this description. Nevertheless,
comparing the micelle shapes using $R_g$ only does not provide an accurate picture of how our micelles truly compare to each other. Our $R_g$ values are larger than what is reported in Ref. [8], but the differences in our simulations can explain this. First, we used an all-atom force field while they used an united-atom force field leading to shorter simulations. Second, the $R_g$ is dependent on the number of molecules in the micelles, and the difference between an all-atom and united-atom force field may affect this value as well. The shape the micelles took may also be affected by the placement of the QS21-A molecules during random insertion unless enough time has passed to negate this difference. Nevertheless, visually inspection does confirm that our micelles could use more time. The complete set of $R_g$ graphs can be found in Appendix E.

<table>
<thead>
<tr>
<th>System</th>
<th>$R_{g,x}$ (nm) Average</th>
<th>$R_{g,x}$ (nm) Std Dev</th>
<th>$R_{g,y}$ (nm) Average</th>
<th>$R_{g,y}$ (nm) Std Dev</th>
<th>$R_{g,z}$ (nm) Average</th>
<th>$R_{g,z}$ (nm) Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-REF</td>
<td>4.70</td>
<td>1.83</td>
<td>4.59</td>
<td>1.89</td>
<td>2.08</td>
<td>1.02</td>
</tr>
<tr>
<td>50-REF</td>
<td>6.50</td>
<td>0.40</td>
<td>6.82</td>
<td>0.64</td>
<td>3.28</td>
<td>0.94</td>
</tr>
<tr>
<td>1-DOPC</td>
<td>0.79</td>
<td>0.15</td>
<td>0.76</td>
<td>0.16</td>
<td>0.79</td>
<td>0.15</td>
</tr>
<tr>
<td>5-DOPC-SI</td>
<td>1.40</td>
<td>1.10</td>
<td>3.15</td>
<td>2.51</td>
<td>3.31</td>
<td>2.66</td>
</tr>
<tr>
<td>10-DOPC-SI</td>
<td>2.78</td>
<td>1.86</td>
<td>2.61</td>
<td>0.43</td>
<td>3.41</td>
<td>1.56</td>
</tr>
<tr>
<td>50-DOPC</td>
<td>4.53</td>
<td>1.00</td>
<td>5.45</td>
<td>0.63</td>
<td>6.71</td>
<td>1.03</td>
</tr>
<tr>
<td>1-DPPC</td>
<td>0.82</td>
<td>0.14</td>
<td>0.78</td>
<td>0.15</td>
<td>0.82</td>
<td>0.14</td>
</tr>
<tr>
<td>50-DPPC</td>
<td>5.76</td>
<td>0.88</td>
<td>6.00</td>
<td>0.70</td>
<td>5.75</td>
<td>0.94</td>
</tr>
</tbody>
</table>

**Table 4.3: $R_g$ and H-bond Average for the Systems**

<table>
<thead>
<tr>
<th>System</th>
<th>$R_g$ (nm) Average</th>
<th>$R_g$ (nm) Std Dev</th>
<th>H-bond to QS21-A Average</th>
<th>H-bond to QS21-A Std Dev</th>
<th>H-bond to Water Average</th>
<th>H-bond to Water Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>10-REF</td>
<td>4.97</td>
<td>1.75</td>
<td>57.62</td>
<td>5.95</td>
<td>382.62</td>
<td>13.68</td>
</tr>
<tr>
<td>50-REF</td>
<td>7.07</td>
<td>0.69</td>
<td>317.45</td>
<td>13.41</td>
<td>1795.30</td>
<td>33.72</td>
</tr>
<tr>
<td>1-DOPC</td>
<td>0.97</td>
<td>0.06</td>
<td>3.91</td>
<td>1.62</td>
<td>46.37</td>
<td>4.40</td>
</tr>
<tr>
<td>5-DOPC-SI</td>
<td>3.50</td>
<td>2.55</td>
<td>26.81</td>
<td>3.17</td>
<td>195.41</td>
<td>8.90</td>
</tr>
<tr>
<td>10-DOPC-SI</td>
<td>3.71</td>
<td>1.53</td>
<td>57.02</td>
<td>5.88</td>
<td>385.48</td>
<td>14.23</td>
</tr>
<tr>
<td>50-DOPC</td>
<td>6.91</td>
<td>1.03</td>
<td>316.49</td>
<td>13.33</td>
<td>1886.84</td>
<td>28.02</td>
</tr>
<tr>
<td>1-DPPC</td>
<td>1.00</td>
<td>0.05</td>
<td>3.00</td>
<td>1.44</td>
<td>48.02</td>
<td>3.78</td>
</tr>
<tr>
<td>50-DPPC</td>
<td>7.19</td>
<td>0.65</td>
<td>319.00</td>
<td>12.36</td>
<td>1861.63</td>
<td>29.10</td>
</tr>
</tbody>
</table>

Table 4.3 also includes data on how the QS21-A to QS21-A and QS21-A to water H-bonds change in the different systems. We noted in Section 4.3 that the 1-DPPC system could be treated as 1-REF because the QS21-A molecule did not interact with the DPPC-Chol bilayer and remained a significate distance from the bilayer. With these values, we can create a plot that compares how these change with a bilayer in the system. We also plotted these values over time.
for each system, but they reflect the same information presented in Table 4.3. Thus, we have placed these figures in Appendix E.

Before detailing the results of our investigation into the effect of a bilayer on QS21-A properties, we note again that more tests are needed with different bilayer compositions. Since there is so little information in the literature on this topic, we can only make speculations based on our own systems. Starting with Figure 4.14 and QS21-A to QS21-A H-bond numbers, the presence of a bilayer did not appear to have an effect on how the QS21-A H-bonded to itself or other QS21-A molecules. If the bilayer did affect these results, then it might imply that QS21-A preferred to interact with the DOPC-Chol bilayer than with other QS21-A molecules.

![Figure 4.14](image)

**Figure 4.14:** A comparison of the number of QS21-A to QS21-A H-bonds over the number of QS21-A molecules in the system in the presence and absence of a bilayer. Since the lines are overlapping with minute differences, the graph implies that bilayers do not affect the QS21-A to QS21-A H-bonding.

The presence of a bilayer did appear to affect the number of QS21-A to water H-bonds. Figure 4.15 shows that as the number of QS21-A molecules increased, the systems with bilayers increased the number of QS21-A to water H-bonds. This result is unexpected as we predicted that the opposite would happen. Our reasoning is that the presence of the bilayer meant fewer water molecules in the system, and thus fewer targets for QS21-A to water H-bonds. The results
began to make more sense since, as we noted before, this could have also been an artifact of the QS21-A placement within the systems and further simulations are needed to elaborate this issue.

![Graph](image)

**Figure 4.15:** A comparison of the average number of QS21-A to water H-bonds over the number of QS21-A molecules in the system in the presence and absence of a bilayer. While there is a difference in the values, this could have been caused by the random placement of the QS21-A molecules in the systems.

In summary, from the data gathered our results are in agreement with what is reported in Ref. [8] and our micelles needed more time. We also noted that the presence of a bilayer, specifically a DOPC-Chol bilayer, may not affect the interactions between QS21-A and itself or other QS21-A molecules.

### 4.5 References


Chapter 5
Concluding Remarks and Future Work

5.1 Gathered Conclusions

To the best of our knowledge, only one published computational study of QS21 exists [1]. In addition, there are only a few structural studies and even the MOA of QS21 remain largely unknown [2]. In the only published computational study, Pedebos et al. parameterized QS21 to be used with the united atom GROMOS96 force field and they focused on the behavior of QS21 in water solution only; the current results are the first published ones for QS21 interacting with membranes. We also provide a new all-atom parameterization using the OPLS-AA force field. Thus, the results presented in this thesis provide new experimentally testable predictions for QS21 and its interactions with model membranes. The force field parameterization will also be released to the public domain.

The reasons why there is such a limited number of studies using QS21 are many. From the computational perspective, the molecule is large and complex. The former implies that simulation of even a single QS21 requires a large number of water molecules for solvation. It also means that diffusion and conformational changes of QS21 are slow and thus long simulations are necessary. The latter is due to many previously not parameterized molecular groups and their interactions. The results from the single QS21-A systems were somewhat surprising in that the QS21-A in the 1-DPPC system negligible interacted with the bilayer. Due to the complex shape and large size of the molecule, it turned out that making any definite claims on how QS21-A interacts with the bilayer, the shape of QS21-A micelles, or the MOA of QS21-A is very difficult. These difficulties arise from the reason that for the large QS21 molecule to comprehensively probe the phase space would take prohibitively long times. We were, however, able to gain some new insights that, as far as we know, have not be recorded before. First, and importantly, we were able to witness a penetration event between a QS21-A molecule and a DOPC-Chol bilayer. This was suggested to be due to the double bonds affecting packing of the lipids. We have presented the first images of a possible configuration QS21-A takes after penetrating the bilayer. The first recordings of the effect of QS21-A reduction on membrane
properties were presented as well. The reduction did cause a temporary change in the membrane properties, and there is an upper limit on how much water there can be in the system before it becomes a hinderance for QS21-A to bilayer interactions and H-bonds. These points help to set guidelines for optimizing the simulation conditions for QS21-A and bilayer systems in the future. There is a possibility that the presence of a bilayer may not affect how QS21-A interacts with itself or other QS21-A molecules, but that more studies are needed to verify if this same presence increases the number of QS21-A to water interactions and H-bonds. Finally, the results in this thesis present the most computational modelling results to-date for QS21-A and bilayers and set a standard for future works.

5.2 Future Work

Firstly, we note that all future studies should start with the smallest possible systems; in terms of atomistic MD, QS21 is a large molecule and the smallest possible systems for QS21 exceed the sizes of typical membrane simulations. It would be fair to say that ~60-70% of our allotted time and resources was spent running the 50 QS21-A systems, leaving ~30-40% to work through the 5 and 10 QS21-A systems, 10 and 50 QS21-A references, and 1 QS21-A systems. We theorize that working with multiple smaller systems would lead to more results, a better opportunity to observe multiple penetration events, and possibly a better understanding of how QS21 interacts with the bilayers.

5.2.1 Testing the Effect of QS21 and Its Isomers

In this thesis, we studied bilayer interactions using different amounts of QS21-A molecules. However, QS21 exists as two isomers, even if most experiments seem to report results for QS21-A only. In future studies, we would like to include both isomers to investigate the similarities and differences between them. We mentioned in Chapter 1 that it is not known if QS21-A and QS21-X have different MOAs, but it is also not known how the isomers interact with each, how they collaborate in a single MOA, or the effect on the micelle when both isomers are present. As QS21 is effective in the monomer state, understanding how the two isomers differ and cooperate is useful for drug and vaccine design.
We would also like to gain more insight into how the QS21 isomers interact in a micelle with and without the bilayer. Having a wider range of QS21-A amounts could give better insight on the effect of the quantity of QS21 on interactions and H-bonding. While QS21 is active in the monomer state, storage is better in a micelle as it can offer protection to the sensitive ester bond between D-Fuc and DMOS [1, 3]. Investigating this process may lead to a better understanding of not only the conditions but also the optimal concentration to store QS21. It would also be the perfect chance to see how the two isomers behave in the micelle versus the bilayer environment. Systems that matched the natural ratio of 2:1 of QS21-A to QS21-X would be of the most interest, but we would also need to run pure QS21-A and pure QS21-X systems to act as the reference systems. We may also have to run a set of systems with a 1:1 ratio to study how the QS21-A majority changes the behavior.

5.2.2 Testing the Effect of Cholesterol

As QS21-A penetrated the bilayer to the Chol headgroup/lipid backbone area, it would be interesting to study the effect of varying ratios of Chol. The human body has Chol existing at 20% to 70% of the membrane [4-5]. By running series of systems with differing amounts of Chol in each, we would test whether QS21 penetration is based on the amount of Chol in the bilayer. These would be single QS21 systems to begin with until we understand monomer QS21 interactions better. Our goal would be to find the minimal Chol level for QS21 to penetrate, or if such level exists at all. Another possibility is that QS21 penetration is an all-or-nothing action where even the smallest amount of Chol will allow for penetration. It might also be that penetration is possible without Chol in the bilayer, but the process is eased or quickened when Chol is there.

5.2.3 Mixed Lipid Bilayer

Our use of simple bilayers was a good starting point given the level of literature on QS21-bilayer interactions. It should be kept in mind, however, that computational modeling does not allow simulations of the actual membranes in the human body that have tens or hundreds of lipid species but instead one is limited to model membranes. Similar limitations also exist for experimental studies of membranes. Current limitations are mostly due to computational capacity, but we can still work with bilayers composed of multiple types (typically 2-4) of the
most abundant and important lipids such as PCs and Chol. One future direction could be to use lipid structure found in dendritic cells to test the most current MOA theory for QS21 [6].

5.2.4 Free Energy Difference

QS21-A and QS21 are hydrophobic molecules, but we did not learn about whether it preferred the environment within the bilayer or within the micelle. Thus, we would look at the free energy difference between QS21 being in a micelle versus being in the bilayer. One possible method to study this would be Umbrella Sampling [7] combined with Weighted Histogram Analysis Method [8].

5.3 References

Appendix A: Verlet and Leapfrog Integrator

The Leapfrog algorithm [1] used in this thesis is derived from the Verlet algorithm [2] proposed in 1967. In the Verlet algorithm, the position \( \mathbf{r} \) at time \( t + \Delta t \), \( \mathbf{r}(t + \Delta t) \), is found using the position and acceleration at \( t \) as well as \( \mathbf{r}(t - \Delta t) \). We use Taylor expansion to find the value of \( \mathbf{r}(t + \Delta t) \) and \( \mathbf{r}(t - \Delta t) \),

\[
\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t \frac{d}{dt} \mathbf{r}_i(t) + \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) + \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t) + O(\Delta t^4) \tag{A.1}
\]

\[
\mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \Delta t \frac{d}{dt} \mathbf{r}_i(t) + \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) - \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t) + O(\Delta t^4) . \tag{A.2}
\]

These are summed together to obtain the position equation

\[
\mathbf{r}_i(t + \Delta t) + \mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) + \Delta t \frac{d}{dt} \mathbf{r}_i(t) - \Delta t \frac{d}{dt} \mathbf{r}_i(t) + \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) + \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) + \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t) - \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t) = 2 \mathbf{r}_i(t) + \Delta t^2 \frac{d^2}{dt^2} \mathbf{r}_i(t) \]

with \( \frac{d^2}{dt^2} \mathbf{r}_i(t) = \mathbf{a}_i \) for acceleration and \( m_i \mathbf{a}_i = \mathbf{F}_i \)

\[
\mathbf{r}_i(t + \Delta t) \approx 2 \mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \mathbf{F}_i . \tag{A.3}
\]

To get the velocities at time \( t \), or \( \mathbf{v}_i(t) \), we subtract Equation A.1 and A.2

\[
\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t) = \mathbf{r}_i(t) - \mathbf{r}_i(t) + \Delta t \frac{d}{dt} \mathbf{r}_i(t) + \Delta t \frac{d}{dt} \mathbf{r}_i(t) + \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) - \frac{\Delta t^2}{2} \frac{d^2}{dt^2} \mathbf{r}_i(t) + \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t) + \frac{\Delta t^3}{6} \frac{d^3}{dt^3} \mathbf{r}_i(t)
\]

with \( \frac{d}{dt} \mathbf{r}_i(t) = \mathbf{v}_i \) for velocity

\[
\mathbf{v}_i(t) \approx \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t - \Delta t)}{2\Delta t} . \tag{A.4}
\]

The Verlet algorithm of Equation A.3 and A.4 and is simple to program, efficient, time-reversible, and symplectic with a truncation error of \( \Delta t^4 \) for positions and \( \Delta t^2 \) for velocities. However, this algorithm is also sensitive to numerical precision and round-off errors. This is
caused by the method used for finding the velocities which uses the difference of two positions with the same magnitude.

The Leapfrog algorithm includes the velocity term in the equations and is also time-reversible and symplectic, but the velocity is found at every half-step. Finding the positions again,

\[ \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) \approx \frac{\mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t)}{\Delta t} \]
\[ \Delta t \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) \approx \mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) \]
\[ \Delta t \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) - \mathbf{r}_i(t) \approx -\mathbf{r}_i(t - \Delta t) \]
\[ \mathbf{r}_i(t) - \Delta t \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) \approx \mathbf{r}_i(t - \Delta t) \]
\[ \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \approx \frac{\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)}{\Delta t} \]
\[ \Delta t \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \approx \mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t) \]
\[ \Delta t \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) + \mathbf{r}_i(t) \approx \mathbf{r}_i(t + \Delta t) \]
\[ \mathbf{r}_i(t + \Delta t) \approx \mathbf{r}_i(t) + \Delta t \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right). \]  

(A.5)

Comparing these results to what was done with the Verlet algorithm, Equation A.6 is the new position equation used in the Leapfrog algorithm. To obtain the velocity equation, we substitute Equation A.5 and A.6 into A.3

\[ \mathbf{r}_i(t + \Delta t) \approx 2 \mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \frac{\Delta t^2}{m_i} \mathbf{F}_i(t) \]
\[ \mathbf{r}_i(t) + \Delta t \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \approx 2 \mathbf{r}_i(t) - \mathbf{r}_i(t - \Delta t) + \Delta t \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) + \frac{\Delta t^2}{m_i} \mathbf{F}_i(t) \]
\[ \Delta t \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \approx 2 \mathbf{r}_i(t) - \mathbf{r}_i(t) - \mathbf{r}_i(t) + \Delta t \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) + \frac{\Delta t^2}{m_i} \mathbf{F}_i(t) \]
\[ \mathbf{v}_i \left( t + \frac{\Delta t}{2} \right) \approx \mathbf{v}_i \left( t - \frac{\Delta t}{2} \right) + \frac{\Delta t}{m_i} \mathbf{F}_i(t). \]  

(A.6)

References

Appendix B: Ewald Summation and Particle-Mesh Ewald

We start by stating some assumptions we will work under as explained in Refs. [1,2]. First, we are working with a collection of $N$ total number of particles housed in a cubic system with a side length of $L$ and volume of $V = L^3$. Second, the system is charge neutral. Third, periodic boundary conditions (PBC) are applied. Fourth, the system and all PBC images are surrounded by an infinite dielectric constant medium. Fifth, the system and image particles will repel each other at short distances.

For this system, the aim is to find the Coulomb contribution

$$ U_{Coul} = \frac{1}{2} \sum_{i=1}^{N} q_i \Phi(r_i) $$

where the prefactor of half is to cancel double-counting, $q_i$ is the charge of atom $i$, and $\Phi(r_i)$ is the electrostatic potential at position $r$ for atom $i$ which is equal to

$$ \Phi(r_i) = \sum_{j=1}^{N'} \sum_{n \in \mathbb{Z}} \frac{q_j}{|r_{ij} + nL|^f} $$

with

$$ r_{ij} = r_i - r_j $$

and $n$ is the three-dimensional integer vector. The prime in the first summation means that the sum is done over all possible images and particles unless $j = i$ if $n = (0,0,0)$. Said in another way, we are including the interactions between each particle and its periodic images but not with itself. Moving forward, we will use Gaussian units for compact notation.

While Equation B.2 is a straightforward summation, the conditionally convergent nature of the equation makes it unusable. To solve this problem, P. P. Ewald used screening and compensating diffuse charges [1]. The former surrounds point charge $q_i$ as $-q_i$ that normally takes the form of a Gaussian distribution
\[ \rho_{\text{screen}}(r) = -q_i \left( \frac{\alpha}{\pi} \right)^{\frac{3}{2}} \exp(-\alpha r^2) \]  
(B.4)

with the width set to \( \sqrt{2/\alpha} \). The compensating charge distribution cancels out the screening charge, leaving the point charge behind. The compensating charge is the sum of all the screening charges

\[ \rho_{\text{comp}}(r) = \sum_{j=1}^{N} \sum_{n \in \mathbb{Z}} q_i \left( \frac{\alpha}{\pi} \right)^{\frac{3}{2}} \exp \left[ -\alpha \left| r - (r_j + nL) \right|^2 \right]. \]  
(B.5)

Unlike Equation B.2, Equation B.5 counts the self-energy or self-interaction term. However, this leads to the equation being a smoothly varying periodic function. The screening charge is solved in the real space using a direct summation. The compensating charge is solved in the Fourier space.

The compensating charge is done in the Fourier space as we need to solve for Poisson’s equation

\[ -\nabla^2 \Phi(r) = 4\pi \rho(r) \]  
(B.6)

and use the Fourier series

\[ f(r) = \frac{1}{V} \sum_{l=-\infty}^{\infty} \tilde{f}(k) \exp[i \mathbf{k} \cdot \mathbf{r}] \]  
(B.7)

\[ \mathbf{k} = \frac{2\pi}{L} \mathbf{l} \]  
(B.8)

where \( 1/V \) is the normalization factor, \( \mathbf{l} = (l_x l_y l_z) \) are the Fourier space lattice vectors, and \( \tilde{f}(k) \) are the Fourier coefficients

\[ \tilde{f}(k) = \int_{V} d r f(r) \exp[-i \mathbf{k} \cdot \mathbf{r}]. \]  
(B.9)

To bring Equation B.6 to the Fourier space, we start with \( \Phi(r) \) and the left side

\[ -\nabla^2 \Phi(r) = -\nabla^2 \left( \frac{1}{V} \sum_{k} \tilde{\Phi}(k) \exp[i \mathbf{k} \cdot \mathbf{r}] \right) \]

\[ = \frac{1}{V} \sum_{k} k^2 \tilde{\Phi}(k) \exp[i \mathbf{k} \cdot \mathbf{r}]. \]  
(B.10)
Doing the same with the right side gives

\[ 4\pi \rho(r) = 4\alpha \left( \frac{1}{V} \sum_k \tilde{\rho}(k) \exp[i k \cdot r] \right) \]

\[ = 4\alpha \frac{1}{V} \sum_k \tilde{\rho}(k) \exp[i k \cdot r] \]  \hspace{1cm} \text{(B.11)}

which combines with Equation B.10 to give the transformed Poisson’s equation

\[ \frac{1}{V} \sum_k k^2 \tilde{\Phi}(k) \exp[i k \cdot r] = 4\alpha \frac{1}{V} \sum_k \tilde{\rho}(k) \exp[i k \cdot r] \]

\[ k^2 \tilde{\Phi}(k) = 4\alpha \tilde{\rho}(k) \]  \hspace{1cm} \text{(B.12)}

We can now solve the compensating charge distribution using Equation B.12 after we perform a Fourier transform on Equation B.5

\[ \tilde{\rho}_{\text{comp}}(k) = \int_V \, dr \exp[-i k \cdot r] \tilde{\rho}_{\text{comp}} \]

\[ = \int_V \, dr \exp[-i k \cdot r] \sum_{j=1}^N \sum_{n \in \mathbb{Z}} q_j \left( \frac{\alpha}{\pi} \right)^3 \exp \left[ -\alpha |r - (r_j + nL)|^2 \right] \]

\[ = \int_{\text{all-space}} \, dr \exp[-i k \cdot r] \sum_{j=1}^N q_j \left( \frac{\alpha}{\pi} \right)^3 \exp \left[ -\alpha |r - r_j|^2 \right] \]

\[ = \sum_{j=1}^N q_j \left( \frac{\alpha}{\pi} \right)^3 \exp[-i k \cdot r_j] \exp \left[ \frac{-k^2}{4\alpha} \right] . \]  \hspace{1cm} \text{(B.13)}

We can now solve Equation B.12 by substituting Equation B.13 into it

\[ k^2 \tilde{\Phi}(k) = 4\alpha \tilde{\rho}(k) \]

\[ \tilde{\Phi}(k) = \frac{4\alpha}{k^2} \tilde{\rho}_{\text{comp}}(k) \]

\[ \Phi(k) = \frac{4\alpha}{k^2} \sum_{j=1}^N q_j \left( \frac{\alpha}{\pi} \right)^3 \exp[-i k \cdot r_j] \exp \left[ \frac{-k^2}{4\alpha} \right] . \]  \hspace{1cm} \text{(B.14)}

With Equation B.14 we move forward assuming that \( k = 0 \) is equal to 0, which is consistent with our assumptions, as Equation B.14 is defined only for \( k \neq 0 \) as a consequence of the conditional convergence the Ewald summation has. Using this equation also lets us find the first part of the Ewald summation by first finding
\[ \Phi(r) = \frac{1}{V} \sum_{k \neq 0} \tilde{\Phi}(k) \exp[ik \cdot r] \]

\[ = \frac{1}{V} \sum_{k \neq 0} \exp[ik \cdot r] \frac{4\alpha}{k^2} \sum_{j=1}^{N} q_j \left( \frac{\alpha}{\pi} \right)^2 \exp[-ik \cdot r_j] \exp \left[ \frac{-k^2}{4\alpha} \right] \]

\[ = \frac{1}{V} \sum_{k \neq 0} \sum_{j=1}^{N} 4\alpha \frac{k^2}{k^2} q_j \exp[ik \cdot r] \exp[-ik \cdot r_j] \exp \left[ \frac{-k^2}{4\alpha} \right] \]

\[ = \frac{1}{V} \sum_{k \neq 0} \sum_{j=1}^{N} 4\alpha \frac{k^2}{k^2} q_j \exp[ik \cdot (r - r_j)] \exp \left[ \frac{-k^2}{4\alpha} \right] \]

and then applying it to find \( U_{\text{Fourier}} \)

\[ U_{\text{Fourier}} = \frac{1}{2V} \sum_{i=1}^{N} q_i \Phi(r_i) \]

\[ = \frac{1}{2V} \sum_{i=1}^{N} q_i \sum_{k \neq 0} \frac{4\alpha}{k^2} \sum_{j=1}^{N} q_j \exp[ik \cdot (r_i - r_j)] \exp \left[ \frac{-k^2}{4\alpha} \right] \]

\[ = \frac{1}{2V} \sum_{i,j=1}^{N} 4\alpha \frac{k^2}{k^2} q_i q_j \exp[ik \cdot (r_i - r_j)] \exp \left[ \frac{-k^2}{4\alpha} \right] \]

\[ = \frac{1}{2V} \sum_{k \neq 0} \frac{4\alpha}{k^2} |\rho(k)|^2 \exp \left[ \frac{-k^2}{4\alpha} \right] \]

where Refs. [1, 2] go into detail on how \( \rho(k) \) is equal to

\[ \rho(k) = \sum_{i=1}^{N} q_i \exp[ik \cdot r_i] . \] (B.17)

Since the Fourier space part includes the self-energy or self-interaction term due to the use of Equation B.5, we need to remove it such that

\[ U_{\text{Real}} + U_{\text{Fourier}} - U_{\text{Self}} = U_{\text{Coul}} . \] (B.18)

The overcounting introduced by Equation B.5 is valued at

\[ \rho_{\text{over}}(r) = q_i \left( \frac{\alpha}{\pi} \right)^{\frac{3}{2}} \exp(-\alpha r^2) . \] (B.19)
As with $\rho_{\text{comp}}$, we can use Poisson’s equation to solve for the electrostatic potential $\Phi_{\text{over}}(r)$. By taking into account that $\rho_{\text{over}}$ is a Gaussian charge cloud with spherical symmetry, Poisson’s equation becomes

$$-\frac{\partial^2 r \Phi_{\text{over}}(r)}{\partial r^2} = 4\pi r \rho_{\text{over}}(r). \quad (B.20)$$

To continue, we will need the error function ($\text{erf}(\beta r)$) and complementary error function ($\text{erfc}(\beta r)$)

$$\text{erf}(\beta r) = \frac{2}{\sqrt{\pi}} \int_0^{\beta r} e^{-x^2} dx \quad (B.21)$$

$$\text{erfc}(\beta r) = 1 - \text{erf}(\beta r) = \frac{2}{\sqrt{\pi}} \int_{\beta r}^{\infty} e^{-x^2} dx. \quad (B.22)$$

Moving forward, performing a partial integration of Equation B.20 gives

$$-\frac{\partial r \Phi_{\text{over}}(r)}{\partial r} = \int_0^r dr \, 4\pi r \rho_{\text{over}}(r)$$

$$= -\int_0^r dr \, 4\pi r q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \exp(-\alpha r^2)$$

$$= -2\pi r q_i \left(\frac{\alpha}{\pi}\right)^{\frac{3}{2}} \int_r^\infty dr^2 \exp(-\alpha r^2)$$

$$= -2q_i \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \exp(-\alpha r^2) \quad . \quad (B.23)$$

We can perform another partial integration on Equation B.23 if we include the error function

$$r \Phi_{\text{over}}(r) = 2q_i \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} \int_0^r dr \exp(-\alpha r^2)$$

$$r \Phi_{\text{over}}(r) = q_i \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \int_0^r dr \exp(-\alpha r^2)$$

$$r \Phi_{\text{over}}(r) = q_i \left(\frac{2}{\pi}\right)^{\frac{1}{2}} \exp(-\alpha r^2)$$

$$\Phi_{\text{over}}(r) = \frac{q_i}{r} \text{erf}(\sqrt{\alpha r}) \quad . \quad (B.24)$$

and when $r = 0$

$$\Phi_{\text{over}}(r) = \Phi_{\text{self}}(r) = 2q_i \left(\frac{\alpha}{\pi}\right)^{\frac{1}{2}} . \quad (B.25)$$
Equation B.25 is then used to find $U_{Self}$

$$U_{Self} = \frac{1}{2} \sum_{i=1}^{N} q_i \Phi_{self}(r_i)$$

$$= \frac{1}{2} \sum_{i=1}^{N} q_i 2q_i \left( \frac{\alpha}{\pi} \right)^{\frac{1}{2}}$$

$$= \frac{1}{2} \sum_{i=1}^{N} q_i^2 \left( \frac{\alpha}{\pi} \right)^{\frac{1}{2}}$$

$$= \left( \frac{\alpha}{\pi} \right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2$$  \hspace{1cm} (B.26)

and complete the Fourier space part of the Ewald summation.

For the real space component of the summation, we use the complementary error function and Equation B.24 to get $\Phi_{Real}$ for $q_i$ as it is surrounded by the screening charge $-q_i$

$$\Phi_{Real}(r) = \frac{q_i}{r} - \frac{q_i}{r} \text{erf}\left(\sqrt{\alpha r}\right)$$

$$\Phi_{Real}(r) = \frac{q_i}{r} \text{erfc}\left(\sqrt{\alpha r}\right)$$  \hspace{1cm} (B.27)

and find $U_{Real}$ using the same method as before

$$U_{Real} = \frac{1}{2} \sum_{i \neq j}^{N} \left( \frac{q_i q_j}{r_{ij}} \text{erfc}\left(\sqrt{\alpha r_{ij}}\right) \right).$$  \hspace{1cm} (B.28)

We can now combine Equation B.16, B.26, and B.28 to find $U_{Coul}$

$$U_{Real} + U_{Fourier} - U_{Self} = U_{Coul}$$

$$\frac{1}{2} \sum_{i \neq j}^{N} \left( \frac{q_i q_j}{r_{ij}} \text{erfc}\left(\sqrt{\alpha r_{ij}}\right) \right) + \frac{1}{2V} \sum_{k \neq 0}^{4\alpha} |\rho(k)|^2 \exp \left[ -\frac{k^2}{4\alpha} \right] - \left( \frac{\alpha}{\pi} \right)^{\frac{1}{2}} \sum_{i=1}^{N} q_i^2 = U_{Coul}. \hspace{1cm} (B.29)$$

The computational cost of the Ewald summation is $O(N^2)$ but it can be reduced to $O(N^{3/2})$. This makes the summation impossible to apply to larger systems. The $O(N^{3/2})$ is, however, still computationally restrictive, but by using Fast-Fourier transforms, the Ewald summation can be modified to create a method called Particle-Mesh Ewald [4-5] which reduces the computational cost to $O(N\log N)$. This is a result of restricting the spatial solution to lattice points, or by solving the problem after the charges have been mapped onto a grid using a Gaussian spread function.
Afterwards, Lagrange interpolation is used to remove the solutions from the grid and assign back to the particles.

**References**


Appendix C: Berendsen and Nosé-Hoover Thermostat

For the Berendsen thermostat [1], let us first consider the Langevin equation [2] that describes the coupling of a system to a heat bath with a fixed reference temperature \( T_0 \)

\[
m_i \dot{v}_i = F_i - m_i \gamma_i v_i + R(t) \tag{C.1}
\]

where \( m_i \) is the mass of atom \( i \), \( \dot{v}_i \) is the first derivative of atom \( i \)'s velocity, \( F_i \) is the force on atom \( i \), \( \gamma_i \) is the damping constant that determines the coupling strength to the heat bath, \( v_i \) is the velocity, and \( R_i \) is a Gaussian stochastic variable with zero mean. The Berendsen thermostat works by rescaling the velocities via weak coupling to an external bath. This affects the kinetic energy which changes the temperature of the system. The equation of motion for the system is changed to

\[
m_i \dot{v}_i = F_i - m_i \gamma_i \left[ \frac{T_0}{T - T_0} \right] v_i \tag{C.2}
\]

with \( T \) as the measured temperature. The rescaling is a proportional scaling per \( \Delta t \) from \( v \) to \( \lambda v \) where \( \lambda \) is the scaling factor

\[
\lambda = \left[ 1 + \frac{\Delta t}{\tau_T} \left( \frac{T_0}{T} - 1 \right) \right]^{\frac{1}{2}} \tag{C.3}
\]

where \( \tau_T \) is the temperature relaxation time. Ref. [1] shows how Equation C.3 is the same as correcting the measured temperature according to

\[
\frac{dT}{dt} = \frac{T_0 - T}{\tau_T}. \tag{C.4}
\]

Equations C.2 and C.4 show that when \( \tau_T \) is very large Equation C.2 becomes Newton’s equation of motion. The weak-coupling leads to the strength of the coupling being as weak as possible, but this thermostat has trouble sampling the canonical ensemble [4]. If \( \tau_T \) is very small, then it can sample the canonical ensemble.
The thesis uses the Nosé-Hoover thermostat [3, 5] which works by extending the system’s Hamiltonian the two extra variables of a thermal reservoir and a friction term in the equations of motion

\[ H_{\text{Nosé–Hoover}} = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + V(r^n) + \frac{\xi^2 Q}{2} + L \frac{\ln s}{\beta} \]  

(C.5)

where \( p_i \) is the momentum of atom \( i \), \( V \) is the potential energy, \( \xi \) is the thermodynamic friction coefficient related to the extra coordinate \( s \), \( Q \) describes the coupling strength as the mass of \( s \), \( L \) is the degrees of freedom in the extended system, and \( \beta = 1/(k_B T) \). This thermostat is able to sample the canonical ensemble and is time-reversible [3].

**References**


Appendix D: Berendsen and Parrinello-Rahman Barostat

The Berendsen barostat [1] follows the same principle as the Berendsen thermostat [1] where control over the variable in question is done through rescaling via weak coupling to an external source. Here, the external source is a pressure bath, or piston, and we are rescaling the atom position \( r_i \) and box size \( L \) to control the pressure

\[
\begin{align*}
  r_i &\rightarrow \mu r_i \\
  L &\rightarrow \mu L 
\end{align*}
\]

where \( \mu \) is the scaling factor

\[
\mu = \left[ 1 + \frac{\Delta t}{\tau_P} (P_0 - P) \right]^{1/3}
\]

with \( \tau_P \) as the pressure relaxation time, \( P_0 \) is the reference pressure, \( P \) is the measured pressure. To find \( \mu \) we start at the pressure correction, which is similar to the temperature correction

\[
\frac{dP}{dt} = \frac{P_0 - P}{\tau_P} .
\]

An isotropic system will have \( P \) equal to

\[
P = \frac{2}{3V} (E_{kin} - \Xi) .
\]

where \( V \) is the measured volume, \( E_{kin} \) is the kinetic energy, and \( \Xi \) is the pair-additive virial

\[
\Xi = -\frac{1}{2} \sum_{i<j} r_{ij} \cdot F_{ij}
\]

with the force on atom \( i \) due to atom \( j \) as \( F_{ij} \), and

\[
r_{ij} = r_i - r_j .
\]

The barostat changes the equation \( \dot{r} = v \) to

\[
\dot{r} = v + \alpha r
\]

where \( \alpha \) is a coefficient related to \( \mu \). We will need the isothermal compressibility \( \beta \) related to the change of pressure
\[ \beta = -\frac{1}{V} \left( \frac{dV}{dP} \right) \]
\[ = -\frac{1}{V} \frac{dV}{dt} \cdot \frac{dP}{dt} \]
\[ \frac{dP}{dt} \frac{dP}{dt} \]
\[ \frac{dP}{dt} = -\frac{1}{\beta V} \frac{dV}{dt} \] (D.9)

and the equation

\[ \dot{V} = 3\alpha V \]
\[ \frac{dV}{dt} = 3\alpha V \]
\[ \frac{1}{V} \frac{dV}{dt} = 3\alpha \] . (D.10)

We can find \( \alpha \) by first combining Equation D.9 and D.10

\[ \frac{dP}{dt} = -\frac{1}{\beta V} \frac{dV}{dt} \]
\[ \frac{dP}{dt} = -\frac{3\alpha}{\beta} \]
\[ \frac{dP}{dt} = -\frac{3\alpha}{\beta} \] (D.11)

and substituting Equation D.11 into D.4 yields

\[ -\frac{3\alpha}{\beta} = \frac{P_0 - P}{\tau_p} \]
\[ 3\alpha = -\frac{\beta(P_0 - P)}{\tau_p} \]
\[ \alpha = -\frac{\beta(P_0 - P)}{3\tau_p} \] . (D.12)

Compressibility enters the algorithm only in conjunction with the time constant, so the exact value is not necessary. We recommend Ref. [1] for a thorough explanation.

This thesis used the Parrinello-Rahman barostat [2-4] because it makes each unit vector of the unit-cell or simulation box independent [2-4] and gives control over stress and pressure. This allows the size and shape of the box to change. As with the Nosé-Hoover thermostat [5,6], it extends the system’s Hamiltonian but in a much more complex form. Refs. [2-4, 5] have information on how is implemented.
References

1) Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F.; Dinola, A.; Haak, J. R. 


Appendix E: Additional Figures from Chapter 4

The following are the figures, both images and graphs, that did not make it into the main body of the thesis. Their order in this appendix matches the order that they would have appeared in Chapter 4.

System After Pre-Equilibration

System at End Point

1-DPPC

50-DPPC

Figure E.1: Images showing the 1-DPPC and 50-DPPC systems after pre-equilibration and at 900 ns and 700 ns respectively.
**Figure E.2:** (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 50-DPPC.
Figure E.3: Images showing 1-DOPC, 50-DOPC, 5-DOPC-SI, and 10-DOPC-SI systems after pre-equilibration and at 900 ns and 700 ns respectively.
Figure E.4: (A) the area per lipid values for (B) membrane thickness, and (C) lipid order parameter for 5-DOPC-SI.
Figure E.5: (A) The area per lipid values, (B) membrane thickness, and (C) lipid order parameter for 50-DOPC.
Figure E.6: The H-bonds between QS21-A and (A) the lipids and (B) the Chol of the 50-DPPC system over time.
Figure E.7: The penetrated QS21-A molecule (A) at the start of penetration 200 ns, (B) at 700 ns in the 10-DOPC-SI system, and (C) with nitrogen, blue atoms, phosphorus, tan atoms, lipid backbone oxygens, gold atoms, and Chol headgroup oxygen, red atoms, shown to illustrate the location of the molecule.
Figure E.8: Images showing the 10-REF and 50-REF systems after pre-equilibration and at 800 ns and 900 ns respectively. The end point image of 50-REF has periodic conditions on for a better image of the micelle.
Figure E.9: The $R_g$ graphs of the micelles in (A) 10-REF and (B) 50-REF.
Figure E.10: The $R_g$ graphs of the single QS21-A molecules in (A) 1-DOPC and (B) 1-DPPC. This was done to observe how a single QS21-A changed over time.
Figure E.11: The $R_g$ graphs of the micelles in (A) 5-DOPC-SI and (B) 10-DOPC-SI.
Figure E.12: The $R_g$ graphs of the micelles in (A) 50-DOPC and (B) 50-DPPC.
Figure E.13: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 1-DOPC system.
Figure E.14: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 5-DOPC system.
Figure E.15: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 10-DOPC-SI system.
Figure E.16: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-DOPC system.
Figure E.17: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 1-DPPC system.
Figure E.18: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-DPPC system.
Figure E.19: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 10-REF system.
Figure E.20: The number of H-bonds over time for (A) QS21-A to QS21-A, and (B) QS21-A to water for the 50-REF system.
Curriculum Vitae

Name: Sarai Guerrero

Post-secondary Education and Degrees: The University of Western Ontario, London, Ontario, Canada

Degrees: 2012-2017 B.Sc.

Honors and Awards: Dean’s Honor List 2017

Related Work Experience: Teaching Assistant The University of Western Ontario 2017-2019

Presentations:
102nd Canadian Chemistry Conference and Exhibition
Québec, Québec, Canada
June 3-7, 2019

2019 Chemical Biophysics Symposium
Toronto, Ontario, Canada
May 3-5, 2019

Centre for Advanced Material and Biomaterials Day
London, Ontario, Canada
April 17, 2019

Western Research Forum 2019
London, Ontario, Canada
March 22, 2019

Electrochemistry Society Student Chapter’s 2nd Research Symposium
London Ontario, Canada
December 12, 2018