Multiscale modelling of bacterial and viral sanitisation

A combined molecular dynamics and nuclear magnetic resonance study

C. Waller

March 2024

Abstract

The cell envelope is a pathogen's first defence against potentially hazardous materials in the immediate environment. Because of this, pathogens evolved to control which chemicals are allowed in and out of the cell selectively. New cell envelope adaptations mean our current sanitisation methods are becoming less effective daily. Lipid compositions, protein mutations, and lipopolysaccharide (LPS) in the outer membrane of Gram-negative bacteria are but a few features that pathogens have developed to help them resist current antimicrobials. Due to the importance of the cell envelope in the survivability of a pathogen, it has become a widespread target of sanitisers, specifically the lipid bilayer or essential proteins. A detailed understanding of these areas is critical if antimicrobial resistance one day renders all of our current methods ineffective. If this were to happen, we must understand why they worked initially, as this will inform our search for the adaptations responsible for resistance. By identifying the areas where these adaptations are found, scientists can engineer new, tailored antimicrobials. These tailored antimicrobials could exploit the properties of the adapted bilayer or new mutations in membrane proteins to induce deformation and disable or destroy them.

In this thesis, Chapters 4 and 5 explored the effect of short-chained alcohols and chlorhexidine (CHX) on phospholipid (PL) membranes with and without LPS using a multidisciplinary approach. I found that the application of sanitising alcohols at concentrations as low as 20% had a significant deforming effect on the cell membrane (SaCM) of Staphylococcus aureus (*S. aureus*), as well as the inner (EcIM) and outer (EcOM) membrane of Escherichia coli (*E. coli*). Furthermore, the application of short-chain alcohols created disruptions in the headgroup region, which allowed CHX to bind more effectively with PL leaflets. In contrast, this same disruption led to budding in the PL leaflet of the EcOM. I validated the simulation data by performing equivalent nuclear magnetic resonance (NMR) experiments. In this way, simulations added visual context to changes in the spectra obtained from NMR experiments. These experiments assessed whether the prevalent effects in the simulations were replicated in vitro beyond computationally accessible time scales. Said NMR experiments found that the CHX's binding position and conformation in molecular dynamics (MD) were consistent with those in vitro. In conjunction, these works showed that although short-chained alcohols only underwent transient interactions with the headgroup-tail interface of membranes, this overall aggregation was essential in the initial phases of membrane sanitisation. Furthermore, this allowed CHX to adopt a "c-shape" binding conformation with its termini embedded in the same interface region.

In chapters 6 and 7, the effect of sanitising agents on proteins and the intricacies of protein-protein interaction during SARS-CoV-2 host infection were explored in all-atom (AA) and coarse grain (CG) resolution, respectively. By applying a CG resolution to large cross-linking systems, I was able to simulate these systems for very long timescales. On the other hand, the AA system was simulated for far less time because this provided a significantly greater resolution, enabling me to explore specific interactions in further detail. The dynamics of the SARS-CoV-2 spike (S) protein were explored in a two-step approach: firstly, by detailing the impact of sanitising alcohols on the ectodomain (ECD) of the S protein in AA detail and secondly, by simulating cross-linking of different-sized S protein and angiotensin-converting enzyme 2 (ACE2) receptor clusters. Simulating this area in these separate approaches allowed for assessing structural and conformational stability. In chapter 6, it was found that, similarly to alcohol-membrane systems, despite the transience of interactions, the deformation process was less stochastic than it initially appeared. This hypothesis was first formed when assessing the trends in amino acids, which tended to see more interactions with alcohol in the surrounding solution. These amino acids tended to be large and hydrophobic with charged or polar regions, not dissimilar to the headgroup-tail interface of a membrane. The properties of amino acids that interacted more with short-chain alcohol suggested that similar to the membrane, short-chained alcohols were partitioned into hydrophilic-hydrophobic interface regions. These interactions resulted in an overall rigidifying of the trimer. Simulation of S-ACE2 cross-linking in Chapter 7 found that higher coordinations were generally less stable but possible. The S trimer's conformational restrictions were seen to restrict its ability to bind to multiple receptors. Tilting in the neck region and an observed rigid-body rotation in the receptor binding domain (RBD) meant that the protein was significantly less flexible and placed particular conformational requirements on how these proteins bind. Ultimately, this resulted in high-order binding being substantially less favourable and suggested that this would not occur in vitro due to the energetic penalties incurred by deviation from the low-order cross-linking conformation.

Contents

List of Tables

List of Figures

Declaration of authorship

I, Callum Waller, declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

Multiscale modelling of bacterial and viral sanitisation

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;

2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;

3. Where I have consulted the published work of others, this is always clearly attributed;

4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;

5. I have acknowledged all main sources of help;

6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;

7. Either none of this work has been published before submission, or parts of this work have been published as:

Chapters 1 and 2: C. Waller *et al.* "Impact on *S. aureus* and *E. coli* membranes of treatment with chlorhexidine and alcohol solutions: insights from molecular simulations and nuclear magnetic resonance.",J. Mol. Biol., 2022, 21, pp. 1-39.

Signed: .....
Date: .....

Acknowledgements

I would first like to thank my supervisors, Prof. Syma Khalid, Prof. Phil Williamson and Senior Principal Investigator Pete Bond. Syma and Phil were invaluable in the initial stages of this PhD, and their insight and enthusiasm for their respective areas allowed me to form a good basis for my research, for which I am extremely grateful. Pete's assistance allowed me to progress my project in directions I was interested in while providing discussion to ensure I maintained relevance and was well-equipped to continue working in an unfamiliar country.

Secondly, I would like to thank the members of the research groups in which I worked, both past and present, for creating such a great working environment, even during the pandemic. Their discussion was extremely helpful in problem-solving, developing my project and maintaining sanity during the more challenging periods of the project.

I would like to thank my family and friends, those I've known for a long time and those I met during my studies. Whether providing advice, discussion, or just keeping me company on Discord while I was writing or working to meet deadlines.

Finally, I would like to acknowledge the support of Laura Phillips. If moving halfway around the world so I could complete my research isn't worthy of acknowledgement, then I don't know what is...

To Laura

How's this for complicated biophysics?

Definitions and abbreviations

AA All-atom

ACE2 Angiotensin-converting enzyme 2

AI Artificial intelligence

ADP Adenosine diphosphate

APL Area per lipid

ATP Adenosine triphosphate

BGU Biguanides

CET Cryo-electron tomography

CG Coarse grain

CHL Cholesterol

CHX Chlorhexidine

CHG Chlorhexidine gluconate

CL Cardiolipin

CPL Chlorophenyl

CSA Chemical shielding anisotropy

DNA Deoxyribonucleic

DTR Decision tree regression

E Envelope protein

ECD Ectodomain

E. coli Escherichia coli

EcOM *E. coli* outer membrane

EcIM *E. coli* inner membrane

ERGIC Endoplasmic reticulum Golgi intermediate compartment

FID Free induction decay

FP Fusion peptide

Gal Galactose

Glc Glucose

GLUC Gluconate

HDX Hydrogen-deuterium exchange

Hep Heptose

HEX Hexane

ISOP Isopropanol

Kdo Ketodeoxyoctonic acid

LC-MS Liquid-chromatography-mass spectrometry

LOS Lipooligosaccharide

LPS Lipopolysaccharide

LPG Lyso-phosphatidyl-glycerol

M Membrane protein

Man Mannose

MAS Magic angle spinning

MD Molecular dynamics

ML Machine learning

MSD Mean squared diffusion

NMR Nuclear magnetic resonance

NOE Nuclear Overhauser effect

NOESY Nuclear Overhauser effect spectroscopy

P2 Phosphatidyl-inositol-bisphosphate

PC Phosphatidyl-choline

PE Phosphatidyl-ethanolamine

PG Phosphatidyl-glycerol

PL Phospholipid

PME Particle mesh Ewald

PROH Propanol

PS Phosphatidyl-serine

RBD Receptor binding domain

RBM Receptor binding motif

RF Radiofrequency

RFT Random forest

Rhm Rhamnose

RNA Ribonucleic acid

S Spike protein

SaCM *S. aureus* cell membrane

SARS-CoV-2 Severe acute respiratory syndrome

SASA Solvent accessible surface area

S. aureus Staphylococcus aureus

SL Sphingolipid

SM Sphingomyelin

VDW Van der Waals

VMD Visual molecular dynamics

VOC Variant of concern enzyme 2

Chapter 1

Introduction

1.1 Biological cells

Cells are the fundamental components of all living organisms and can be specialised to perform specific functions as part of a collective or individually. Two of the fundamental cellular organisations are prokaryotes and eukaryotes[1]. The distinction between these two different types of cells is that a eukaryote contains a nucleus and membrane-bound organelles, while prokaryotes contain neither of these features. To depict the simplicity of a prokaryotic cell, a simplified representation of a prokaryotic and eukaryotic cell are shown in Figure 1.1. Bacterial cells, one of the main research subjects in this thesis, belong to the category of prokaryotic cells. These pathogens present a significant challenge in developing antimicrobial treatments due to the vast diversity in their membrane's lipid and protein components[2]. As such, it is crucial to stay informed about the latest advancements in antimicrobial treatments and take preventive measures to prevent the spread of pathogens and safeguard overall health.

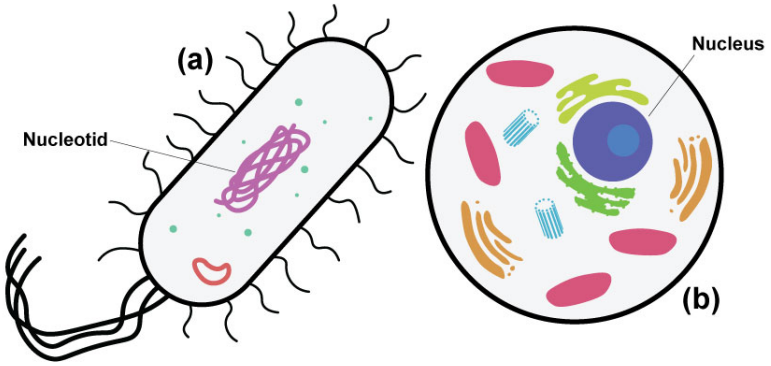


Figure 1.1: A simplified depiction of a (a) prokaryote and (b) eukaryote (not to scale).

1.1.1 Biological cell membranes

Biological membranes serve as the first line of defence for cells, protecting them from harmful environmental chemicals. The plasma membrane is primarily formed by a double layer of phospholipids (PLs), which provide a physical barrier between the interior and exterior of the

cell. The membrane also contains crucial proteins that maintain cell structure, regulate cell contents, and enable pathogens to infect hosts[3, 4, 5, 6]. The composition of lipids and proteins in these membranes varies widely among different membranes and different species, allowing cells to become well-specialised in the environment in which they exist. Prokaryotic membrane regions, such as those of bacteria, differ from eukaryotic ones in multiple ways, as the plasma membranes of prokaryotes are substantially simpler than those of eukaryotes. Eukaryotes often contain a wider range of lipids and include cholesterol (CHL), which is absent in prokaryotic membranes[7, 8]. Prokaryotic plasma membranes, on the other hand, generally consist of a smaller variety of types of PL and membrane-bound proteins. This simplicity of prokaryotic membranes is visible in their proteins and transport mechanisms, which are simpler and less comprehensive than those of eukaryotes due to their lack of membrane-bound organelles. One of the most prominent examples of a prokaryote is bacteria, one of the research focuses in this thesis. Bacteria alone present a wide variety of variations in lipids and proteins present in the cell membrane, even among different strains of the same bacteria. This wide variation in lipid and protein composition presents a significant challenge when developing effective broad-spectrum antimicrobial treatments as mutations often arise specifically to coutneract them[9, 10, 11].

1.2 Bacteria

Bacteria are single-celled microorganisms and, as mentioned previously, prokaryotes due to their lack of a membrane-bound nucleus and other organelles. Due to their relatively simple yet efficient structure, they are abundant in nature and thrive in many environments. Said abundance is primarily due to their ability to rapidly replicate, evolving to exist in some of the most inhospitable environments on Earth. These environments consist of, but are not limited to, extreme temperature, high pressure, extreme pH, high salinity, high radiation, low nutrients and general harsh chemical environments[12]. Bacteria have also evolved to exist parasitically, and some species can pose significant risks to humans if left untreated[13]. Generally, bacteria can be categorised into several structures which they can take: Coccus (spherical, Figure 1.2(a)), bacillus (rod-shaped, Figure 1.2(b)) or spiral/twisted shaped (Figure 1.2(c)). The

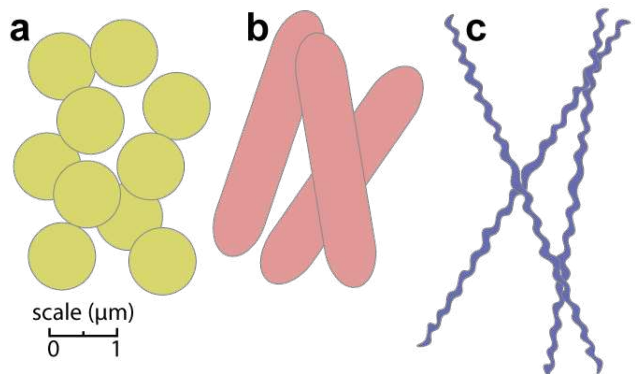


Figure 1.2: Examples of the three most common bacterial shapes, showing (a) coccus, (b) bacillus, and (c) spiral/twisted shaped

role of bacteria in various ecosystems and fields such as medicine and biotechnology means we must understand their characteristics, behaviours and potential impacts. The structures shown in Figure 1.2 are by no means exhaustive of bacterial structures. The structures in this figure are fundamental to bacterial taxonomy and classification; however, further variation exists within these and other bacterial shapes[14].

1.2.1 Bacterial cell envelopes

Although significantly simpler than eukaryotic membranes, the cell envelope of bacteria is still complex[15]. PL membranes significantly contribute to the bacterial cell envelope, forming the primary partition between the cell and its environment[9]. The importance of these structures is two-fold: firstly, they provide structural stability and keep harmful chemicals from entering the cell, and secondly, they can supply fluidity while maintaining the stability necessary to exist in the cell's natural environment[16]. Although the need for these structures as a partition is obvious, the need for species-specific fluidity is less intuitive. Membrane fluidity can specialise bacterial species to survive environmental factors such as temperature, pH and toxicity.

Membrane-bound proteins are crucial to the function of bacterial cell envelopes. These proteins perform many functions, including the transportation of molecules, removal of harmful chemicals, production of energy, and even infectivitye[17]. One of the most notable antibacterial resistance proteins is the efflux pump[18]. The efflux pump is a large transport protein found in the envelope region of many strains of bacteria. This transport protein actively removes antimicrobials from the cell by transporting them out. Although this mechanism is generally insufficient to resist bacteriocidal concentrations of antimicrobials, it has proven to be an effective mutation in resisting therapeutic concentrations of antimicrobials[19]. Similarly, the PACE transporter protein is found in bacterial membranes. The expression of this protein (PA2880) was found to be responsible for the efflux of CHX via proton motive force[20]. The presence of this protein is directly responsible for CHX resistance and is another example of emerging antimicrobial resistance in bacterial proteins.

1.2.1.1 Peptidoglycan and Gram staining

Bacteria are further subdivided by whether they are Gram-positive or -negative. A critical difference between these is the quantity of peptidoglycan present within their cell envelope, as Gram-positive bacteria contain a substantially thicker cell wall[6]. Peptidoglycan is a large, rigid, covalently bonded structure essential for cell structure and stability[21]. Due to the abundance of this structural feature in Gram-positive bacteria, it is possible to perform a simple test to distinguish between these types of bacteria with and without the exceptionally thick cell wall present in Gram-positive bacteria.

The Gram stain technique was developed in 1884 by H. C. Gram and is still used today[22]. A crystal violet dye is added to a sample during the Gram staining test. Gram-positive and -negative bacteria will take up this dye. The sample is then washed with a decolourising solution, which dissolves the PL component of the bacterial cell envelopes. A counterstain (e.g., safranin or fuchsine) is added, and only Gram-positive bacteria will retain the stain

colour[23].

The stain colour is retained in Gram-positive bacteria as the cell wall is over ten times thicker (20–80 nm) than in Gram-negative (1–2 nm)[24]. The cell wall contains pores as wide as 60 nm, which span the majority of, if not the entire, cell wall[25]. When the peptidoglycan cell wall becomes dehydrated, large pores in this region will close, retaining the dye[26]. The presence of peptidoglycan is far more significant in Gram-positive bacteria because it makes up the entirety of the envelope's structural component as they do not possess a LPS containing outer membrane, whereas Gram-negative bacteria do[6].

1.2.1.2 Gram-negative and -positive envelopes

In this thesis, Gram-negative and -positive bacteria are studied; however, cell envelope differs substantially between Gram-negative and -positive bacteria (Figure 1.3). The envelope region of Gram-negative bacteria (e.g. *E. coli*) possesses an inner and outer membrane, separated by a periplasmic space containing the peptidoglycan cell wall[9]. The periplasm is an aqueous region densely packed with osmolytes, soluble proteins and other essential components[27]. The outer membrane has an asymmetrical lipid distribution between leaflets and generally contains beta-barrel protein structures[17]. The inner leaflet of the outer membrane is composed of a mixture of PLs, and the outer is primarily LPS. The inner membrane of Gram-negative bacteria generally contains helical structure proteins[28]. Until recently, the EcIM was thought to have a symmetrical distribution of lipids. However, this has recently been shown to be untrue[29]. Bogdanov *et al.* found that the distribution between leaflets in the EcIM is asymmetric. Furthermore, this distribution correlates with whether the bacteria are coccus or bacillus.

The envelope structure of Gram-positive bacteria (e.g. *S. aureus*) differs from that of Gram-

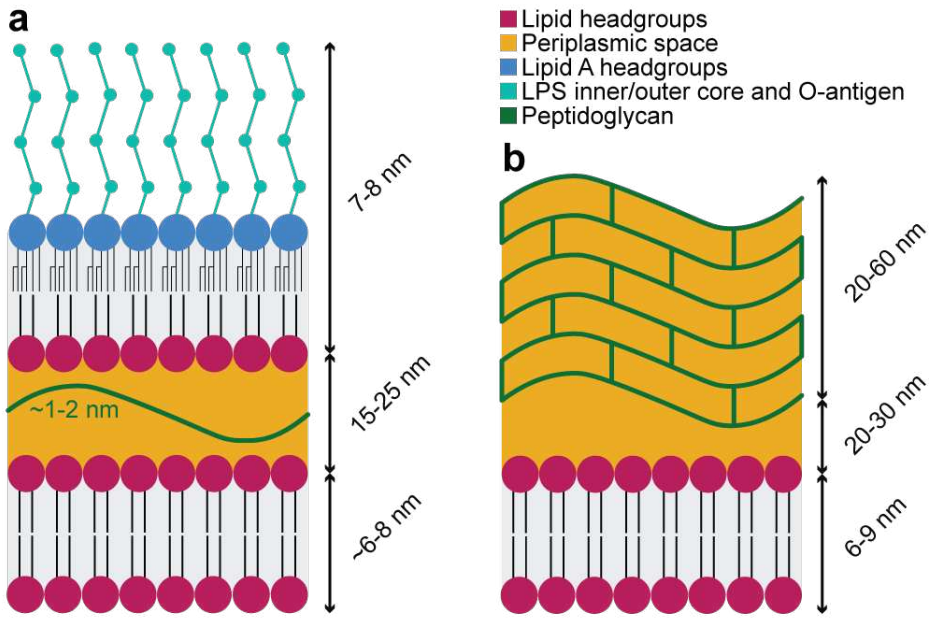


Figure 1.3: Cross-sections representations of the (a) Gram-negative and (b) Gram-positive cell envelopes.

negative. A key difference is that Gram-positive bacteria do not possess any outer membrane[6]. To compensate for this lack of structure, the envelope of Gram-positive bacteria possesses a thick peptidoglycan cell wall, responsible for a positive Gram-staining result. This cell wall is a rigid, covalently bonded structure spanning 20-60 nm, depending on the strain of bacteria[25]. The relative difference in size between the two types of bacteria can be seen in Figure 1.3.

1.3 Viral particles

Whereas prokaryotes are simpler than eukaryotes, viruses are extremely simple particles compared to bacteria and are smaller. For perspective, the smallest bacteria have a diameter of approximately $0.4\ \mu m$ in diameter, while viruses are generally 0.02-$0.25\ \mu m$[30, 31]. Viruses exploit the capabilities of the host cell for processes such as replication, whereas bacteria are equipped to do this by themselves. Because of this, viruses do not need to contain relatively large structures, which are essential in bacteria[32]. Exploiting a host cell for complex processes removes the requirement for large, complex structures, making them strictly parasitic. This tradeoff for structural simplicity also makes them reliant on hosts to exist. Because viruses lack cellular structures and cannot perform metabolic processes by themselves, they are neither prokaryotic nor eukaryotic. Viruses are not classified as living organisms because they cannot perform their metabolic processes[33]. Viruses can pose a significant risk as they are highly efficient in entering host cells and rapidly replicating[34]. Due to the size and efficiency of virions, this makes them very difficult to stop or treat[32].

1.3.1 Viral capsids and envelopes

Unlike bacteria, viruses can be enveloped or non-enveloped; these viruses are referred to as naked viruses[3]. Instead, a virus possesses a protein coat called a capsid surrounding its genetic material; this generic material can be ribonucleic acid (RNA) or deoxyribonucleic acid (DNA)[3]. The capsid is a protein shell that encloses the viral genome and is responsible for protecting and transporting it so that it can be delivered to the appropriate region in the host cell's cytoplasm[35]. The capsid comprises protein subunits called capsomers and can be helical (Figure 1.4(a)) or icosahedral (Figure 1.4(b)) in shape, however they do not necesarily always adopt an ordered shape[32]. The shape allows for optimal packing of the type and amount of genetic material contained in the virus. A specific shape can also allow the capsid to mimic structures on the host cell surface, making host recognition and attachment easier[36].

Some viruses also possess an outer envelope that surrounds the capsid; these are called enveloped viruses (Figure 1.4(c)). The outer envelope is generally derived from host cell or organelle membranes and contains viral proteins and glycoproteins[37]. Deriving the PL composition of this envelope from the host cell helps the virus avoid immunodetection as the PL composition is recognised by the immune system[38]. Glycoproteins are essential for host cell infection by viral particles and, further, the particles' ability to avoid detection by the immune system.

1.3.1.1 Glycoproteins

A distinguishing feature of viral envelopes is the abundance of glycoproteins; these are surface proteins with covalently attached carbohydrates[39]. Although these structures exist in prokaryotic organisms, they are rare[40]. Glycoproteins are crucial for viral particles to function effectively, aiding them in viral entry, attachment, fusion and immune evasion. Glycans aid viral proteins in immune evasion through glycan shielding. Glycan shielding makes it harder for the host immune system to create antibodies that effectively bind to the viral protein by hiding it under glycans[41]. An example of why glycan shielding is necessary is shown in Figure 1.5, where the RBD of the S protein is in the up state and is visible. This area is not glycosylated as it needs to bind to the host cell receptor. To avoid this being targetted by the immune system, however, it remains buried in the ECD until needed to bind[42]. The broad range of viral glycoproteins reflects the diverse mechanisms viral particles use to infect host cells successfully. The most common amongst these are envelope proteins (E), membrane proteins (M) and S protein[43]. Glycoproteins like S are imperative in the virion-host cell binding mechanism of viruses[44]. Glycosylation is a post-translational modification that can significantly affect aspects of the proteins, such as folding, stability, receptor binding and immune recognition, as mentioned previously. Glycan chains are also sometimes necessary for creating an effective binding site for host receptors, allowing for the initiation of cell infection[45]. Despite the importance of glycoproteins to viral particles, they are also essential in developing vaccinations. Generally, vaccines are specifically developed for glycoproteins in the target virus. Because of this, we must understand them well to create effective vaccines[46].

1.3.2 SARS-CoV-2

SARS-CoV-2 is a coronavirus which emerged around late 2019, belonging to the coronaviridae family of enveloped viruses[47, 48]. This virus' rapid global spread and impact highlighted the risk of pathogenic outbreaks and our inability to deal with them effectively. Due to the virulence and infectivity of SARS-CoV-2, it has proven extremely difficult to apply effective, rapid-response preventative measures short of widescale vaccination. Because of this, we must have an excellent understanding of all aspects of the virus. If we hope to stop similar outbreaks

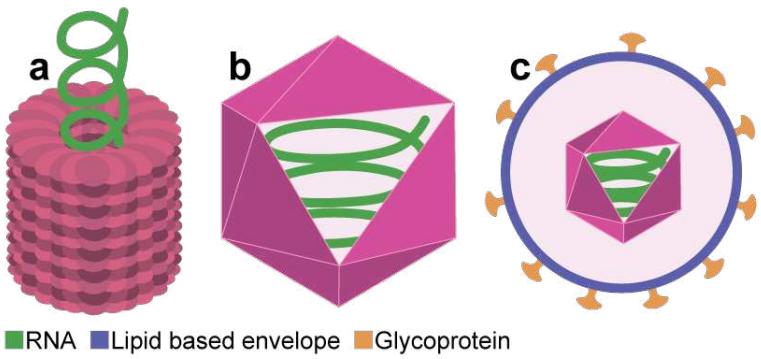


Figure 1.4: Types of virus, showing a (a) helical virus, (b) icosahedral virus and (c) an example enveloped virus.

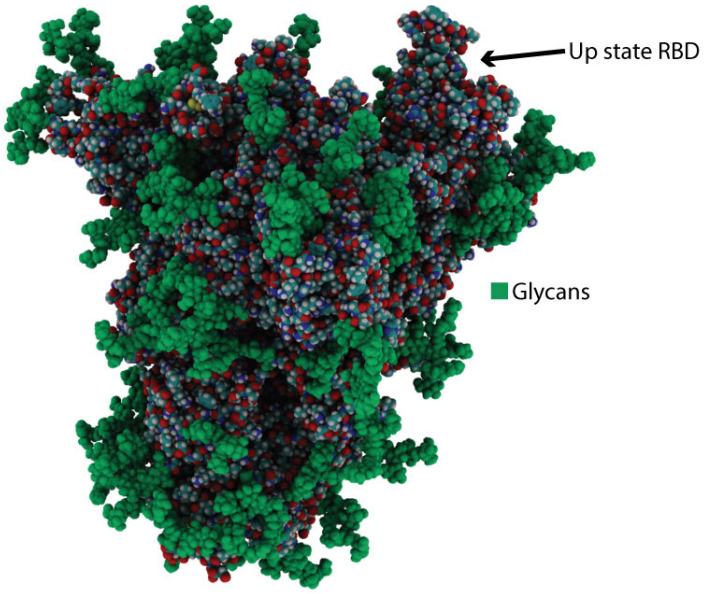


Figure 1.5: A visual representation of the wildtype Wuhan strain of the S protein ECD coloured by element with glycans coloured in green.

in the future, we must first understand why the SARS-CoV-2 outbreak was so damaging.

SARS-CoV-2 is a highly efficient virus with a diameter as small as 50-140nm. The particle is equipped to fuse with mammalian cells before inserting its RNA and using the host's organelles to reproduce. The particle contains several vital proteins (Figure 1.6), such as M, which is responsible for assembly and morphogenesis[49]. As well as this, there is also E, which is a small protein responsible for many aspects such as assembly, budding, envelope formation and pathogenesis[50]. The most well-known of these proteins is the S protein, responsible for attaching SARS-CoV-2 to a host cell before membrane fusion. The S protein is located in the viral envelope of SARS-CoV-2 and is necessary for the virus to latch onto human cells by binding to receptors such as ACE2 and others[51].

Once S has latched onto the host cell, the viral particle will then fuse with the host cell[52]. The capsid will release the RNA, which hijacks the host cell; this causes the host cell to produce viral proteins rapidly[53]. Some of these are responsible for the structure of new SARS-CoV-2 particles, and others are responsible for the replication of viral RNA. Viral RNA is transcribed and replicated to generate new viral particles and create proteins. New SARS-CoV-2 particles then bud from the host cell and are released into the host to spread the infection further[54]. Generally, the damage caused by this process will then signal the immune system and trigger an immune response[52].

A significant challenge faced in the SARS-CoV-2 pandemic was the emergence of new variants[54]. Random genetic mutation in SARS-CoV-2 occasionally results in variations which make the particles more effective, enabling them to out-compete previous, less effective variants[55]. These significant variants are termed variants of concern (VOCs) and may possess similar risks to previous variants with the added concern of vaccine resistance and increased transmissibility. Significant mutations often make changes to the S protein[56]. These modi-

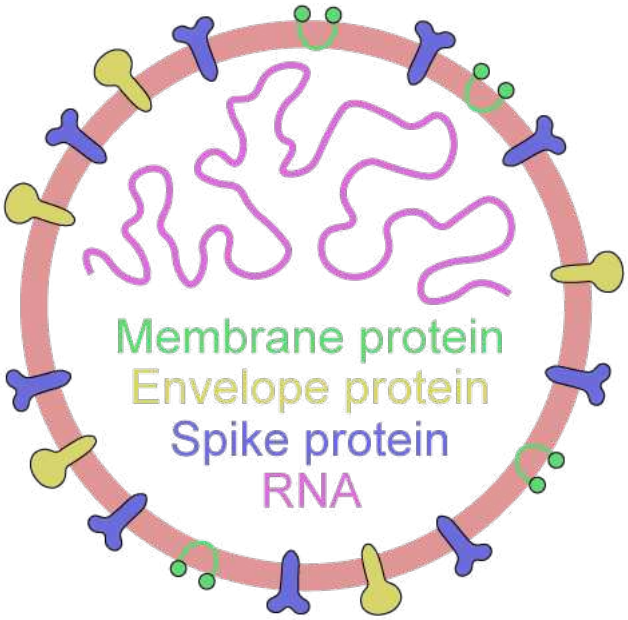


Figure 1.6: A simplistic representation of the SARS-CoV-2 viron.

fications improve the ability of the protein to bind effectively to host cell receptors and better evade the immune system[57].

1.4 Lipids

Lipids are self-assembling amphiphilic molecules that form a crucial component of biological membranes[58]. Lipids relevant to the work discussed in this thesis are shown in Figure 1.7. As these molecules make up the majority of a membrane's composition, their ratio, structure and properties significantly affect the overall properties of the membrane[59]. Broadly, lipids are subdivided into three classes: PLs, sphingolipids (SL) and CHL[10]. PLs are lipids composed of fatty acid chains attached to a glycerol backbone bonded to a phosphate group with an attached R group. This R group can significantly affect many properties, such as the membrane's charge, polarity or stability[60, 61]. For instance, in the *S. aureus* cell membrane (SaCM), the headgroups of Phosphatidyl-Glycerol (PG) and Lyso-Phosphatidyl-Glycerol (LPG) can form hydrogen bonds which make the membrane less susceptible to sanitisers[62]. The fatty acid tails of lipids can also significantly affect the bilayer, determining the membrane's rigidity or fluidity[63]. SLs are similar to PLs but contain a sphingosine backbone instead of a glycerol one, effectively increasing membrane stability[64]. These lipids are necessary for cell recognition, so they also appear in viral membranes, as features like this allow them to remain undetected by the immune system[35]. CHL is a bulky, cyclic molecule that effectively increases membrane rigidity at high temperatures while allowing it to remain relatively fluid at low temperatures; this is important for a cell's survivability[65].

In this work, the endoplasmic reticulum Golgi intermediate compartment (ERGIC) representing a viral membrane and epithelial membranes were studied. These are used to model systems containing the SARS-CoV-2 S and ACE2-B^0AT1 heterodimer of dimers, respectively[3, 35]. The

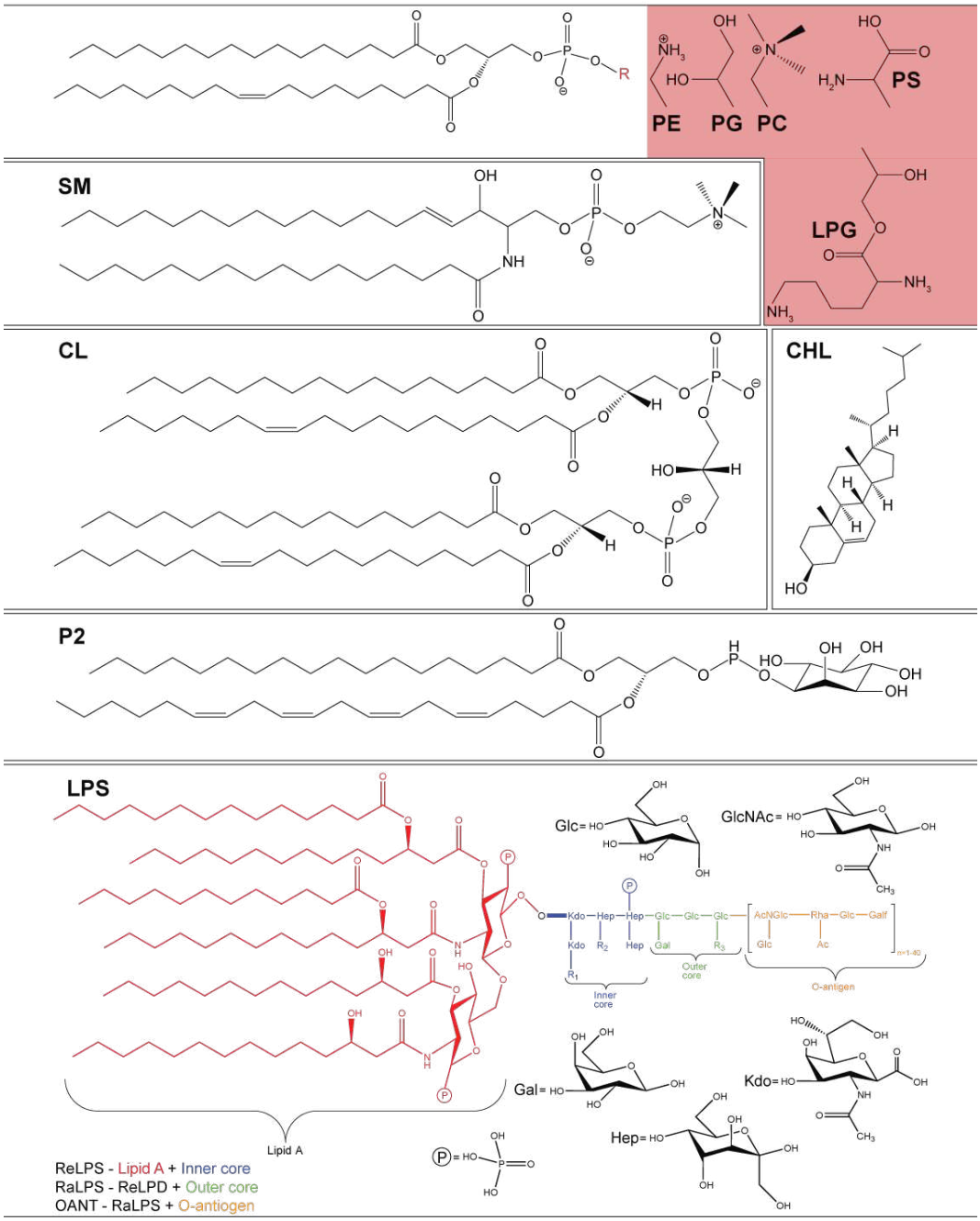


Figure 1.7: structure of the lipids used to model the *E. coli*, *S. aureus*, ERGIC and epithelial membranes. PE, PG, LPG, PC and PS are shown as their differences in the R group.

ERGIC membrane was used to model a viral envelope, as when viruses use host cells to reproduce, they derive the envelope composition from the host. The ERGIC is located between the endoplasmic reticulum and Golgi apparatus and is responsible for processing newly synthesised proteins[50]. The membranes of this organelle are composed of phosphatidyl-choline (PC), PE, phosphatidyl-inositol-bisphosphate (P2), phosphatidyl-serine (PS) and sphingomyelin (SM). As this region is also associated with processing newly formed lipids, it is difficult to determine a definitive lipid ratio. The epithelial membrane comprises the same components as the ERGIC membrane with the addition of CHL. However, the ratio depends on the cell's location

and function. As SARS-CoV-2 is known to hijack human cells and use the ERGIC region to multiply, this is also generally considered a good model for the viral membrane.

1.4.1 Lipopolysaccharide

LPS is a complex lipid with six hydrocarbon tails and a large polysaccharide chain extending from the headgroup (Figure 1.7)[66]. These molecules are only found in the outer membrane of Gram-negative bacteria and play a key role in structural stability, avoiding immunodetection and protecting the bacteria from harmful chemicals in the surrounding environment[67]. LPS comprises four main regions: lipid A, the inner core, the outer core and O-antigen. The polysaccharide region (inner core, outer core and O-antigen) extends into the extracellular environment from the cell[68]. It forms a relatively large sugar region, which is extremely difficult for potentially harmful molecules to cross due to the abundance of hydrophilic sugars. These sugars form strong interactions and create a densely packed region that is not readily translocated; this supplies the envelope with a permeability barrier that protects the cell.

The lipid A component of LPS anchors the molecule to the membrane through interactions with other lipid A moieties[69]. It has also been shown that lipid A is bound together laterally by interactions between LPS and calcium ions in the LPS leaflet[70]. Lipid A comprises 6-7 fatty acid chains connected by a diglucosamine backbone; however, these can vary in length and saturation between different strains and species. The structure of lipid A is generally conserved between species[71]. This region of LPS provides a hydrophobic barrier that will exclude hydrophilic chemicals that make it past the polysaccharide region.

The polysaccharide regions of LPS extend from lipid A, comprising the inner core, outer core and O-antigen. As mentioned, this is a large sugar region that extends into the extracellular environment, making it difficult for hazardous molecules to enter the cell. Contrary to the lipid A region, this provides a large hydrophilic barrier that will exclude unwanted hydrophobic chemicals from the cell. ReLPS comprises lipid A and the inner core, whereas RaLPS also contains the outer core[68].

The inner core is the portion of LPS located closest to lipid A, embedded in the outer membrane. The inner core of LPS is relatively well conserved between different strains of Gram-negative bacteria, consisting of a series of sugar residues; typically, these are Kdo (3-deoxy-D-manno-octulosonic acid), heptose (Hep), and glucose (Glc)[72]. This region is necessary for stability but also provides a connection point for the outer core and, by extension, O-antigen[73, 74].

The outer core is significantly less conserved between species and can vary widely between bacterial strains. This section can contain various sugars, such as mannose (Man), galactose (Gal), and rhamnose (Rhm), but is not limited to these. In conjunction with the strong lateral interactions between lipid A and calcium, this creates a highly immobile wall which chemicals will struggle to get past[70].

The outermost LPS region is O-antigen, a highly immobile wall of polysaccharides extending into the extracellular environment. O-antigen is generally not conserved between bacterial

strains as this allows for different strains to evade the immune system. As O-antigen generally differs between strains, it is harder for the immune system to generate antibodies that effectively target all possible O-antigen structures[75]. The O-antigen region is by far the largest part of LPS. However, it is not present in all Gram-negative bacteria and LPS that do not possess O-antigen can be referred to instead as lipooligosaccharide (LOS)[73].

1.4.2 Self-assembly

Self-assembly is a property of lipids crucial to their effectiveness. These molecules self-assemble principally for two reasons: to minimise the potential energy from interactions with their environment and, secondly, to maximise the system's entropy. The structures formed by this process focus on minimising contact between polar and hydrophobic regions. The simplest arrangement that allows for this is the formation of a micelle (Figure 1.8(a)). Here, self-assembly occurs by packing hydrophobic regions together, shielding them from the polar environment with the polar headgroups[76]. Similarly, lipids can also form a lipid bilayer (Figure 1.8(b)), which allows the lipids to shield the polar and hydrophobic environments from one another by placing them in the bilayer centre. To avoid leaving the ends of the bilayer exposed, the bilayer can wrap around to create a lipid vesicle (Figure 1.8(c)); this creates a stable and adaptable barrier for cells and viral particles. This arrangement creates a physical barrier and, through molecules such as channel proteins, transporters and receptors, allows the cell to control what comes in and out (nutrition and antimicrobials)[77]. Figure 1.8(d) represents a self-assembling structure similar to that of Figure 1.8(a); however, the environments are invested. This may occur if lipids were placed in a hydrophobic environment rather than a hydrophilic one.

As mentioned, self-assembly occurs to minimise the enthalpic contribution to the system whilst maximising entropy. The system enthalpy is decreased by self-assembly due to the absence of interactions between hydrophobic and polar regions. To avoid unfavourable interactions between the polar water and hydrophobic lipid tails, water would form elaborate hydrogen bonding networks around it, known as water cages. This allows water to minimise contact to the hydrophobic region and decrease enthalpy by forming hydrogen bonds between them. Although this is enthalpically beneficial, the order this induces is detrimental to the system's entropy. Self-assembly avoids interactions of hydrophobic and polar regions entirely, thus avoiding the formation of water cages and providing an entropic benefit. Further, self-assembly

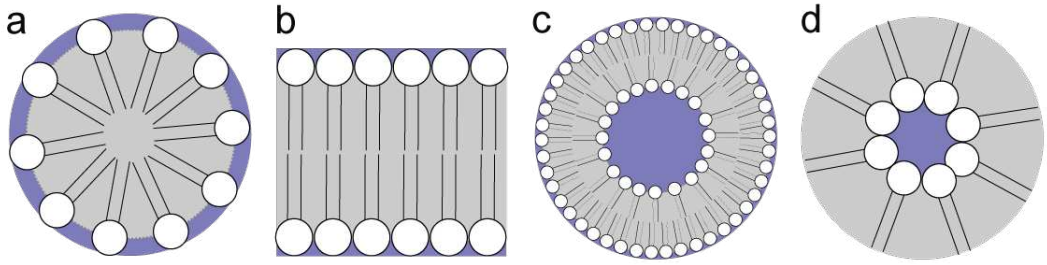


Figure 1.8: Examples of the four most common, relevant self-assembly structures of PLs, showing hydrophobic environments in grey and hydrophilic environments in blue. (a) Micelle. (b) Bilayer. (c) Vesicle. (d) Inverted micelle.

also increases hydrogen bonding and VDW interactions between lipids, furthering the decrease in enthalpy.

1.4.3 Bacterial lipid compositions

Due to differences in their PL composition, bacterial membranes may behave very differently[5, 6, 9, 78]. Further, in recent years, multiple bacterial species have seen increased antimicrobial resistance through changes in their membrane composition and LPS[79, 80].

1.4.3.1 *E. coli* membranes

E. coli is a Gram-negative species of bacteria with an envelope containing both the EcIM and EcOM. The lipid compositions of membranes can vary largely but generally fall close to certain ratios, and for the EcIM, this is approximated 75:20:5 Phosphatidyl-Ethanolamine (PE), PG and Cardiolipin (CL), respectively[81]. The PL composing the EcIM is decidedly saturated, meaning there are no kinks in the aliphatic tails, and the bilayer is tightly packed and rigid. This rigidity allows for *E. coli* to thrive in diverse environments where a more fluid membrane may be susceptible to factors such as osmotic pressure[82].

1.4.3.2 *S. aureus membranes*

As opposed to *E. coli*, *S. aureus* is Gram-positive, containing only the SaCM and the peptidoglycan cell wall in its envelope. Wild type *S. aureus* has a PL ratio of approximated 50:30:20 PG, LPG and CL, respectively[83, 84]. As opposed to *E. coli*, the PL composition of the SaCM contains many unsaturated aliphatic chains, resulting in a loose, fluid bilayer. This fluidity allows the bacterium to move proteins about the cell membrane while responding to environmental changes. This fluidity has been attributed to a susceptibility to antipathogens[85].

1.4.4 Chlorhexidine

CHX is a commonly utilised sanitising agent which is a member of a chemically related group of antimicrobials called bisbiguanides and forms a di-cation at physiological pH[86] (Figure 1.9). The structure of CHX consists of two chlorophenyl (CPL) functional groups, each bonded to separate biguanides (BGU), which are linked via the hexane (HEX) functional group. From end to end, this hydrophilic-hydrophobic-hydrophilic structure is thought to be critical to its bacteriostatic or bactericidal properties, depending on concentration[87, 88]. CHX is widely used as an antiseptic and disinfectant due to its broad efficacy, affordability and safety and is commonly found in mouthwashes and surgical scrubs[89, 90]. It is effective against both Gram-positive (e.g. *S. aureus*) and Gram-negative (e.g. *E. coli*) bacteria. CHX is thought to achieve bacterial cell death by disrupting bacterial membranes, leading to leakage of cellular contents, hence why it is often combined with alcohol, which is theorised to work in a similar way[91]. Under physiological pH, the two bisguanide groups carry a cationic charge of $1e$, each. It is thought that the cationic nature of the molecules plays a significant role in their interaction with the negatively charged bacterial membranes[92].

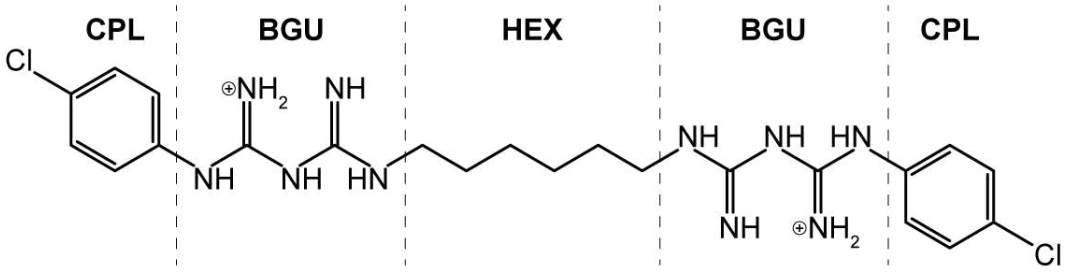


Figure 1.9: Separation of CHX into key functional groups, which are necessary for the interaction of CHX with a lipid membrane.

CHX is theorised to bind at the headgroup region of membranes via the CPL and BGU functional groups as a 'wedge', parting the headgroups and creating gaps that make the membrane 'leaky'[93]. This charge interaction between CHX and the bilayer was shown by Van Oosten *et al*[94]. This results in binding further from the membrane centre but more strongly to the head group region in negatively charged membranes. This interaction is particularly strong; studies have shown that, unlike other antiseptics, CHX can remain active after being washed away, remaining bound to the epidermis[95, 96]. CHX is active against *E. coli* (Gram-negative) and *S. aureus* (Gram-positive).

It was previously thought that the adverse effect of CHX on pathogens was due to its inhibition of ATPase, a group of enzymes necessary for the hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and the release of energy. This would deny a pathogen the energy it needs to perform essential processes that the cell relies on; however, this has recently been shown not to be the prevailing reason CHX is effective. The research conducted by Kuyyakanond *et al.* measured the pH over 4 minutes and showed that although upon the addition of CHX, there is an initial decrease in pH, it recovers shortly afterwards. In contrast, upon adding polymyxin, the pH decreases and does not recover, suggesting that this operates via a different mechanism[97].

1.5 Molecular dynamics simulation of biological systems: A brief review

In modern-day research of microbiological systems, MD is an invaluable tool that allows us to probe the intricacies of systems that are too small or have interactions that are too brief to be explored in a laboratory. This is essential as laboratory techniques are often technically cumbersome or lack the resolution required to study these systems. For example, NMR is beneficial for assessing solution or solid-state dynamics but generally lacks a simple way of visualising systems without intricate techniques. Visual depiction is essential when determining phenomena such as binding or partitioning as data otherwise leaves us to form conclusions which may be subject to confirmation bias. This is not to say that either MD or NMR are better; however, both have caveats. By taking a multidisciplinary approach, we can address the shortcomings of each technique and play to the strengths of multiple techniques.

Over the last few decades, MD has expanded from a niche, theoretical approach in its infancy to

a disruptive technique at the forefront of science[98]. As such, this field has seen the sampling techniques, force fields and structures it relies upon expand rapidly in detail, functionality and complexity. Initially, it was thought "good enough" to model biological membranes as one lipid type, generally PC. However, since then, our simulation systems have become significantly more robust; the lipid composition of the EcIM, for instance, is modelled as 90:5:5 (PE:PG:CL)[99], while the cell membrane of *S. aureus* is modelled as 50:30:20 (PG:LPG:CL)[83, 84]. The importance of these distinctions is furthered when considering membranes like the EcOM which possesses a leaflet composed predominantly of LPS[100], or the SaCM, which relies on PG-LPG interactions for headgroup stability[101]. This is but one example of how our understanding of how to best model systems we are interested in has improved over the years.

In its infancy, the number of molecules that there were models for was minimal, let alone the number of larger structures such as membranes that could be simulated[102]. Over the years, this lack of detail was addressed by developing comprehensive parameterisations[103], tools to predict these parameterisations[30], and more robust force fields better equipped to model molecules accurately[104]. In recent years, the model of the cell envelope has been taken to its extreme in computational simulations, which models the Gram-negative cell envelope, from membranes and cell walls[105], to LPS lateral diffusion[70] to the often overlooked crowding of the cell envelope[106]. This development of comprehensive models has not been limited to bacteria, as our ability to simulate substantial aspects of the viral envelope has also come along significantly[107]. In 2021, this was taken as far as using artificial intelligence (AI) to drive simulations of an entire viral vesicle with embedded S proteins to assess how it binds to ACE2 receptors[108]. This only foreshadows the advances that will surely be made in the following decade, when the simulation of entire viruses may become standard practice. Extensive research in these areas over recent years has provided essential insight into the complexity of bacteria and viruses, understanding which we can exploit to further our defences against them[109].

1.6 Solid-state NMR of biological systems: A brief review

Solid-state NMR is a well-established technique often used to assess the structure and dynamics of biological species such as proteins or lipids. The structural insight this technique can give us is beneficial when determining molecular-level interactions in vitro, as this level of detail can be extremely difficult or impossible to access with most laboratory techniques. By examining how systems return to equilibrium, we can extrapolate properties such as molecular conformation, localisation and effect on larger structures such as lipid vesicles. By applying many NMR techniques, we can gain valuable insight into molecular-level interactions such as protein and bilayer dynamics, ligand interactions, molecular structure and many other aspects.

A significant challenge in the development of solid-state NMR was the anisotropy of solid-state systems, as they are generally highly ordered. This is not an issue in 1D ^{31}P NMR as the resulting lineshape is a sum of all anisotropic environments present in the vesicle, which are essential when determining the stability of said vesicles. However, when trying to obtain 1D or 2D ^{1}H NMR, anisotropy will result in spectra that are extremely difficult or impossible to resolve as peaks appear for each proton in many different anisotropic environments. This

was eventually overcome by the development of the magic angle spinning (MAS) technique, which resolved this problem by essentially mimicking the tumbling a molecule would experience in solution[110]. By spinning the sample along the diagonal of a cube relative to the NMR machine's magnetic field, the x, y and z environments are averaged out. This development allowed for the analysis of solid-state systems using methods previously only accessible to liquid-state samples.

Although this technique is significantly older than MD, it is only in the last 40 years that it has become commonplace in microbiology[111]. This development began with the first spectra obtained for biomolecules in the 1980s, followed by MAS's development in the 1990s. It was in the 21st century that the first solid-state NMR spectra of membrane-bound proteins were resolved[112, 113]. This rapid development has led us to a point where the previous 20 years have been focused on improving the pulse sequences used in these experiments or developing new ones that allow us to probe microbiological systems further.

1.7 Aims

Bacterial and viral pathogens constantly threaten our health, as they can quickly adapt to adverse conditions and become resistant to treatment. This is why it is essential to understand the intricacies of these pathogens and develop effective ways to combat them. While our knowledge of these systems has grown considerably over the years, many aspects remain poorly understood and in need of further research.

My thesis aims to shed more light on the effect of sanitisers on pathogenic envelopes, which are essential structures that protect pathogens and enable them to cause disease. By employing a multidisciplinary approach, I hope to gain a more comprehensive understanding of how sanitisers interact with these structures and how we can better exploit their features to destroy or disable pathogens.

Specifically, I will investigate how sanitisers bind or partition into membranes and how pathogenic envelopes respond to these treatments. By applying techniques from multiple disciplines, I aim to ratify my results while developing a multifaceted understanding of a range of aspects that are currently targets of sanitisation or potentially could be in the future.

I aim to expand our understanding of how sanitisers work and how mutations could potentially render them obsolete. By identifying the mechanisms responsible for the effects we see when we apply sanitisers, we can better predict and prepare for future outbreaks and epidemics. Moreover, my research will explore the dynamics of SARS-CoV-2 binding, the virus responsible for the COVID-19 pandemic. Additionally, my research explores the dynamics of SARS-CoV-2 binding as this provided a relevant area where imperfections in a pathogen's binding mechanism may provide regions that sanitisation methods could exploit.

The following chapters can be summarised as follows:

- Chapter 2: Introduction to the theory, concepts and approaches used in MD simulations, NMR experiments and machine learning.
- Chapter 3: Methods used for MD simulation setup, NMR experiment setup and heat mapping module development.
- Chapter 4: Deciphering the mechanical effect of chaotropic sanitisers on PL membranes *via* all-atom molecular dynamics simulation.
- Chapter 5: Determining binding conformations and locality of chaotropic sanitisers to the inner membrane of *E. coli*.
- Chapter 6: Protein deformation resulting from the application of alcohol to the Wuhan variant of the S protein *via* all-atom molecular dynamics, with analysis assisted by decision tree regression machine learning.
- Chapter 7: Assessing the stability of S protein-ACE2 receptor heterodimer cross-linking *via* coarse grain simulations.
- Chapter 8: Concluding remarks and future perspective

Chapter 2

Theory

2.1 Introduction to MD

MD employs mathematical models to model small-scale representations of real systems using experimentally derived parameters. The significant benefit of this technique is that it allows us to simulate systems at atomistic levels of detail. Since these simulations are models, we can also explore purely theoretical systems by removing or adding features to a system and assessing the effect of this on the system. This allows us to explore the nuances of systems by only simulating specific properties or regions. These techniques have become essential for efficient work in drug design, pathology, and exploratory research in recent years.

2.1.1 Classical MD

MD is one of the most prominent areas of computational chemistry and relies solely on classical mechanics, principally upon Newton's second law of motion (Equation 2.1).

By equating Newton's second law of motion with the potential energy of a particle in motion, the acceleration of a particle can be calculated as a function of the particle's mass, potential energy and position. Equation 2.2 shows the potential energy of a particle in motion. Rearranging this equation to give the force (Equation 2.3) means that $\vec{F}$ can be substituted into

$$\vec{F} = m\vec{a} \tag{2.1}$$

$$dE_P = -\vec{F}d\vec{r} \tag{2.2}$$

$$\vec{F} = -\frac{dE_P}{d\vec{r}} \tag{2.3}$$

$$\vec{a} = \frac{1}{m}\frac{dE_P}{dr} \tag{2.4}$$

Equating Newton's second law (Equation 2.1) with force derived from the potential energy gradient of a particle (Equations 2.2) to find the force incident on a particle (Equation 2.3), where $\vec{F}$ is force, m is mass, $\vec{a}$ is acceleration, E_P is the potential energy, r is position and d denotes a differential.

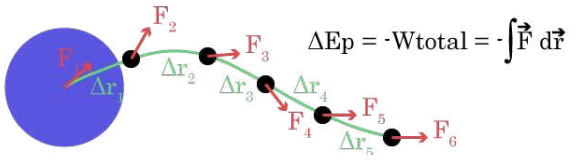


Figure 2.1: A visual depiction of the approximation made when calculating the change in potential energy of a particle from the forces acting upon it, where the particle is shown in blue, forces in red and changes in trajectory in green.

Newton's second law, yielding Equation 2.4; this enables us to calculate a particles acceleration. As this relied upon us knowing the potential energy gradient of a particle, we must first know the forces acting upon it, as shown in Equation 2.2. This relates the position of the atoms to their potential energy and allows us to determine how the particles should move in a simulation. This differential equation for acceleration works exceptionally well for MD as the changes in position and force are stepwise. A visual depiction is supplied in Figure 2.1 to aid the reader. To know the force acting upon a particle and by extension, the potential energy gradient, a force field is used (Section 2.1.3).

By integrating over all timesteps for any required period, we can come up with a reasonable estimate of what we expect the molecules in our system to do by knowing in which direction they are expected to accelerate at what speed and for how long. It should be noted that as it is upon the user to provide the force field, a simulation is only as good as the information provided. To elaborate on this, a visual representation of the stepwise process an MD simulation uses is provided in Figure 2.2.

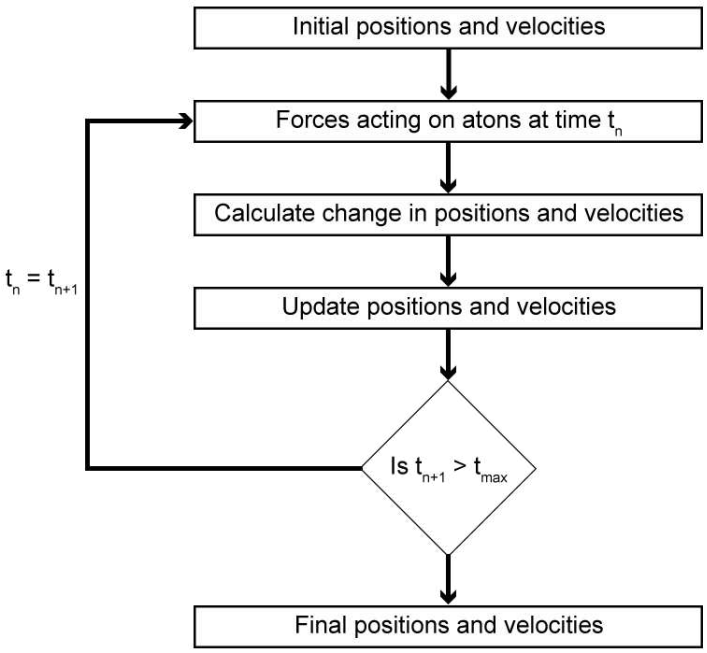


Figure 2.2: A flow chart representing the process of calculating the positions and velocities of particles in the next timestep in the simulation.

2.1.1.1 Coarse graining

Coarse graining (CG) focuses on computationally modelling systems by simplifying their representations to increase sampling with an accepted trade off in accuracy (Figure 2.3(a)). This is commonly implemented by representing pseudo-atoms instead of atoms to decrease the number of particles which need to be explicitly sampled. There are also varying resolutions to which a user can coarse grain their system by simply combining hydrogen atoms into hetroatom's, grouping amino acid atoms or as many as 4 hetroatoms into large pseudo-beads.

The Martini forcefield was initially developed for the simulation of lipids but has since been expanded and can now simulate more most molecules with sufficient parameterisation, such as proteins, ribonucleic acids or various other molecules. In work performed here, Martini 2 coarse graining is used which simulates 4 hetroatoms as a separate pseudo-atom. The aim of this mapping is to reproduce the thermodynamic properties of the system in as much detail as possible, while significantly decreasing the computational expense that the system will incur.

Whereas the average AA simulation is on the scale of a few hundred nanoseconds, Martini models can explore mechanisms that occur over multiple microseconds. This grouping of atoms into beads results in an overall smoothing of the potential energy landscape as the structure of molecules is far less specific. Because of this the potential energy landscape is sampled faster and events in simulation happen faster than they would in AA resolution (Figure 2.3(b)). This results in kinetic barriers being overcome faster as they are 'smoothed' by the decreased specificity. Because of this, the simulation will also progress faster with respect to the timesteps covered. This, as well as the more accessible longer timeframes, means that thermodynamically unlikely processes are far more accessible in CG simulation. It should, however, be noted that this smoothing of the potential energy surface can result in small states being inaccessible. Concerning the forcefield, this entails defining beads and their associated properties rather than atoms.

2.1.2 Integrators

MD simulations are created in the same way as a flip book. Stepwise snapshots of the simulated species are created by calculating the forces in a frame and using this to find the acceleration. This is used to calculate the positions of the particles after a small change in time, resulting in a series of coordinates that, when played at speed, represent what is expected to happen. We do this instead of calculating the motion analytically because of the vast number of particles in the system. As motion is so dependent on its surroundings, calculating the motion of these particles leads to a many-bodies problem. As such, the prediction of this needs to be approached with a finite difference method. A finite difference method assumes that positions, velocity and acceleration can all be approximated via a Taylor expansion. MD will generally do this via the Verlet of velocity Verlet method; however, work performed here uses the leap-frog integrator[114].

2.1.2.1 Timestep

When considering the timestep for an integrator in a simulation, we have to consider the fastest motions in the system to ensure that we are capturing this detail. This can have different meanings depending on whether we are explicitly representing each atom in the system or coarse-graining, as this will determine the fastest motions we aim to capture. For simulation systems, this can mean that beads would overshoot positions within their vibration range. This can lead to beads landing on one another, causing exponentially large potentials, causing the system to explode.

2.1.2.1.1 All-atom In AA simulation, the goal is primarily to capture even the fastest atomistic motions, such as bond vibrations, fast atomic motions and high-frequency dynamics. Due to Nyquist's theorem, the sampling rate (timestep) should be at least half the fastest motion we need to capture accurately to avoid aliasing. For example, the period of a C-H bond vibration is approximately 11 fs, so we should sample the system at least every 5.5 fs. Although this would technically be acceptable, using a significantly smaller sample rate as low as 10-20% of the fastest motion is generally recommended. As a result, the timestep in AA simulations is generally on the order of 2-3 fs. This is a large hit to computation time as it means positions need to be calculated extremely frequently, even if we choose not to save them this often.

2.1.2.1.2 Coarse-grain In CG systems, the fastest motion is significantly slower. This is because we do not consider atoms explicitly but instead consider beads, representing 3-4 heteroatoms. This results in significantly heavier beads, and due to the conservation of momentum, the same energy will result in slower vibration frequencies. As a result, the timestep can be significantly larger, which further saves computation time as Martini 2 is parameterised for timesteps of 20-40 fs. This represents another significant benefit of CG simulations for long timescales. When applied correctly, CG MD is an invaluable technique enabling us to sample long simulations quickly; however, this comes with associated caveats. Due to the simplification of structures and the effect this has on the potential energy landscape, interactions can be unrealistically short or missed entirely. The asterisk in Figure 2.3 marks a point where a potential energy well can be missed due to smoothing the potential energy landscape.

2.1.2.2 Leap-frog integrator

The leap-frog integrator is the default used by Gromacs[115]. It makes calculations of the displacement and velocity alternately, first performing the displacement calculation at t and then the velocity calculation at $t+\frac{1}{2}t$, essentially jumping over itself, hence why it is named the leap-frog integrator; this is done by using two equations. The integration will initially make a calculation using Equation 2.5, using the information we provide the system with (initial velocity, initial position, timestep) and arrive at the new position at time $t + \Delta t$. Following this we can then calculate $v(t + \frac{1}{2}\Delta t)$ again using our input data (timestep, mass, force, Equation 2.6), however it is immediately obvious that we do not know $v(t + \frac{1}{2}\Delta t)$ as we are

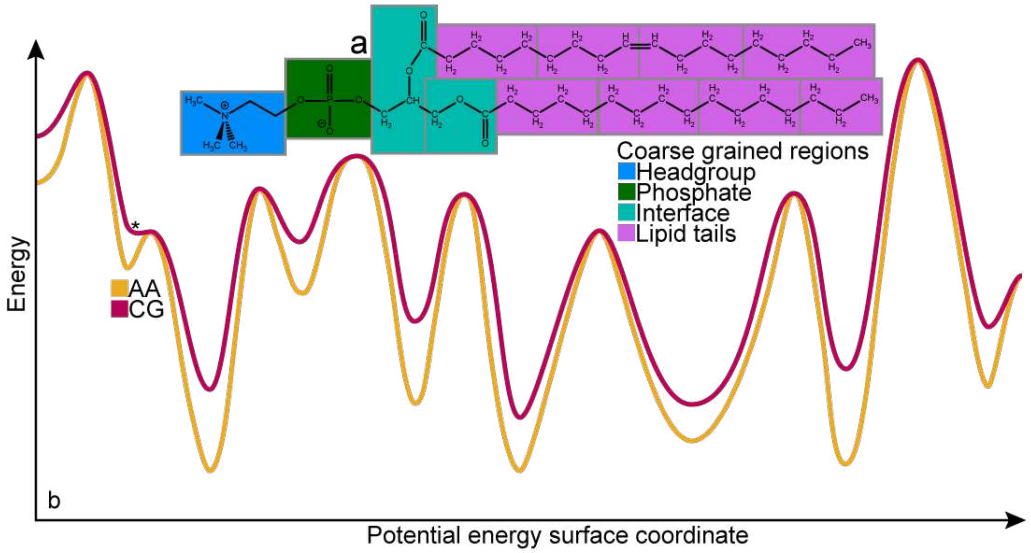


Figure 2.3: (a) A representation of the mapping from an AA structure of PE to a CG structure, with aliphatic beads in green, ester beads in red, phosphate beads in pink and ammonium beads in blue. (b) A 2D representation of the potential energy surface of a system, where an example of an AA system is shown in yellow and a CG system in red. The position marked by an asterisk indicates a potential well which missed due to smoothing of the potential energy landscape in CG.

currently at $t = \frac{1}{2}\Delta t$ we don't know the velocity half a timestep before the simulation started. We can approximate $v(t + \frac{1}{2}\Delta t)$ by using a single step of the Euler approximation (Equation 2.7). This provides us with a velocity for the half timestep and thus will allow us to calculate the first leap in velocity, passing the position calculation and giving us the velocity at $\frac{3}{2}\Delta t$.

2.1.3 Force fields

As stated earlier, to calculate the acceleration of a particle in an MD simulation, the potential energy gradient of said particle must be known. This is calculated by equating the forces acting upon said particle; an MD simulation calculates these forces by using the force field. A force field provides the simulation with experimentally derived data so particles can incur energy penalties from interactions. Force field interactions are broadly separated into bonded and non-bonded, split further into four bonded and two non-bonded terms. The data in the force field and how it is written determines how accurate, efficient and system-specific the simulation will be. The main drawback with this method is that we cannot simulate dissociation as we do not consider electrons; this is due to the extreme level of computational power nanometre-sized systems would require for this. This is not to say that one force field is necessarily the best, as some are parameterised better for different molecular species such as proteins, lipids, or any other molecular species in theory. The research performed here utilises the Gromacs simulations software, and as such, the details of how computational methods work are consistent with this package [116, 117, 118].

In work performed here, the CHARMM36 forcefield was used. CHARMM36 is a forcefield which is commonly used in molecular dynamics, originally parameterised to generally simulate proteins

$$\vec{v}\left(t+\frac{1}{2}\Delta t\right)=\vec{v}\left(t-\frac{1}{2}\Delta t\right)+\frac{\Delta t}{m}\vec{F}(t) \tag{2.5}$$

$$r(t+\Delta t)=r(t)+\Delta t\vec{v}\left(r+\frac{1}{2}\Delta t\right) \tag{2.6}$$

$$\vec{v}\left(t+\frac{1}{2}\Delta t\right)=\vec{v}(t)+\frac{\Delta t}{2m}\vec{F}(t) \tag{2.7}$$

The velocity equation used by the leap-frog integrator (Equation 2.5), the position equation used by the leap-frog integrator (Equation 2.6) and the Euler approximation (Equation 2.7). Where $\vec{v}$ is the velocity of a particle, t is the current time, Δt is the timestep, m mass of the particle and $\vec{F}$ is the force acting upon the particle due to interactions with other particles in the system.

and nucleic acid, however it is now parameterized to simulate a wide range of biomolecular molecules. This forcefield was parameterised on experimental data for molecular properties such as bond vibrations, crystal structures and thermodynamics. As new scientific techniques are developed, the CHARMM forcefield has increased significantly in accuracy and adaptability. It is now easy to parameterise new molecules using the CHARMM General Force Field (CGenFF).

CGenFF is a forcefield developed to cover a wide range of chemical functional groups present in biomolecules and drug-like molecules. This forcefield is extremely useful for generating molecular structures, enabling users to rapidly generate and simulate their own structures through generalised, additive potential energy functions, which can describe the potential energy of a structure through itsstructural coordinates.

2.1.4 Bonded terms

2.1.4.1 Bond length

Force fields deal with bonds by treating them as if they were springs with an associated spring constant (k), oscillating about some optimal distance (r_0). The best way of doing this is using a Morse potential. However, to decrease computational cost, a form of Hooke's law is used (Equation 2.8). Although this is a slight difference in the complexity of the equation when integrating over the entire system, these subtle differences add up and lead to substantial increases in computational time.

$$\text{Hooke's law: } E_H=\frac{1}{2}k_H(r-r_0)^2 \tag{2.8}$$

$$\text{Bond length potential energy: } E_{bond}=k_b(r-r_0)^2 \tag{2.9}$$

Hooke's law (Equation 2.8) and the equation determining the potential energy of a bond (Equation 2.9), where E_H and E_{bond} are the potential energy, k_H and k_b are the spring constants and r is the spring length and r_0 is the optimal spring length.

Although Hooke's law is technically less accurate and does not model the exponential increase in potential energy with increasing distance, this method's precision can be improved by including higher-order terms in a way similar to a Taylor expansion. Most bonded terms with an

optimal value are modelled from a harmonic oscillator (Equation 2.9). Although this approximates how the bond length energy should behave around the optimal bond length, deviations from this result in exponentially large energies. This can cause significant issues in keeping the system energy constant. To avoid this problem, systems are carefully equilibrated and assuming simulations are well parameterised, bond lengths will not deviate far from the optimal value. This means we do not generate significant errors due to large bond lengths, as if the system is properly equilibrated, these won't occur. The nuances of this difference are shown in Figure 2.4.

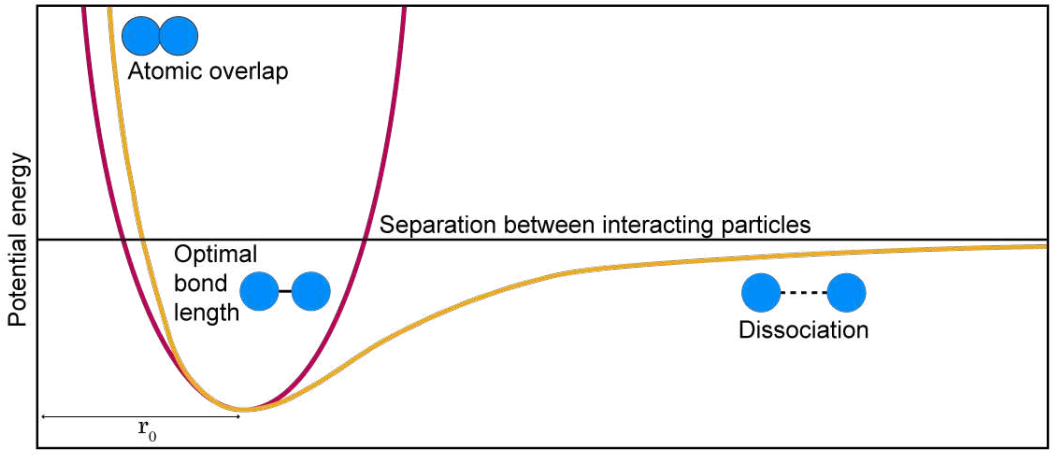


Figure 2.4: The (yellow) Morse and (red) Hooke potentials over distance. Showing optimal bond length (r_0).

2.1.4.2 Bond angle

Bond angles behave the same as bond lengths, oscillating about some optimal value. Because of this, the potential energy equation for this is the same as that of bond lengths. This equation's similarity to bond lengths can be seen in Equation 2.10.

$$\text{Bond angle potential energy: } E_{angle} = k_\theta(\theta - \theta_0)^2 \tag{2.10}$$

The equation used to determine the energy of a bond angle, where E_{angle} is the potential energy of an angle in the simulation, k_θ is the spring constant of the angle, θ is the bond angle and θ_0 is the optimal bond angle.

2.1.4.3 Torsion angle

The torsion angle (also referred to as the dihedral angle) term defines energies of rotations along a bond in terms of peaks and troughs. The troughs correspond to optimal values, and the term is continuous as it can theoretically rotate a full 360°. The term for this must be relatively flexible as the potential peaks and troughs are directly related to the functional groups surrounding the carbons on either side of the bond. Due to the cyclic nature of torsion angles, the equation we use to deduce the potential energy term must consider a modified cosine function to describe the peaks and troughs over 360° accurately and allow this to be

changed easily to describe multiple torsions. It should be noted that although a torsion angle can theoretically rotate 360°, this is generally not common, and conformations remain bound by a local minimum. Equation 2.11 shows the equation used for this term.

$$\text{Torsion potential energy: } E_{torsion} = A[1 + cos(n\phi - \phi_0)] \tag{2.11}$$

The equation used to determine the energy of a torsion angle, where $E_{torsion}$ is the potential energy of a torsion angle in the simulation, A describes the barrier heights, n describes the number of barriers, ϕ is the torsion angle and ϕ_0 is phase shift.

The equation for the potential energy is described in this way so that variables describing the bond can be used to detail many different torsions. Here, the bond constant (A) determines the steepness/height of the potential barriers involved; a greater constant results in more well-defined states. For example, if the R groups were single protons, then the barriers could be expected to be significantly smaller than if they were bulky hydrocarbons. As the torsion angle can rotate a full $360°$, n is also required to describe the number of potential energy barriers created by groups of atoms at either end of the bond interacting as the molecules rotate about the bond. A simple way of considering this is as a Newman projection where the potential barriers are represented by staggered groups overlapping. Figure 2.5 shows a visual representation of this.

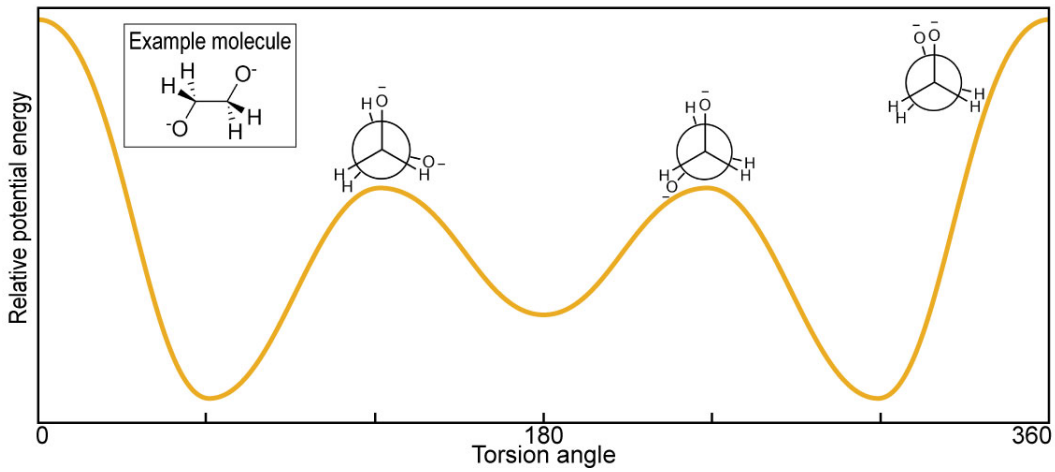


Figure 2.5: The potential of the dihedral term over dihedral angle, relative to example Newman projections of clashes which are higher and lower in energy relative to one another.

2.1.4.4 Improper torsion

The physical states defined by this term initially appear somewhat different from those previously described. However, this is another harmonic oscillator term. The improper torsion term is used to maintain specific conformations of ring structures or other crowded molecule areas. By monitoring the angle between two separate planes of three atoms, important stereochemical information is maintained. For example, this will maintain benzene rings or sugar conformations. It is also essential to ensure that functional groups do not flip into another

orientation; when working with molecules such as proteins, amino acids may flip between their L and D stereochemistry without this term. The potential energy term is calculated using another equation derived from Hooke's law to maintain this. The similarity of this to the bonded and angular terms can be seen in Equation 2.12.

$$\text{Improper torsion potential energy: } E_{improper} = A(\phi - \phi_0)^2 \tag{2.12}$$

The equation used to determine the energy of an improper torsion angle, where $E_{improper}$ is the potential energy of an improper torsion angle in the simulation, A describes the barrier heights, ϕ is the improper torsion angle and ϕ_0 is the ideal improper torsion angle.

2.1.5 Non-bonded terms

2.1.5.1 Complications in calculation

Barriers to consider when dealing with Non-bonded interactions are generally related to their long-range nature. As they are unbounded, unlike bonded terms bound by a covalent bond, they should technically be calculated for all particles at all distances in the system. This is not how we treat them in MD, as this would result in unrealistic computation times, self-interaction and infinite interaction problems due to the PBC. Instead, cutoff models with algorithms designed to estimate the penalties beyond them are applied.

2.1.5.2 Van der Waals

This term models short-range Van der Waals (VDW) interactions, ideally with a Lennard-Jones potential, an equation also used to model the same interaction in physical chemistry. Equation 2.13 shows that at very small distances, the $(r_{ij}^0)^{-12}$ term dominates, describing the increase in potential energy caused by the repulsive force from the overlap of electron orbitals. However, while at relatively large distances, the negative $(r_{ij}^0)^{-6}$ term will dominate, describing the attractive forces between two atoms.

$$\text{VDW potential energy: } E_{VDW} = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] \tag{2.13}$$

The equation used to determine the energy of VDW interactions, where E_{VDW} is the potential energy, ϵ describes the well depth, σ is the contact distance and r_{ij} the separation distance between particles I and j.

A visual representation of this (Figure 2.6) shows the position of the optimal bond length and the penalty for deviating from this.

When dealing with VDW interactions, a dispersion correction must be applied, as the Lennard-Jones potential does not sufficiently account for long-range interactions. While a cutoff is excellent for computation efficiency, it neglects some long-range interactions. Dispersion arises from temporary fluctuations in the electron density surrounding an atom and can induce similar dipoles in nearby atoms, which causes further dipoles. Although the Lennard-Jones model

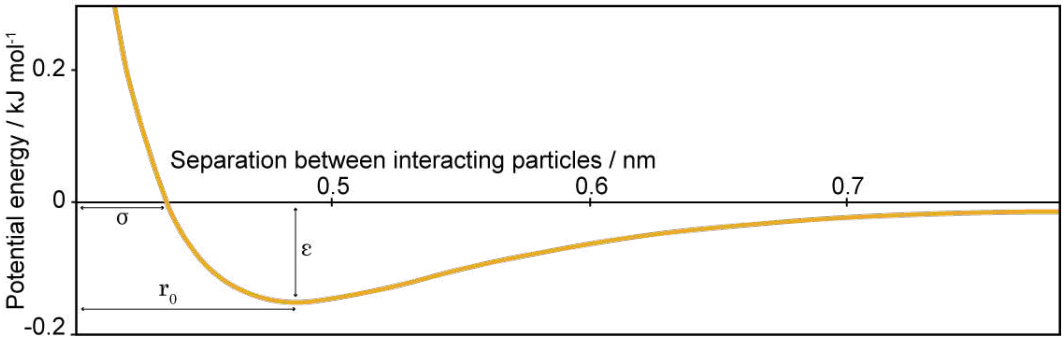


Figure 2.6: The potential of the (yellow) Lennard-Jones potential over distance. Showing contact distance (σ), optimal separation (r_0) and bond strength (ϵ).

treats short-range interactions well, this is not the case for long-range ones. The dispersion correction corrects the potential energy for dispersion forces beyond the cutoff, better approximating real-world interactions.

2.1.5.3 Permanent electrostatics interactions

The energy penalty due to electrostatics is based on Coulomb's law, and the difficulty in representing this accurately lies in the parametrisation of formal charges in species such as halides, alkali metals and alkaline earth metals.

$$\text{Electrostatics potential energy: } E_{el} = \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \tag{2.14}$$

The energy equation for permanent electrostatics where E_{el} is the potential energy, q is the charge on a particle, ϵ_0 is the permittivity of a vacuum and r_{ij} is the separation distance between particles i and j.

In Equation 2.14, the interaction energy is shown to be the charges of the atoms (q_i and q_j) divided by the multiplication of the dielectric constant and their separation. The dielectric constant represents the polarizability of the species in question. In Figure 2.7, the potential energy and separation can be seen to have a second-order exponentially decaying relationship.

The significant barrier we must consider when dealing with electrostatics is their slow convergence. Convergence refers to the attainment of consistent electrostatic forces and energies with respect to increasing simulation length. This is important as until the point of convergence, the electrostatics interactions in the system will not have reached equilibrium as they have not had sufficient time to influence the arrangement and behaviour of charged particles in the system. For this reason, electrostatic convergence is generally achieved in the equilibration phase of the simulation, as this means the production run is not subject to inaccuracies such as unrealistic distributions, artefacts, incorrect thermodynamic properties or deviation from experimental observations. As implied, the effect of this extends to the long-range dependencies of electrostatics interactions as they are not bound.

With employing such a rigid approach, we need to consider how this may create artefacts in our experimental data, which could have extensive implications for the results. The first thing

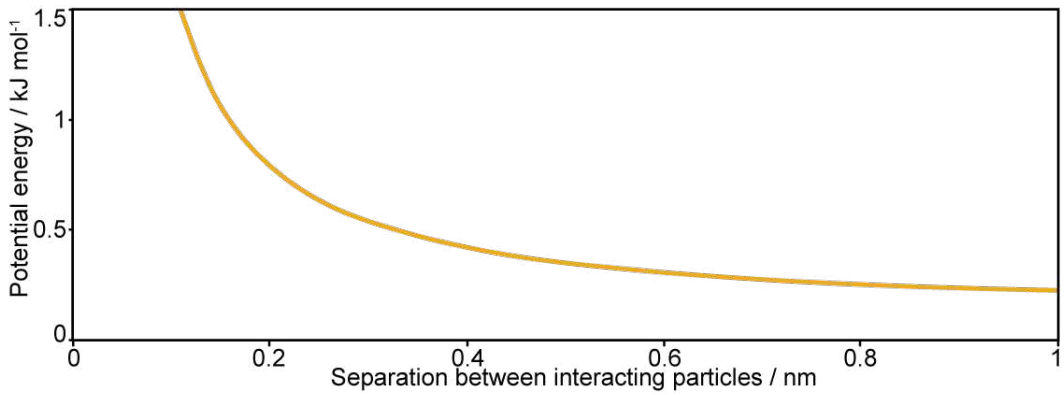


Figure 2.7: The potential energy of the coulomb potential equation over distance. Distances and potential energies are shown explicitly for comparison with the Lennard-Jones potential.

to consider is what the distance should be. Generally, a cutoff distance of 1.2 nm is applied as it can be seen from Figure 2.7 that toward 1 nm, the long-range approaches a constant value. However, It should be noted that it does not become negligible, and defining it as 0 J immediately beyond 1.2 nm will create large inaccuracies in our results. A particle-mesh Ewald (PME) approach is applied to AA systems, and a reaction field is usually applied to CG to remedy this. These methods split the calculation into two calculations, which converge more rapidly than the default electrostatics equation.

2.1.5.3.1 All-atom Unlike VDW interactions, we cannot simply cut off the effect of electrostatics over long distances. To address this, a PME is applied after a cutoff distance to account for these long-range interactions while still accounting for their effect. This is why electrostatics are handled very carefully in calculations. A compromise needs to be made between accuracy and computational expense, as the calculations obviously need to be sufficiently accurate. The PME will essentially group and average charges beyond the cutoff distance by assigning them to cells in a mesh. Charges are smeared over the cell by an interpolation function, so they are no longer formal but rather a charge density. This is then converted into reciprocal space by a fast Fourier transform (FFT) algorithm; electrostatics are calculated and converted back into real space. Using these calculations, the force and energy calculations are then performed. This significantly reduces the computational cost of this term while maintaining accuracy.

2.1.5.3.2 Coarse-grained PME works very well for a fully atomistic system where charges are well-localised. However, in CG systems, charges are less well-defined. The solution is to apply a reaction field rather than a PME approach. This has a minimal effect on computational expense and has been shown to create reliable results. The reaction field broadly separates the system into two environments for the sake of reducing the computational load for calculating long-range interactions. The reaction field introduces a dielectric boundary, beyond which interactions are approximated as a uniform, smooth, decaying dielectric continuum. The inner region contains real particles for which the forces and energetics are calculated.

Generally, PME is not used in CG simulations. PME is computationally expensive, and for CG scale systems, this can become costly due to the size. CG simulations are, by definition, less detailed; the computational overhead introduced by PME would result in slower computation and decrease the appeal of CG. It should be noted, however, that PME is occasionally applied in CG simulations. The application of PME, while increasing computational load, can provide more detailed electrostatics of a more reliable approach in inhomogeneous systems.

2.1.6 Energy minimisation

Before performing an MD simulation, it is first important to ensure that the molecules in our simulation have realistic bond lengths and angles that will not create exponentially large energy penalties. Ideally, performing a minimisation will result in the system being at an energetic minimum. This is achieved by iteratively making very small changes to the position of beads in the simulation system, which results in a net decrease in the system energy. Generally, this is done with a steepest descent method, which will assess the energetic gradient at a given point and move the system toward an energetic minimum by following the negative gradient.

2.1.7 Periodic boundaries

Computational simulations are inherently limited by the number of atoms we can calculate forces for, which raises the question of how we then consider systems much larger than the relatively small number of atoms we can simulate. In this way, we can apply boundaries to the small simulation boxes, which will consider the system as if it were joined to itself and then repeated infinitely in all dimensions. In these ensembles, atoms on the boundary of the simulation box will experience both bonded and non-bonded interactions with those on the opposite face. For instance, the atoms on the left face will interact with those on the right and large molecules will appear half off one side and re-enter the other. It should be noted here that although species technically leave the box constantly, they re-enter immediately, and as such, the number of atoms in the box remains constant. Figure 2.8 shows an extremely simplistic depiction of how periodic boundaries act, showing that as one molecule moves out of the box, it re-enters on the opposing face.

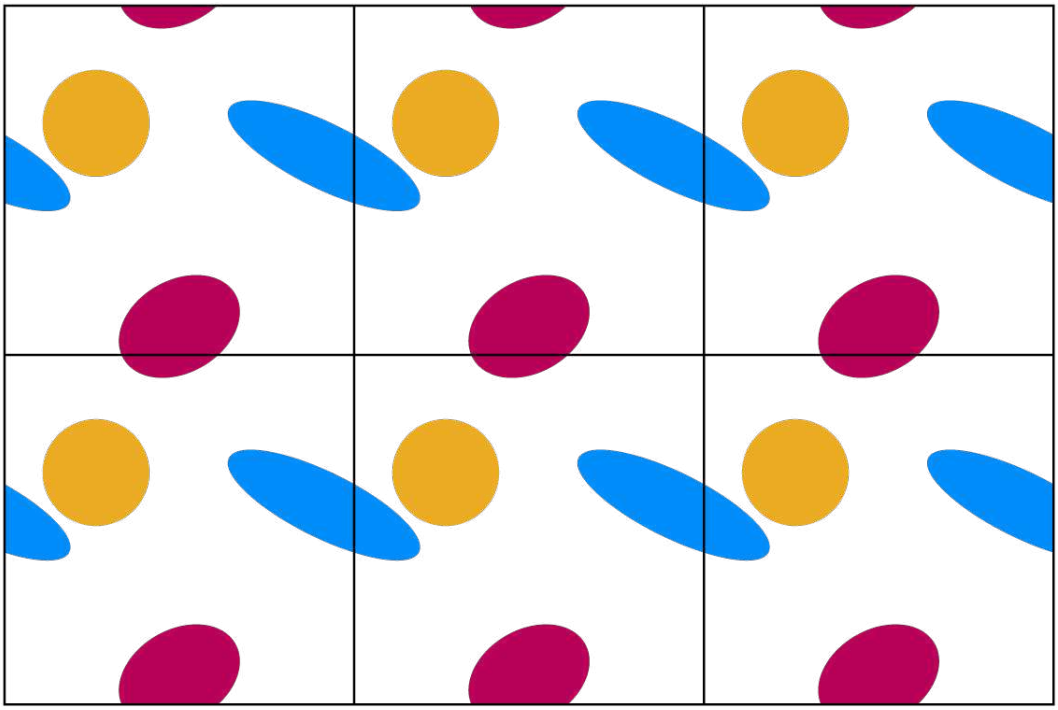


Figure 2.8: A visual representation of a periodic boundary-based system where coloured objects represent molecules that are passed off of one side and return on the opposing side.

2.1.8 Thermostat and barostat

To create accurate models of real-life systems, we must have well-maintained temperature and pressure. However, what this means in a closed, periodic system is not immediately obvious.

2.1.8.1 Thermostat

Two of the most common thermostats in MD are v-rescale and Nosé-Hoover. The reason we have more than one which are readily available is due to the limitations and successes of these thermostats. The goal of a good thermostat is not to fix the temperature of the entire system as it is, as this would mean fixing the kinetic energies of particles in the system, which is inherently unrealistic and would render simulations pointless. Instead, the thermostat aims to maintain the overall average temperature of the box by sampling different regions, which should average out to the instantaneous temperature, as maintained by the thermostat.

Firstly, the v-rescale thermostat; this thermostat scales the velocities of species in the system so that they average out to the desired temperature. This is necessary as during the equilibration phase of a MD system; the temperature must be adjusted so it is at the value specified in the input files. By scaling all velocities in the system proportionally, the average temperature is raised directly, providing a simple means of reaching the temperature. Although very convenient, this technique does not allow for fluctuations during the scaling process, and as such, areas of higher temperature cannot be avoided, so this is generally only in the setup phase.

The second commonly used thermostat is the Nosé-Hoover method, which utilises Lagrangian mechanics to maintain a system's temperature. This is a particularly accurate means of

performing these calculations as Lagrangian mechanics are a more formal means of performing Newtonian mechanics but more concerned with conservative forces. This is important as it means the system will not lose energy over time to non-conservative processes. Although this method is rigorous and strictly deterministic, it is largely only used in production runs as the v-rescale method provides a far more trivial means of altering the temperature. Whilst still easily implemented, the Nosé-Hoover technique is limited mainly by the fact it is not strictly 'ergodic', meaning that given a sufficiently large amount of time, in large systems, species may not mix well as there is no assurance this will sample all space in the box over time.

2.1.8.2 Barostat

Similar to how multiple thermostats can be applied to MD in different ways, different barostats are particularly good at performing different functions. Namely, two of the most used barostats are the Berendsen and Parrinello-Rahman methods, both of which have inherent strengths and weaknesses.

The Berendsen technique is particularly good in terms of maintaining a desired temperature over the course of setup, as feedback from the calculations made can also maintain the temperature of the system. Pressure and temperature are closely related, so that sudden increases in pressure can result in increased kinetic energy of the system; this would increase the average temperature. If this is detected in the Berendsen method, velocities will then be rescaled, similarly to the v-rescale method, so that the system remains at a desired temperature. Although this method is well-founded and very effective, it is limited in the same way as v-rescale temperature coupling, as the rescaling over the system can create pockets of higher and lower kinetic energy. This is also generally restricted to setup.

Parrinello-Rahman barostats expand the constant-pressure method devised by Anderson *et al*[119]. This method focused on simulating systems with constant pressure that still allowed for fluctuations in the kinetic energy of species in the box, which was achieved by implementing stochastic collisions. This was expanded upon by allowing the box dimensions to change while still maintaining pressure[120, 121, 122]. Generally, the Berendsen pressure coupling is applied during setup as a weak means of reaching the desired pressure before fixing it with Parrinello-Rahman.

2.2 Introduction to nuclear magnetic resonance

NMR reports on the interactions of nuclei in a sample by applying a magnetic field, perturbing the sample and monitoring how it returns to thermal equilibrium. The data reported by this technique can be used to infer properties of the system or molecules, such as dynamics, structure, exchange rates and much more. The benefit of this technique is that it supplies us with atomic-level details of large systems in vitro. This is done by monitoring how the spins of the nuclei in a sample exposed to a strong magnetic field return to equilibrium when perturbed by applying a radio frequency (RF). The NMR machine will report this information as a complex signal composed of many monochromatic signals. These monochromatic signals correspond to

$$\omega_0 = \gamma B_0 \tag{2.15}$$

The Larmor equation, relating the Larmor frequency (ω_0) to the gyromagnetic ratio (γ) and the magnetic field strength (B_0).

the relaxation of nuclei in different environments in a sample. Once perturbed by a radiation pulse, nuclei spins decay back to equilibrium. In an NMR experiment, the precession of spins back to equilibrium is measured and used to infer information about the nuclei and their local environments. In modern-day molecular biology, this technique is essential to understanding the transient, nuanced interactions which large, intricate systems depend upon.

2.2.1 Nuclear spin

Spin is a fundamental, quantum mechanical property of a nucleus. Nuclear spin arises from the intrinsic angular momentum of protons and neutrons within the nucleus, which gives rise to a magnetic dipole moment of the nucleus. This quantised property can only take specific values, being numeric half intervals. If the magnetic moment is non-zero, then the nucleus is NMR active. If the magnetic moment is 0, the nucleus will not interact with an applied magnetic field.

2.2.1.1 The Larmor frequency

When spin-active nuclei are exposed to a magnetic field, they begin to precess about it. The Larmor frequency describes the rate at which a magnetic moment (spin-active nucleus, in this case) will precess a magnetic field. This value can be related to the magnetic field strength (B_0) and the gyromagnetic ratio (γ) via the Larmor equation, as shown in Equation 2.15. The gyromagnetic ratio describes the ratio between a nucleus's magnetic moment and angular momentum. Because nuclei with a magnetic moment of 0 do not interact with the magnetic field, they do not possess a Larmor frequency. The Larmor frequency is essential in NMR as this is used to excite nuclei in the experimental sample selectively. When an RF pulse is applied at the Larmor frequency, it tilts the nuclear spin of the Larmor frequency away from alignment with B_0. This places the spin in an excited state, which will precess back to thermal equilibrium. As the Larmor frequency is a product of the gyromagnetic ratio and the magnetic field experienced by the nucleus, the frequency of precession during relaxation can also be changed slightly by the local electron density of the nucleus. As electrons possess their own magnetic moment, a larger electron density will generate its own local magnetic field, shielding the nucleus and changing the magnetic field experienced by the nucleus.

2.2.2 Free induction decay

As mentioned, NMR reports on systems by monitoring the signal produced when the nucleus spins relax from an excited state back to thermal equilibrium. The signal recorded while nuclear spins decay back to an equilibrium state is called a free induction decay (FID). The equilibrium state refers to the orientation of nuclei spins in a system when at rest in an applied magnetic

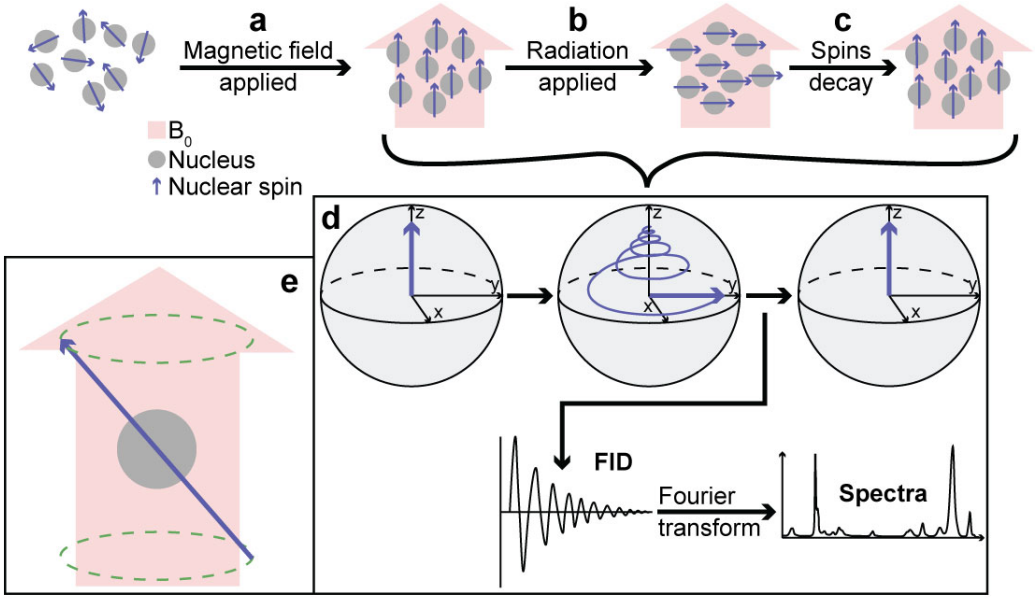


Figure 2.9: A simple, generalised depiction of a 1D NMR experiment. Showing (a) the generation of the equilibrium state, (b) the initial excitation of spins in the sample, (c) the decay of spins and (d) the origin of the FID signal, shown as a vector representation. (e) A representation of spin precession in an applied magnetic field.

field (Figure 2.9(a)). The applied magnetic field is referred to as B_0 and is a measurement of the applied field strength in Tesla. When the magnetic field is applied, nuclei spins will align either parallel (spin-up) or anti-parallel (spin-down) to the magnetic field. The excited state is created when an RF pulse perturbs the nuclear spins from their alignment to B_0 (Figure 2.9(b)). The excitation pulse applied is at the Larmor frequency of the nucleus of interest. Following this, the spins precess about the magnetic field until they return to their thermal equilibrium state, either spin-up or spin-down (Figure 2.9(c-d)). While precessing, the decay of nuclear spins creates a time-dependent magnetic field as the spin returns to the equilibrium state. In NMR, this changing magnetic field produces a voltage in the receiver coil. This changing voltage is monitored and generates an FID signal. As this changing magnetic field depends on the frequency of precession of the nuclei, the output signal will be a complex signal composed of many different monochromatic signals created by precession from many different nuclei. Because of this, the signal can then be Fourier transformed to split it into the intensities of monochromatic signals which contributed to it and, by extension, the atomic environments that contributed.

2.2.3 Chemical shielding

The FID is Fourier transformed into a spectrum that reports the intensity of all monochromatic frequencies that compose the FID; these were the precession frequencies for different nuclei in the sample. In an NMR experiment focusing on one element, if in equivalent environments, all of the nuclei in the sample would process at the same frequency, creating a single frequency in the FID and one large peak in the resulting spectra. This is not the case, as the local electron density of each nucleus will be slightly different depending on their local environment.

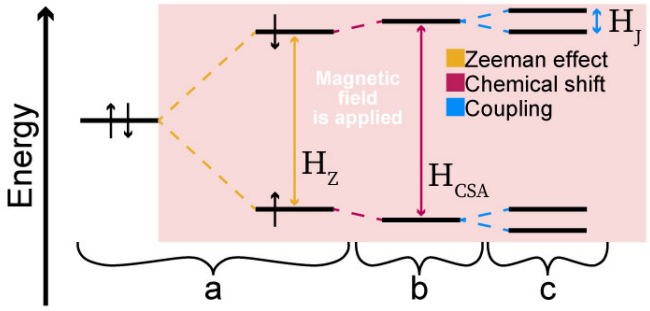


Figure 2.10: Splitting of degenerate energy levels during an NMR experiment due to different factors contributing to the Hamiltonian.

For example, due to the induction effect, a proton bonded to an oxygen will be significantly more deshielded than one bound to a methyl group. This difference in electron density will affect the local magnetic field experienced by the nucleus, thus changing the Larmor frequency (Equation 2.15) and changing the frequency recorded by the receiver coil. This results in the different chemical shifts observed for nuclei of the same element in different environments.

The Hamiltonian operator represents a system's total energy due to its dynamics and quantum mechanical properties. In an NMR experiment, we can use this to describe the change in energy of degenerate spin-up and spin-down states when the magnetic field is applied, as shown in Equation 2.16. Figure 2.10 shows that nuclei are split when the magnetic field is applied to those in the spin-up and spin-down states due to the Zeeman effect. H_Z (Zeeman Hamiltonian) describes applying a magnetic field of strength B_0 to a state where spin-up and spin-down states are degenerate. This is preceded by H_{CS} (chemical shielding Hamiltonian), which describes the change in these states due to the local magnetic field experienced by a nucleus due to its chemical environment. The final term in Equation 2.16 is H_J (J-coupling Hamiltonian). This describes the change in energy levels from the coupling of nearby nuclear spins. This arises from the spin interactions of magnetic dipole-dipole interactions, which couple nuclei together.

2.2.4 Relaxation time constants

Relaxation time constants describe the time an excited state requires to return to thermal equilibrium from an excited state in terms of longitudinal and transverse relaxation.

2.2.4.1 T_1 relaxation time

T_1 is referred to as the spin-lattice or longitudinal relaxation time. This describes how fast the system's z-axis components of nuclear spin relax back to the thermal equilibrium state (Figure 2.11). A long T_1 time constant would suggest that a system slowly relaxed from the excited

$$H = H_Z + H_{CS} + H_J \tag{2.16}$$

The Hamiltonian corresponding to the change in state energy levels when a magnetic field is applied during an NMR experiment. Where

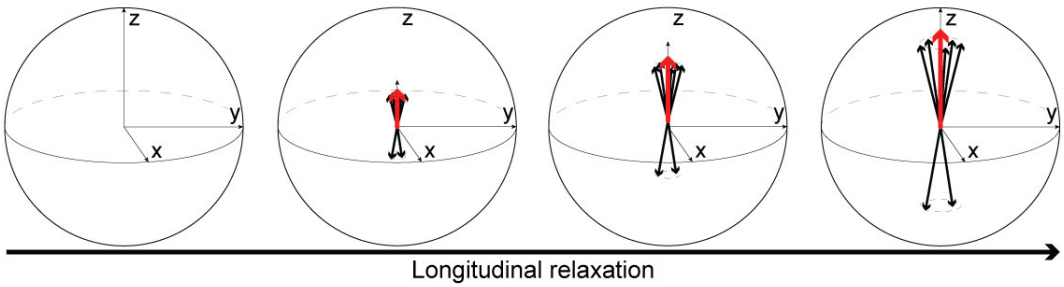


Figure 2.11: A vector-based depiction of longitudinal relaxation from a saturated state to thermal equilibrium.

state to thermal equilibrium. This is usually the case in rigid, immobile systems where there are little interactions between molecules. On the other hand, a short T_1 value would suggest a system had high mobility. This is due to molecules in the system being able to transfer energy between themselves rapidly.

2.2.4.2 T_2 relaxation time

T_2 is referred to as the spin-spin or transverse relaxation time. T_2 describes the decay of an NMR signal in the transverse dimension rather than the longitudinal dimension, as T^1 does. This means in physical terms that T_2 measures the time spins take to fall out of phase following a 90° RF pulse. As spins disperse, the net transverse component of the nuclear spin gradually decreases, as shown in Figure 2.12. A long T_2 time constant means spins remain coherent for longer, which could suggest less interaction, interference or an inhomogeneous applied field. This could indicate that a system is well-ordered or rigid. Short T_2 time constant suggests a system is heterogeneous or highly dynamic, creating many interactions between molecules that result in the loss of spin coherence.

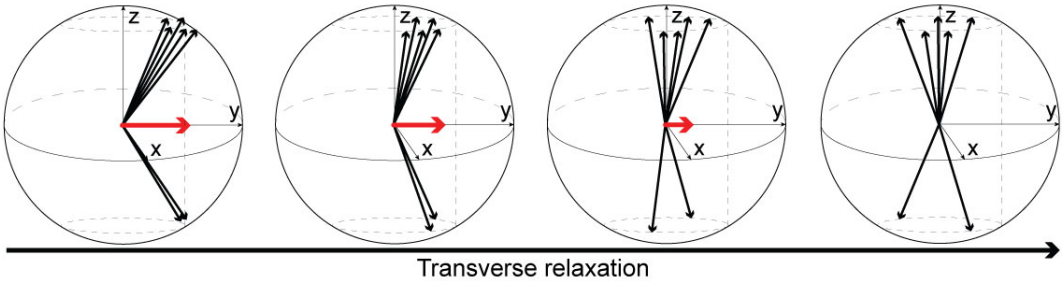


Figure 2.12: A vector-based depiction of transverse relaxation from an excited state to thermal equilibrium.

2.2.5 One-dimensional NMR

Figure 2.13 shows the pulse sequence of a simple 1D NMR experiment. It should be noted that the vector diagram representing the process can only fully explain the quantum mechanics involved in a single spin system. This illustrates the FID's origin and most information necessary to interoperate 1 and 2D data reported in this thesis. Applying an RF pulse rotates the bulk magnetisation by 90°, translating it onto the y-axis. This perturbation creates an excited state

from which the system will gradually return to thermal equilibrium. As described in Section 2.2.2, the time-dependent magnetic field this creates will be recorded as an alternating voltage in the NMR machine receiver coil. By performing a Fourier transformation on the FID, these shifted frequencies can then be separated and plotted as the intensity of their chemical shift. This produces the characteristic 1D NMR spectra commonly seen with the chemical shift reported in parts per million (ppm).

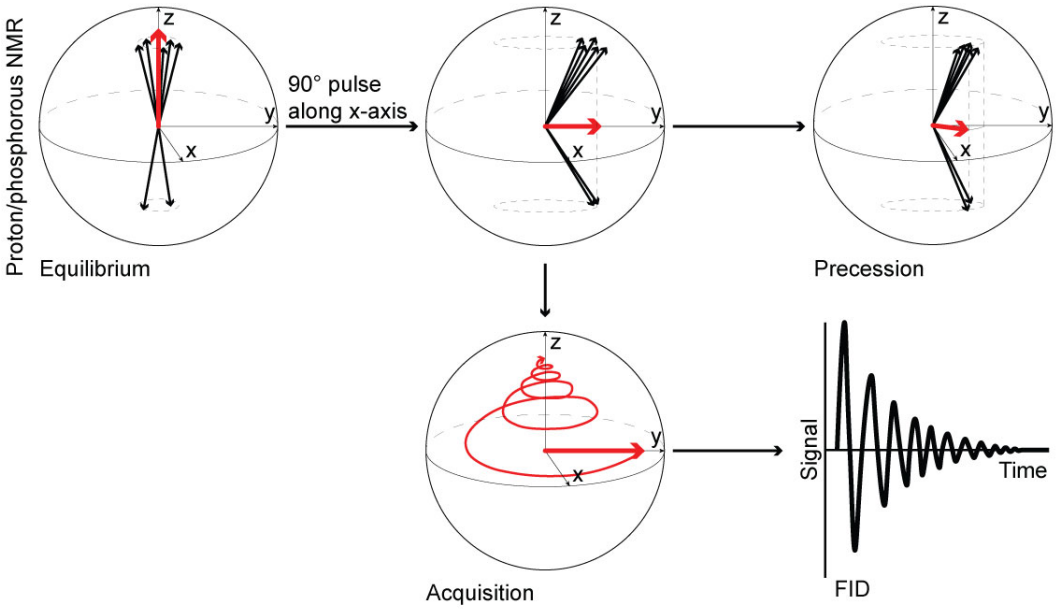


Figure 2.13: The vector model of the changes made to bulk magnetisation in a 1D NMR experiment. Showing the sample's bulk magnetisation in red and individual nucleus contributions as black arrows. Spin evolution during the acquisition is marked by a red line. The diagram marked "precession" is not a different pathway but a snapshot of the system shortly after it begins to decay back to equilibrium during "acquisition".

2.2.5.1 Saturation recovery

The saturation recovery experiment relies on a short RF pulse, equally populating the up- and down-spin states (Figure 2.14). This removes any net magnetisation from the nuclei. The nucleus magnetisation is then allowed to recover partially during the recovery time, in which the magnetisation in the z-axis will grow while returning to thermal equilibrium. More of the magnetisation will recover depending on how long this recovery time is. When a $90°$ pulse is performed to acquire the FID, the spectra will be more intense when left for longer. This is achieved for a series of recovery times, and the intensity is then plotted against time. By then fitting an exponential plateau to the data, the T_1 time constant can be determined, as is shown in Figure 2.14(b). Longitudinal relaxation happens due to the thermal radiation of energy from the bulk magnetisation because of collisions between atoms, as mentioned in Section 2.2.4.1. This can be used to measure how chaotic the environment around a species is, as more chaotic movement will result in far more frequent collisions and thus, the system will return to thermal

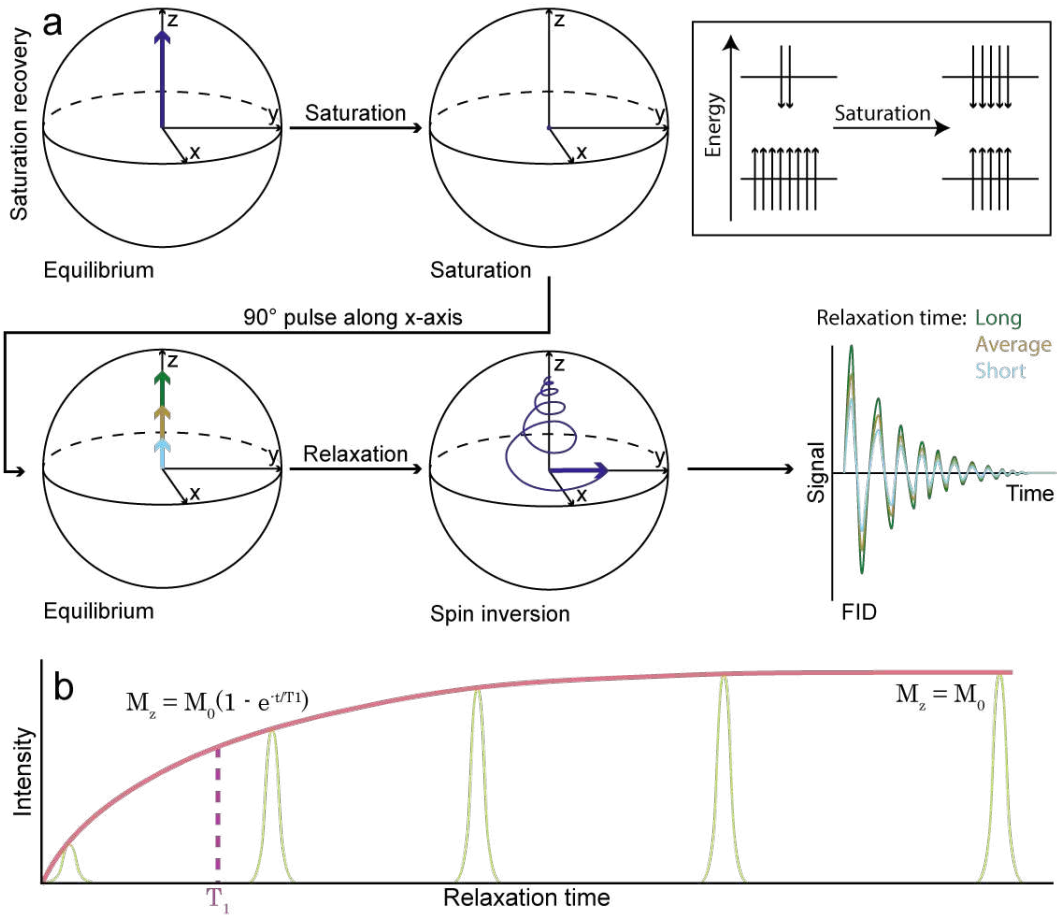


Figure 2.14: (a) The vector model of a saturation recovery NMR experiment. Showing the (purple) bulk magnetisation and its recovery after a (light blue) short, (tan) medium and (green) long relaxation time. Spin evolution during the acquisition is marked by a spiral. (b) The T_1 recovery curve displaying the dependence of bulk magnetisation (M_Z) on the equilibrium magnetisation (M_0) and the T_1 time constant at the current time (t). The T_1 time constants value is marked at the point where the bulk intensity has recovered by about 63%.

equilibrium faster.

2.2.6 A brief introduction to solid-state NMR

Traditionally, NMR dealt with molecules in the liquid state. This was because liquid-state systems have high mobility of molecules in solution, which results in sharp, well-defined peaks. Because of the rapid tumbling of molecules in a liquid state, anisotropic interactions with the magnetic field are averaged, resulting in well-defined peaks that represent an average of these orientations. If molecules being analysed did not tumble, then different orientations of anisotropic environments relative to B_0 would result in different peaks in the spectra reported. This was a significant caveat of solid-state NMR, as contributions to the FID from many environments oriented differently to B_0 would create broad peaks which were hard to resolve. Although this can be extremely useful for techniques such as static ^{31}P NMR spectra of lipid vesicles, this makes experiments performed on many systems difficult or impossible to resolve. Despite the lack of tumbling, the magic angle spinning (MAS) technique has since been established to create well-resolved spectra from solid-state samples. This caveat can also be a significant advantage of solid-state NMR as it provides the ability to report on structural features without molecules of interest tumbling in the solution or being interfered with by the solvent. The lack of solvent means molecules are constrained in a solid-state and remain in less dynamic positions; this can give better structural insight due to the lack of chaos introduced by tumbling. The absence of solvent also removes interference in the spectra from the solvent.

2.2.6.1 Magic angle spinning

MAS spins a sample along the diagonal of a cube ($54.7356°$) to remove anisotropic effects from a solid-state sample (Figure 2.17(a)). Similar to how molecules tumbling in solution will average all orientations relative to the magnetic field, MAS will similarly average conformations by changing their orientation relative to the applied magnetic field. By doing this, the chemical shielding anisotropy (CSA) is averaged along all three axes. This effect on the perturbation of peaks in NMR spectra can be seen in Equation 2.17.

When spun along the diagonal of a cube, the three principal components of the CSA tensor

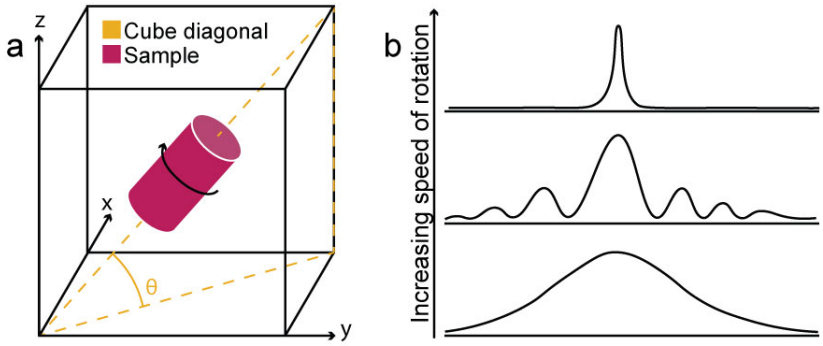


Figure 2.15: A simple, visual depiction of the sample setup of the magic angle spinning technique.

$$H_{CSA} = \frac{1}{2}\sigma_{CSA}(3I_z^2 - I(I+1)) \tag{2.17}$$

The equation for the CSA Hamiltonian, where H_{CSA} is the CSA Hamiltonian, σ_{CSA} is the CSA tensor, I_z is the z-component of the spin operator and I is the nuclear spin quantum number.

(σ_{xx}, σ_{yy}, σ_{zz}) which describe the anisotropy of the nuclei are averaged across all three dimensions. This causes the H_{CSA} to become progressively smaller as the rate of rotation is increased, thus decreasing the contribution to H_{CS} in the NMR Hamiltonian (Equation 2.16, Figure 2.17(a)). This technique can also simplify spectra to a single peak for techniques such as T_1 recovery experiments, where it is easier to monitor the recovery of a single peak rather than an entire lineshape.

2.2.7 Static ^{31}P NMR

As mentioned, static ^{31}P NMR takes advantage of the lack of tumbling in solid-state systems to report on the dynamics of lipid vesicles as a combination of all anisotropic environments. Figure 2.17 depicts the lineshape reported in a static ^{31}P NMR experiment. As this technique is static, the spectra produced are lineshape comprised of many peaks (shown in blue), all corresponding to similar ^{31}P environments oriented differently relative to B_0. Simply represented in Figure 2.16, lipids can be oriented parallel to B_0 (z-axis), perpendicular, or somewhere in between, logically meaning that there will be significantly more lipids representing environments which are closer to an equatorial orientation. As stated in Section 2.2.6.1, the CSA tensor describes anisotropy in the chemical shielding of nuclei. Because there is so little mobility in solid-state systems, ^{31}P nuclei oriented differently to B_0 will proceed differently as the nucleus CSA is oriented differently, relative to B_0. This means nuclei oriented differently to B_0 will have a different contribution from H_{CSA} as the local magnetic field experienced by the nuclei is different due to the orientation of the nuclei and the nucleus CSA. This creates a different

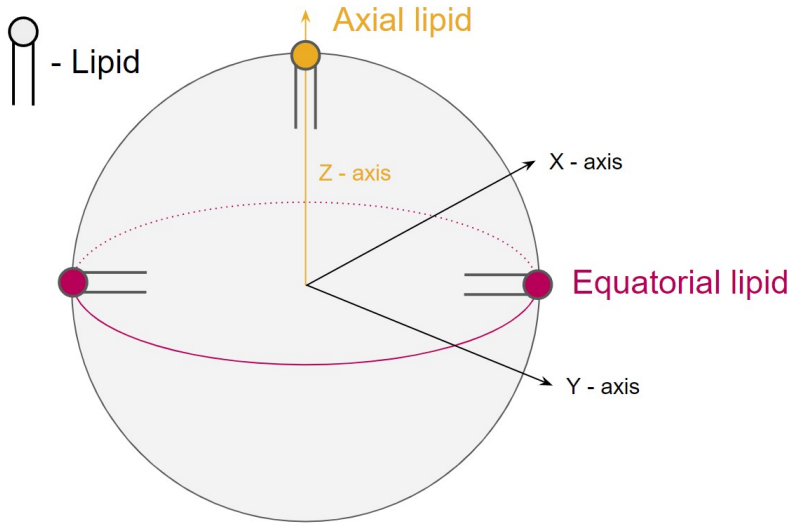


Figure 2.16: A simplified representation of lipids in an (yellow) axial and (red) equatorial orientation relative to a lipid vesicle.

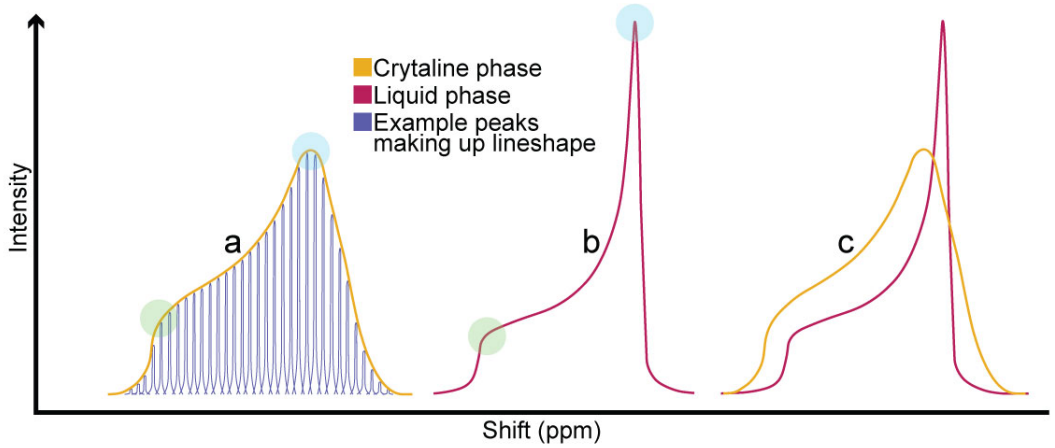


Figure 2.17: A visual representation of basic lineshapes reported by a ^{31}P NMR experiment for lipid vesicles. Highlighting the downfield edge in green and the upfield edge in blue.

separation of their spin-up and spin-down states. The difference in these states across all orientations means the ^{31}P nuclei in the lipid vesicle will process at a range of frequencies corresponding to all orientations in the lipid vesicle.

Highlighted in Figure 2.17 are points in green and cyan. The point highlighted in cyan corresponds to an environment in the vesicle where many ^{31}P nuclei exist in very similar environments. This is generally because there are significantly more equatorial lipids (relative to B_0) in the vesicle, all existing in similar environments. A well-ordered, rigid bilayer will contain lipids in well-defined conformations. This creates a broad lineshape as many well-defined conformations contribute to the spectra. On the contrary, a more fluid bilayer will allow lipids to change conformation more readily, averaging the CSA more and contributing to greater averaging of conformations. This increased mobility, therefore, narrows the lineshape, as CSA is partially averaged.

2.2.8 Nuclear Overhauser effect spectroscopy

2.2.8.1 Nuclear Overhauser effect

The nuclear Overhauser effect (NOE) is an energy transfer from one nucleus to another nearby via cross-relaxation. When the sample is excited, nuclei will couple via dipole-dipole interactions, transferring energy between them. Cross-relaxation is energy transfer, enabling spin-states to redistribute from the excited state to thermal equilibrium. This cross-relaxation affects the relaxation rate of the nuclei involved and can increase (positive NOE) or decrease (negative NOE) the signal from the respective environment. A positive NOE affects the rate at which nuclei return to thermal equilibrium by enhancing T_1 relaxation. A negative NOE will do the opposite, suggesting a decrease in T_1 relaxation. Positive NOE relies upon dipole-dipole interactions, which fall off rapidly over longer distances. Because of this, the NOE is very effective over short distances and can be used to report spatial relations in systems. As discussed in Sections 2.2.4.1 and 2.2.4.2, nuclei in a system will transfer energy between themselves and affect the rate at which they return to thermal equilibrium. These same interactions

that determine the spin-lattice time constant allow us to determine the spatial relationship of nuclei.

2.2.8.2 Technique

Nuclear Overhauser effect spectroscopy (NOESY) is an NMR technique which relies principally upon the NOE. The NOESY technique supplies sufficient mixing time during the experiment for nuclei to relax through cross-relaxation. During this relaxation time, nuclei will collide and transfer energy via spin-lattice relaxation. The characteristic NOESY spectra are created by acquiring an FID at two points during the experiment. Initially, a $90°$ RF pulse is applied along the x-axis ($90°$x pulse). While the excited state decays in this period (t_1), an FID is acquired as the spins evolve under the influence of interactions within the system. A second $90°$x pulse is then applied, after which the spins were left to develop in the NOE mixing time (t_m). Following t_m, a final $90°$y pulse is applied to avoid loss of transverse coherence. This transverse magnetisation is the same as that referred to in Section 2.2.4.2 regarding spin-spin relaxation. Following the coherence-generating pulse, a second FID is acquired for period 2 (t_2). The FIDs for the t_1 and t_2 periods are then Fourier transformed to create the spectra F_1 and F_2, respectively. These spectra are then processed and plotted along the x and y axis to create a contour plot. These spectra show a series of peaks along the diagonal, corresponding

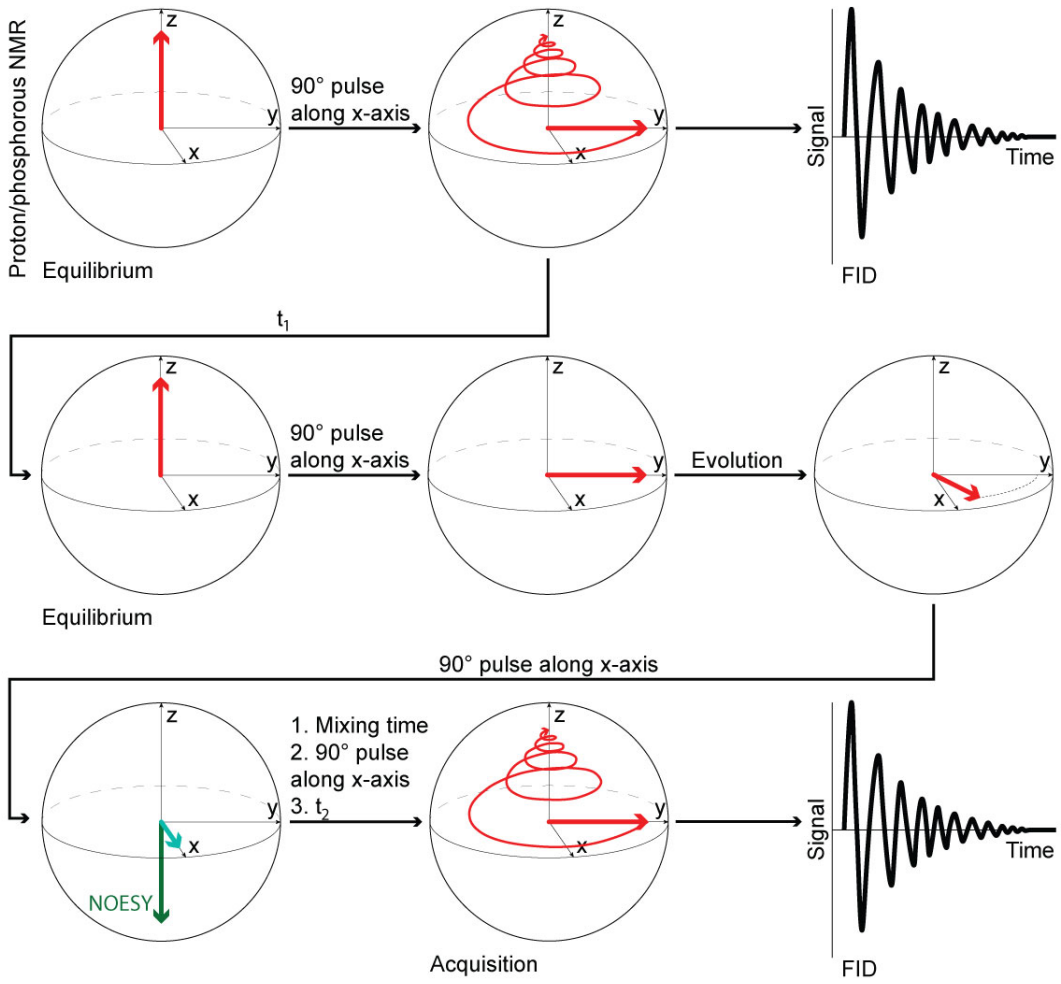


Figure 2.18: The vector model of the changes made to bulk magnetisation in a 2D NOESY NMR experiment. Showing the (red) bulk magnetisation and its separate components, including (green) NOESY. Spin evolution during the acquisition is marked by a spiral.

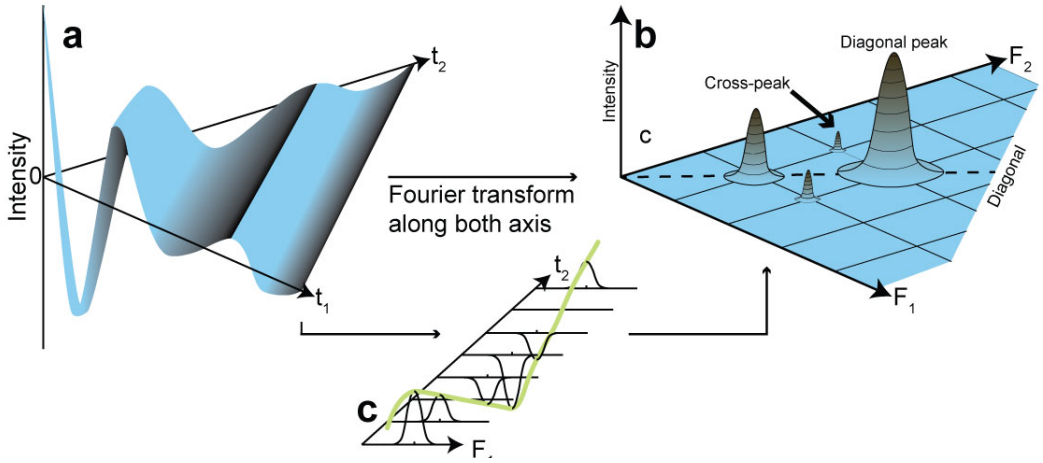


Figure 2.19: (a) A visual representation of a 2-dimensional FID and the corresponding spectra. (b) A resulting 2D NOESY contour plot indicating the position of diagonal and cross-peaks. (c) A midpoint in the 2-dimensional Fourier transformation where t_1 has been Fourier transformed, but t_2 has not been.

to each environment in the sample. Peaks on the off-diagonal indicated points where signals for peaks in F_1 and F_2 were enhanced due to the NOE generated during t_m.

2.2.8.3 Magnetisation exchange rate

Following the generation of 2D NOESY spectra, the magnetisation exchange rate between environments can be calculated. This is possible as this rate is directly proportional to the volume of the cross-peak generated between the two environments. The exchange rates are calculated using the same method used by Feller *et al.*[123], stemming from the work of Holte *et al*[124]. The Solomon equations are used in NMR to calculate the relaxation rate in two-spin systems[125].

$$\frac{dI_{1Z}}{dt} = R_Z^1(I_{1Z} - I_{1Z}^0) - \sigma_{12}(I_{2Z} - I_{2Z}^0) \tag{2.18}$$

$$\frac{dI_{2Z}}{dt} = R_Z^2(I_{2Z} - I_{2Z}^0) - \sigma_{12}(I_{1Z} - I_{1Z}^0) \tag{2.19}$$

$$\frac{dI_{1Z}I_{2Z}}{dt} = R_Z^{12} 2 I_{1Z} I_{2Z} \tag{2.20}$$

The Solomon equations for a two spin system describing the change in the spin states populations over time with respect to their self-relaxation rates. I_Z is the population of the respective spin state along the z-axis, R is the longitudinal relaxation rate of the respective spin state, I_Z^0 is the population of the respective spin state along the z-axis at thermal equilibrium and σ_{12} is the coupling constant of the two spins.

The Solomon equations, as shown in Equations 2.18 to 2.20, described the change in the population of the spin states in a two-spin system over time. If written for all spins in a system, this can be formalised as a series of differential equations describing the relaxation rate of all spins represented in the spectra. At any given mixing time, a measure of the relaxation rates for all diagonal and cross-peaks are provided by the more succinct expressions

shown in Equation 2.21 and its rearrangement in Equation 2.22, following from the Solomon equations.

$$\mathbf{A}(t_m) = exp(-Rt_m)\mathbf{A}(0) \tag{2.21}$$

$$\mathbf{A}(t_m)\mathbf{A}(0)^{-1} = exp(-\mathbf{R}t_m) \tag{2.22}$$

The matrix-based approach to the Solomon equations for a system containing multiple proton environments, where $\mathbf{A}$ is the peak volumes, $\mathbf{R}$ is the rate of magnetisation exchange between sites and t_m is the mixing time. Variables denoting a matrix are shown in bold.

The rearranged left-hand side of Equation 2.22 is diagonalisable, and by diagonalising this in Equation 2.23, it becomes possible to solve for the relaxation matrix.

$$\mathbf{A}(t_m)\mathbf{A}(0)^{-1} = \mathbf{X}\mathbf{D}\mathbf{X}^{-1} \tag{2.23}$$

$$\mathbf{R} = -\frac{1}{t_m}\mathbf{X}log(\mathbf{D})\mathbf{X}^{-1} \tag{2.24}$$

The relaxation matrix is derived from the matrix-based version of the Solomon equation. Where $\mathbf{A}$ is the matrix of peak intensities at a given time, $\mathbf{X}$ is the matrix of eigenvectors, $\mathbf{D}$ is the diagonal matrix of eigenvalues, $\mathbf{R}$ is the matrix of magnetisation exchange rates and t_m is the mixing time.

As seen in Equation 2.24, this results in a simple means of calculating the relaxation rate of all diagonal and cross-peaks.

Chapter 3

Methods development

3.1 Development and use of the Heatmapper python module

3.1.1 Abstract

This section outlines the development of a Python module that explores the interactions of a probe molecule with other molecules in a system. The focus of this module is on recording the contacts between the probe molecules and larger structures such as proteins or bilayers. In complex systems containing alcohol or numerous small molecules undergoing transient interactions, the average of these interactions can be difficult to monitor. To ensure consistency and ease in assessing these contacts, I developed a module that performs calculations and presents results interactively in HTML format.

3.1.2 Introduction

Understanding the contacts involved in processes such as the destruction of larger structures by sanitisers is key to understanding the mechanism of this process. These processes heavily rely on specific contacts between large structures and sanitising molecules. This section introduces a Python module developed to streamline the analysis of these contacts, providing a consistent and efficient way to explore this complex interaction network through visual depiction as a heat map.

3.1.3 Challenges in analysing pathogen-sanitiser contacts

Analysing systems involving sanitising agents, especially those containing alcohol, presents several challenges. Due to the size of these molecules, they undergo transient interactions, which are hard to monitor due to how rapid they are. Because of this, determining the significance of interactions can also be difficult, as these interactions may persist for only a few frames.

3.1.4 The interactive heat map module

I present an interactive heat map module to address the challenges in analysing pathogen-sanitiser interactions. This module initially calculates the number of contacts per residue or atom. This contact data can be exported and applied to analysis data separately or pushed to a heat map of the structure in the molecule. Pushing the contact data to a heat map will produce an interactive structure that can be rotated in the web browser, as shown in Figure 3.1.

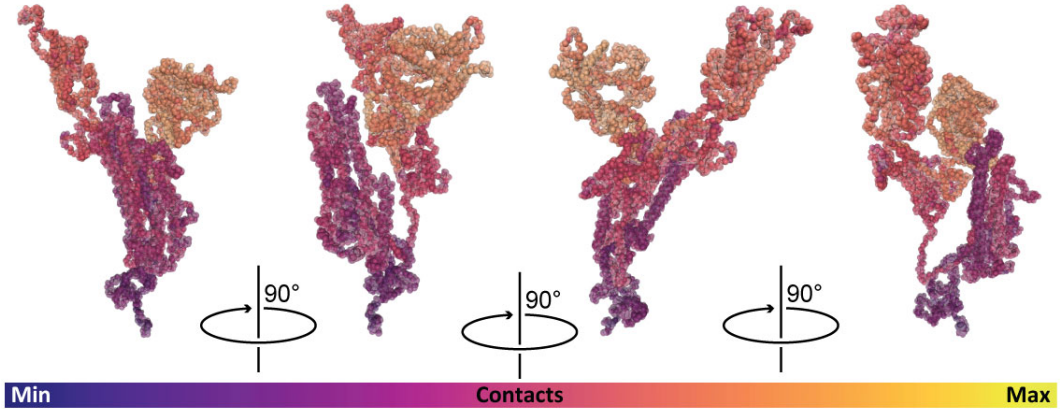


Figure 3.1: A visual depiction of the interactive HTML element of the heat mapper module. Showing a monomer of the S protein ECD reported in Chapter 6 rotated through 360°

.

3.1.5 Benefits of the interactive heat map

The interactive heat map offers several advantages. Firstly, this allows for exploration of the data in 3D. This can be extremely useful for structures such as proteins where features such as pockets are hard to distinguish in a 2D image. The Python library used for visualisation (NGLview) also provides the ability to highlight atoms and draw distance, similar to VMD. Secondly, this enables rapid identification of points of interest. In work such as that presented in Chapter 6 this made it easy to quickly identify areas of the protein which were experiencing more interaction with the solution than others. Finally, this method removes the risk of applying a confirmation bias to observations in the simulation. In expecting to see interactions which were hypothesised before running the simulation, it is possible that other unexpected interactions could be overlooked. Applying a heat map in this way makes it obvious which interactions are significant.

3.1.6 Conclusion

In conclusion, the development of this Python module for interactive heat maps provides a valuable tool for researchers and analysts studying pathogen-sanitiser interactions. It addresses the challenges of transient interactions in complex systems and offers a data-driven approach to identifying key aggregation points and interaction dynamics. This tool enhances the accuracy and objectivity of the analysis, ultimately contributing to a deeper understanding of the mechanisms involved in pathogenic species destruction.

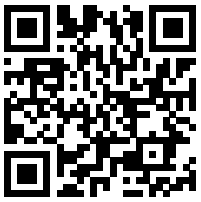
https://github.com/callumj321/Heatmapper 

Figure 3.2: The Github repository containing the code for the heat mapper module.

3.2 Investigating structural properties of a protein via importance values of decision tree regression

3.2.1 Data collection

The first step of analysis involved the collection of relevant data. This involves collecting a dataset of features and labels relevant to the study. This dataset consisted of time-series data of the contacts to each residue in the protein during production, as produced by the heat map module described in Section 3.1.

3.2.2 Data preprocessing

Before conducting any analysis, the data must be preprocessed to ensure the quality and consistency of the dataset. This preprocessing included:

- **Data normalisation or scaling**: I standardised the numerical features to be integers to improve the model's stability.
- **Qulaitative data noramlisation**: Qualitative data such as AA type or other qualities were standardised to integers to improve stability.

3.2.3 Feature importance calculation

The preprocessed data was used to train a random forest (RFT) machine learning algorithm to predict the number of contacts per residue. This model was trained with the contacts per residue as a label and properties of the residue as features. Using this I calculated the feature importance scores using the trained model to determine the areas of interest. The importance scores were obtained through tree-based feature importance which allowed me to extract feature importance scores directly from the model's internal structure for tree-based models, like RF.

3.2.4 Visualization and interpretation

The feature importance scores were visualised using appropriate graphs, such as bar plots. We interpreted these scores to identify the areas of interest with the most important features. An important feature means that the model relied upon it heavily when making decisions as to the number of contacts for a residue.

Chapter 4

All-atom MD simulations of the action of alcohol and chlorhexidine on the membranes of *E. coli* and *S. aureus*

The work presented in this chapter is published in the Journal of Molecular Biology[126].

4.1 Abstract

Membranes form the first line of defence of bacteria against potentially harmful molecules in the surrounding environment. During the initial phases of infection, it is imperative to the pathogen's effectiveness that the envelope is maintained until the point where it has successfully infected the host. Because of this, the most common means of defence against pathogens has become a preventative method based on destroying the envelope. By applying sanitisers to ourselves and contact points in environments prone to outbreaks, we can limit the size of outbreaks or, ideally, avoid them altogether. Understanding the protective properties of these membranes represents an important step toward developing targeted antibacterial agents such as sanitisers. Propanol (PROH), isopropanol (ISOP) and CHX can heavily decrease the threat imposed by bacteria and viruses in the face of growing antimicrobial resistance via membrane disruption mechanisms.

Here, I have employed MD simulations to specific lipid membrane models of the inner and outer membranes of *E. coli* as well as the cell membrane of *S. aureus*. This research was performed to determine how sanitisers affect bacterial membranes. By exploring the impact of CHX and alcohol on these membranes, I identify key aggregation points. In doing this, I determine whether there are specific interactions or conformations upon which this process relies. This shows that the initial stages of sanitisation are more nuanced than previously thought, drawing attention to areas that could potentially develop antibacterial-resistant adaptations in the

future. Knowing how sanitisers work is essential, as one day, scientists may have to determine why new adaptations in bacteria render them less effective.

4.2 Introduction

Pathogenic bacteria pose an enormous threat to human, animal, and plant health, especially given the rate at which resistance to current antibiotics is developing[127, 128]. Effective methods of sanitisation, such as antimicrobial hand and body scrubs, are essential in the control of widespread diseases as they offer a route to decreasing the rates of infection through human contact[129, 130, 131, 132]. Understanding the mechanisms of action against different bacterial membranes is essential in developing new and more effective sanitising agents.

4.2.1 Sanitisation and alcohol

Short-chain alcohol molecules are the most commonly used sanitising agent and are theorised to kill bacteria primarily by denaturing the membrane[92]. The effectiveness of alcohol as an antibacterial agent is well known and has been proven as an effective means of eradicating a plethora of pathogens in vivo, including *E. coli* and *S. aureus*[129, 130]. This deformation is largely attributed to macroscale perturbations imposed upon the membrane, a model elucidated by Feller *et al*[133].

Further, it has been shown that introducing short-chained alcohols to a PL membrane can significantly affect the surface area per lipid (APL) and membrane thickness[134]. More recently, this phenomenon was explored by Ghorbani *et al.*[135], who found significant evidence in MD simulations to support these theories. Their work also showed that following diffusion from the bulk sanitiser solution, there is minimal diffusion of alcohol once distributed into the membrane. This lack of diffusion implies alcohol is partitioning into a bilayer region, which is thermodynamically preferable.

Alcohols contain an OH functional group bound to an aliphatic carbon. Depending on the nature of the bonding to this carbon, the alcohol will be primary, secondary, tertiary, or methanol. In the case of primary alcohols, the aliphatic chain length can have a significant effect on how the molecule behaves, with shorter chains creating effective sanitisers and anaesthetics, whilst longer chains are used as detergents and surfactants[136, 137]. Surfactants actively trap hydrophobic species and dissolve them in aqueous solutions. This happens because of the hydrophilic-hydrophobic nature of these species; these are fatty alcohols with a long hydrocarbon tail. This process is spontaneous and occurs purely due to hydrophobic-hydrophilic interaction, both enthalpically and entropically favourable[138].

Short-chain, sanitising alcohols are also effective due to their chaotropic nature; instead of forming micelles, they form dispersive interactions in a cell membrane when used as sanitisers. As PLs are also amphiphilic in nature, small alcohols form mutually energetically beneficial interactions with them beneath the head groups. However, this packing results in membrane thinning, adversely affecting the pathogen[135]. These small chaotropic molecules are also theorised to destabilise the secondary and tertiary structure of membrane-bound proteins es-

sential to the survival of a pathogen. By providing alternative interactions, protein structures can form alternate interactions, which disrupt the secondary structure and result in the denaturing of the protein[132]. Short-chain alcohol molecules kill pathogens primarily by disrupting the membrane, similarly to CHX[92]. The value of topically applied alcohol as an antibacterial agent is well known and has been proven as an effective form of defence against many pathogens in vivo, including *E. coli* and *S. aureus*[129, 130, 131].

4.2.1.1 Chaotropic agents

Chaotropic molecules such as alcohol can disrupt the structure of water and alter how the hydrophobic effect stabilises larger, complex structures such as proteins, membranes or macromolecules[139, 140]. The chaotropic nature of these molecules means they can change how these larger structures solvate, a phenomenon with many example publications cited in the literature[141, 142, 143, 144, 145, 146, 147]. Chaotropes do this by creating a shell surrounding the ordered molecule, providing favourable interactions between said shell and the environment[148]. Bilayers and proteins rely upon specific interactions to maintain their structure. Depending on the concentration, this means that when chaotropes are applied, the ordered structure becomes distorted or falls apart entirely. This effect has been observed to affect both protein and lipid structures. As a result, adaptations to diminish this effect have been seen in bacterial membranes and proteins[149].

4.2.2 Alcohol-lipid interactions

Alcohol is well known to act as an effective sanitiser[150]. As mentioned, alcohol's effectiveness is principally a result of its disruptive effect on lipid bilayers (Section 4.2.1). In this section, I explore the initial interactions responsible for the overall destruction of membranes by short-chain alcohols.

Lipids are simplistically composed of polar regions (headgroups) and hydrophobic regions (tails). The boundary of these two regions is where alcohol interactions occur, as both molecules are amphiphilic. Interactions between short-chain alcohols and lipid headgroups

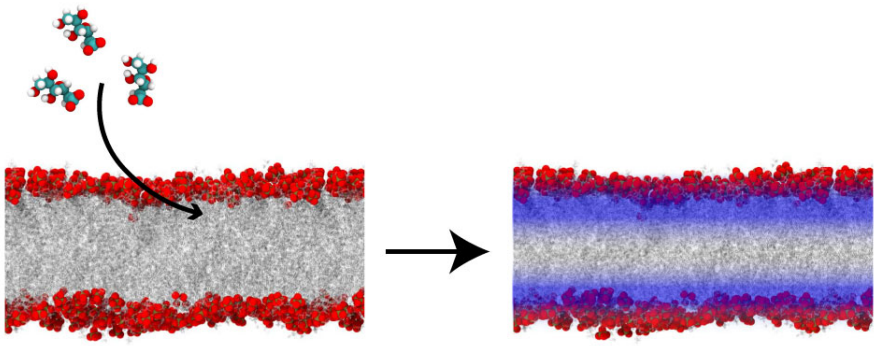


Figure 4.1: The proposed distribution of PROH into a simple lipid bilayer without showing disruption for the sake of simplicity. The final expected distribution is shown as a blue gradient.

are a vital feature in the mechanism of membrane deformation. Patra *et al.* detail interactions between methanol and ethanol with the headgroups of DPPC. They found that the alcohol molecules form interactions with the lipids around the ester groups at the top of their lipid tails, as shown in Figure 4.1[134]. This hydrogen bonding is essential in the initial stages of alcohol-bilayer interaction, suggesting that bilayer deformation depends upon alcohol partitioning into a specific bilayer region. This finding aligns with the effects mentioned in Section 4.2.1.

Unfortunately, research on the interactions between short-chain alcohols and lipid tails is sparse. Research in this area often describes the effect of increasing alcohol chain length on membranes[151, 152]. This research finds that as alcohol aliphatic chains increase in length, the APL of the membrane increases much more when they partition into the bilayer. Although these alcohols are much longer than the ones researched in this thesis, they detail an effect that we later find to be similar, that aliphatic chains in alcohols stack with lipid tails to minimise their interaction with the surrounding hydrophilic environment.

4.2.3 Chlorhexidine-lipid interactions

CHX is a bisbiguanide sanitiser. Although the effect of CHX has been the subject of previous research, detailed work in this area still needs to be more extensive. What is known is mainly summarised very well by Christopher G. Jones[153]. The structure of CHX, as shown in Section 1.4.4, is shown again in Figure 4.2 for the reader's convenience.

A key mechanism utilised by CHX to form initial interactions with membranes is charge-charge interactions between the cation and the membrane surface. Previous research has shown that changes in the lipid composition of membranes can lower the minimum inhibitory concentration (MIC) of CHX required[154]. This effect is further elaborated upon by Khokhar *et al.*, who explains that CHX is theorised to destabilise the membrane, causing the cell to leak[155]. The importance of the charge-charge interaction between the cationic CHX and anionic membrane components was further detailed in work by Van Oosten *et al.*[94]. In this work, they simulate CHX being pulled through membranes in AA and CG resolution and find that with a higher positive charge on CHX, there is less resistance to its binding at greater penetration depths. Further to the importance of charge interactions, it is known that the phenomenon responsible for actual cell death is generally a product of these charge interactions[156]. Work performed before this by Kilelee *et al.* states that a decrease in outer leaflet charge of the SaCM creates a

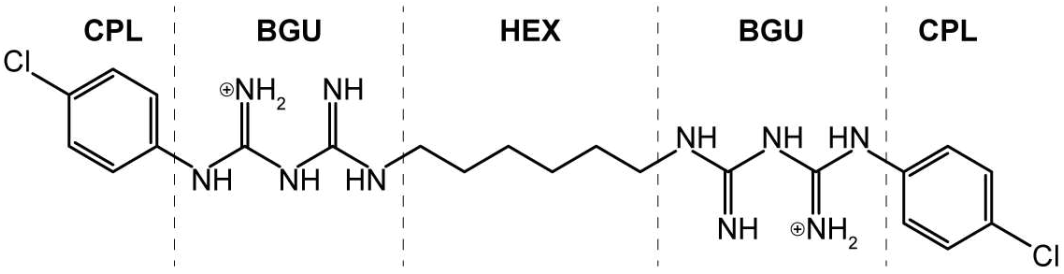


Figure 4.2: Separation of CHX into key functional groups, which are necessary for the interaction of CHX with a lipid membrane.

decreased likelihood of binding of antimicrobial peptides; research which is not unique in finding these resistances[11, 62]. Antimicrobial peptides utilise VDW and hydrophilic interactions to form interaction; these findings are vital if we hope to understand the mechanisms of other antimicrobials, such as those utilised by CHX when binding in the bacterial envelope[157].

Further to these interactions, CHX has been shown to work as an effective cooperative sanitiser with other chemicals such as Vancomycin; CHX decreases the MIC of Vancomycin required[158]. Here, they suggest that the presence of CHX can enhance the effectiveness of structural modifications caused by Vancomycin, while Vancomycin also improves the effectiveness of CHX. Further, they suggest that the effectiveness of CHX is mainly due to interactions that it has with lipid A, performing MD simulations of CHX with a membrane where it was docked to two lipid A molecules. Although this presents a very simplistic approach and does not assess how these molecules would pass the polysaccharide regions of LPS, their findings suggest a binding mode. They postulate that the BGU groups bind to the phosphate region while the CPL groups bury themselves in the hydrophobic region. This finding appears to agree with the previously suggested method of binding of CHX, detailed in Section 1.4.4.

4.2.4 Lipid-lipid Interactions

Interactions between different lipids in the membrane are responsible for the structural stability and fluidity, as mentioned in Section 1.4.3. These macroscale properties result from small, nuanced interactions between the molecules comprising the membrane. In this section, I explain the effect of lipids' interactions with one another on sanitiser effectiveness.

Murzyn *et al.* discusses the interactions between PE and PG in lipid bilayers and assesses how they affect the stability of said bilayer[159]. They discuss primarily hydrogen bonding between PE and PG in their simulated bilayers and note that PE will act as a hydrogen bond donor. Generally, these hydrogen bonds are PE $\cdots$ PE and PE $\cdots$ PG, while they note that PG $\cdots$ PG is exceptionally uncommon as PG very rarely acts as a H-bond donor; this is due to the alcohol functional groups in the headgroup rather than the amine in PE. Between these lipids, they also find that a common phenomenon is the formation of water bridges. The phenomenon of inter-lipid hydrogen bonding is elaborated upon by Doherty *et al.*, who finds water bridging between NH_3 and PO_4 to be common[160]. This is important as PE and PG comprise the majority of *E. coli* membranes. Understanding these interactions will allow us to assess how the addition of sanitisers affects them. The structure of PG and PE are shown in Figure 4.3.

With regards to interactions resulting from antipathogen-resistant adaptations, changes in the

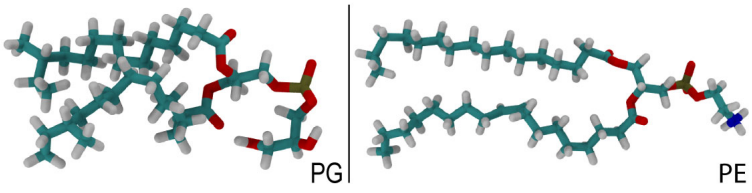


Figure 4.3: Comparison of the structure of PG and PE to show the headgroups regions proposed to form water bridges and hydrogen bonds between them.

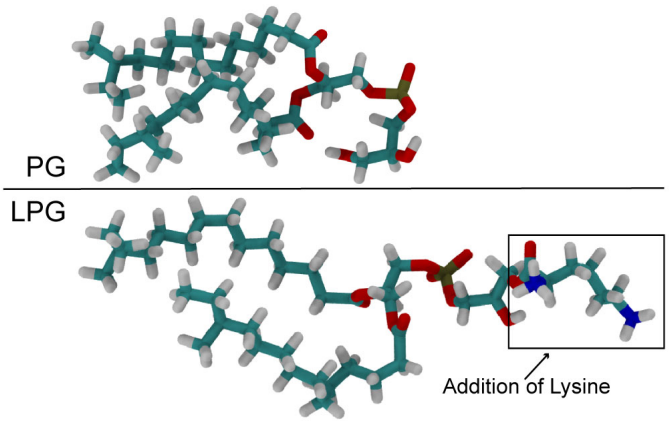


Figure 4.4: The structure of PG and LPG, with the additional Lysine in LPG, is highlighted to detail the modification.

lipid composition of the cell membrane have been instrumental. Bacterial species such as *S. aureus* are well-known to modify PG to LPG and then translocate it to the cell membrane[101]. PG is modified to LPG in a process called lysinylation, performed by the MprF protein. This protein mutates in resistant strains of *S. aureus* (Figure 4.4)[161]. The purpose of this process is to decrease the overall negative charge of the cell membrane; PG, which composes most of the outer leaflet, is negatively charged. The effect of this decrease is that cationic antibacterials, such as CHX, are less likely to bind to the outer leaflet. As such, the bacterial cell is less susceptible to our antibacterials. This inclusion of LPG in the SaCM and its effect on the bilayer stability is elaborated upon by Piggot *et al.* who finds it, although less resistant than the EcOM, to be very effective at maintaining membrane stability[99]. The structure of PG and LPG are shown in Figure 4.4.

4.3 Methods

4.3.1 Simulation system setup and model parameters

PL membrane compositions were based on those used by Piggot *et al*[99]. This is also true for the outer membrane's inner leaflet, while the outer leaflet is 100% Ra-LPS. These membranes were generated using CHARMM-GUI's membrane builder tool[162, 163, 164]. Membranes were then solvated in 0.15 M KCl and subjected to NVT and NPT equilibration steps before a 200 ns production run without sanitisers. All systems in this chapter were constructed by extracting an equilibrated membrane structure from a 200 ns production run of that membrane alone in 0.15 M KCl. Alcohol solutions were also simulated for 200 before addition to simulation systems to ensure they were well mixed. The systems were combined by extracting the coordinates of the membrane system at $t = 200$ ns and removing all water from the simulation system. The water was then replaced with the aqueous alcohol solution. Systems containing CHX in water had CHX added to the system by randomly inserting it into the solution. The experiment where CHX was placed in a membrane pore to assess whether it remained there was created by applying an electric field that caused poration of the membrane before stopping the simulation, positioning CHX and then resuming the experiment. CHX was parameterised via charm-GUI with standard parameters for CGenFF, and charge distribution was determined by the charge, which has previously been found to be most effective[165, 166].

4.3.2 Simulation protocol

Simulations were set up and performed using the GROMACS[167, 115] MD software package (version 2020.2) with the CHARMM36 force field[168, 169]. Systems were maintained at 310 K, using the Nosé-Hoover[170, 171] thermostat with a time constant of 1 ps. The system's pressure was maintained at 1 atm, with a time constant of 2 ps, using semi-isotropic pressure coupling with the Parrinello-Rahman[172, 121] barostat. All van der Waals interactions were cut off at 1.2 nm, and a smooth PME[173] algorithm was used to treat electrostatic interactions with a real space cut off of 1.4 nm. Simulation parameters were chosen based on similar published studies of Piggot *et al*[99]. Each system was subjected to 500 ps of NVT simulation for equilibration, followed by 2 ns of NPT. Positional restraints (1000 kJ mol^{-1} nm^2) were placed on the membrane head group atoms during NVT and NPT equilibration. Production simulations were then performed without any positional restraints. The results were analysed using GROMACS tools and in-house Python scripts. Visualisation was performed using the visual molecular dynamics (VMD)[174] software package. Electroporation of the membrane was performed with a 0.125 V nm^{-1} electric field applied along the membrane normal.

4.3.3 Calculating proportion of sugar interactions with the chlorophenyl region of chlorhexidine

The entire 200 ns of production was parsed through a Python script written in Jupyter Notebooks[175], which checked each timeframe for contacts within 2.5 Å of the CPL region via MDAnalysis[176] and then used this information to create a bar chart with Matplotlib[**Hunter2007**].

4.4 Results and discussion

In Table 4.1, a brief overview of the simulations discussed in this section is shown. These membranes were all approximately 10nm wide in both the x and y plane and approximately 15nm in the z axis (membrane normal).

System	Solute	Length (ns)	Repeats	Electric field along membrane normal (Vnm^{-1})
SaCM	None	200	3	None
SaCM	CHX	200	3	None
SaCM	CHX+PROH	200	3	None
EcIM	None	200	3	None
EcIM	CHX	200	3	None
EcIM	CHX+PROH	200	3	None
EcIM	CHX	200	3	0.125
EcIM	GLUC	200	3	None
EcOM	None	200	3	None
EcOM	CHX	200	3	None
EcOM	CHX+PROH	200	3	None

Table 4.1: A brief overview of the simulations performed and analysed in this section.

4.4.1 Chlorhexidine in aqueous solution: Action on bacterial membranes

Simulations of 200 ns were performed in triplicate where 0.5% w/v CHX in 0.15 M KCl was applied on either side of the three membranes (SaCM, EcIM and EcOM). Visual inspection of the system after 200 ns revealed minor perturbation of the membranes in any of the simulations. The insertion depth was evaluated by calculating the CHX density in the z-dimension (perpendicular to the membrane's normal). Figure 4.5 shows greater penetration into the sub-head group region of the SaCM compared to either of the *E. coli* membranes. There was a shift in the CHX density towards the centre of the membrane as the simulation proceeded, while no such shift was observed for the *E. coli* membranes. The deeper penetration of CHX was due to the higher negative charge in the headgroup region of the SaCM (-0.3 e per lipid), compared to the *E. coli* PL leaflets (-0.15 e per lipid) enabling stronger stabilising electrostatic interactions. This agreed with previously reported findings by Van Oosten *et al.*[94] that CHX bound more strongly with increasing charge disparity. There was no penetration into the sub-headgroup region of the LPS leaflet of the EcOM despite the high charge of LPS. The slow-moving, almost impenetrable nature of LPS has been well-documented in both experimental and simulation studies[70].

For the PL membranes, throughout 200 ns, the CHX molecules transitioned from partial interaction with the membrane on one terminal BGU-CPL region to mostly being intercalated and bound by both termini, as illustrated in Figure 4.6. In this configuration, the BGU and CPL functional groups were buried in/beneath the phosphate region toward the ester functional groups, while the HEX group lay along the bilayer surface. The interaction observed here was termed a 'c-shape' mode of binding and was reminiscent of the wedge proposed by Komljenović *et al*[93]. It should be noted, however, that the wedge they proposed was an inverted conformation, with the CPL-BGU region located in the lipid headgroups and the HEX functional group buried in the lipid tail region.

The mean squared diffusion (MSDif) of lipids in the SaCM and EcIM when exposed to CHX was

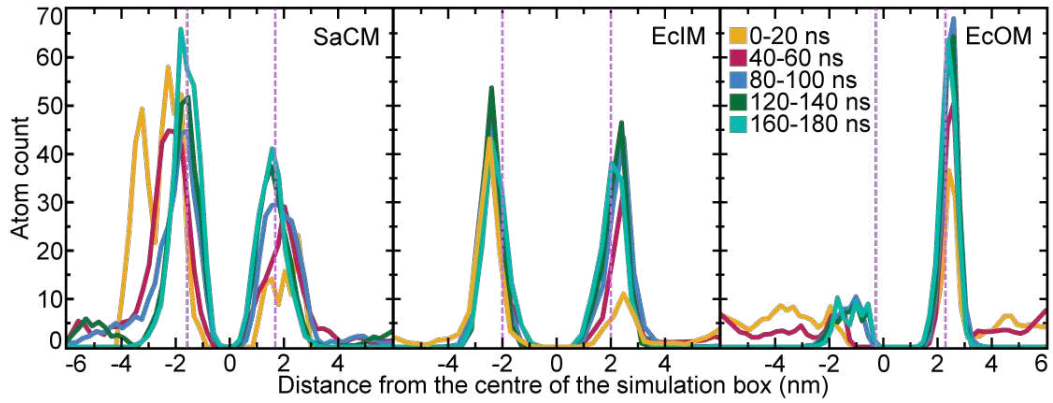


Figure 4.5: 20 ns block average density of CHX at 40 ns intervals along the z-axis of the simulation box for the entire 200 ns production run with average head group phosphate position as a pink, dotted line. Please note that these densities are calculated from a single repeat to avoid convoluting the figure; however, these findings are representative of repeats.

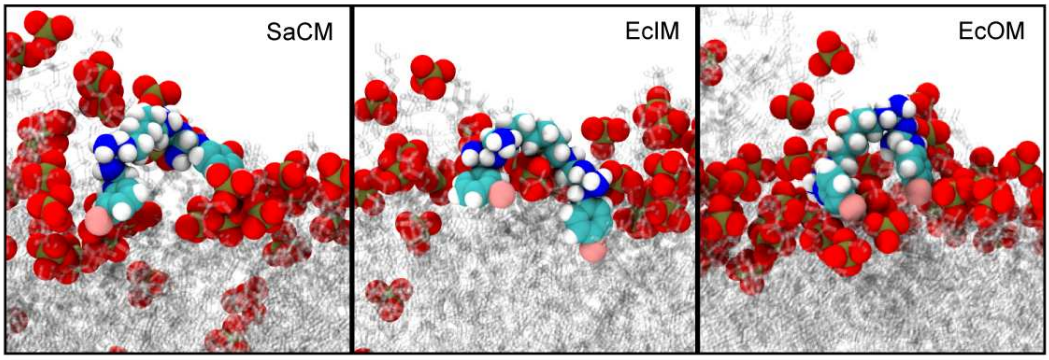


Figure 4.6: Snapshots of CHX C-shape binding in the SaCM, EcIM and EcOM systems are shown as spheres with phosphate oxygen in red, phosphate phosphorous in gold, carbon in cyan, nitrogen in blue, chlorine in pink, hydrogen in white and lipids as translucent.

tracked to assess the effect of c-shape binding on lipid mobility (Figure 4.7). Due to headgroup interactions between PG and LPG in the SaCM, the MSDif was much lower. However, both membranes had a lower MSDif after 200 ns of production with CHX than they did without, especially the EcIM. This confirmed an effect suggested in previous research that CHX would bind lipids together and decrease mobility[154]. To assess the stability of said conformation, the CPL-CPL distance of CHX in the SaCM and EcIM systems was measured. This showed a considerable degree of fluctuation, but the average distance was approximately 13 angstroms after 200ns of production in both the SaCM and EcIM system (Figure 4.8).

The solvent-accessible surface area (SASA) of the CHX was calculated for the simulated trajectories to assess any potential clustering of CHX molecules, but this revealed no net change upon membrane interaction. This indicated that their interaction with lipids did not induce CHX aggregation. Some CHX-CHX interactions between the BGU and CPL regions were observed, but these did not persist for longer than a few nanoseconds (Figure 4.9 (a-b)). The average APL of each membrane was calculated. The SaCM and EcIM membranes only exhibited a marginal difference when CHX was added, which is further elaborated upon in sections about these membranes in alcohol (Figure 4.9(c)). This showed an increase of 0.09 and 0.01 $Å^2$ when CHX was added to the SaCM and EcIM systems, respectively; this minimal

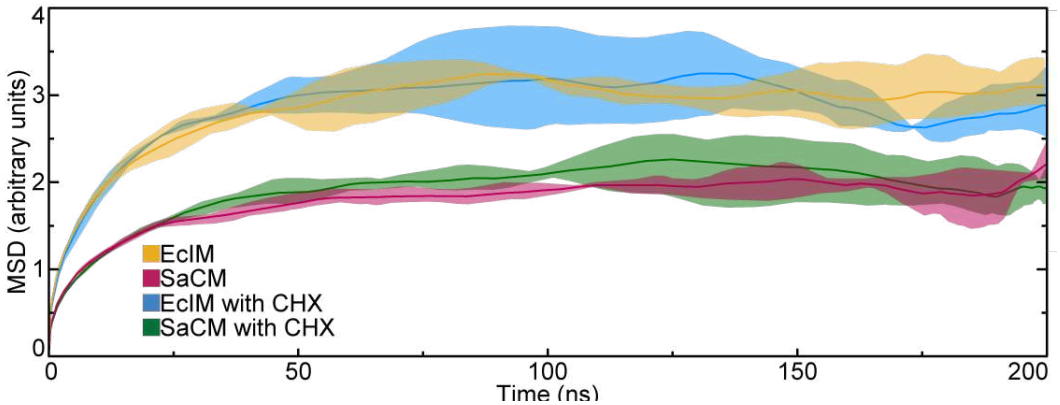


Figure 4.7: The running average of the mean squared diffusion coefficient of lipids in the EcIM and SaCM both with and without CHX, showing the error as a shaded region.

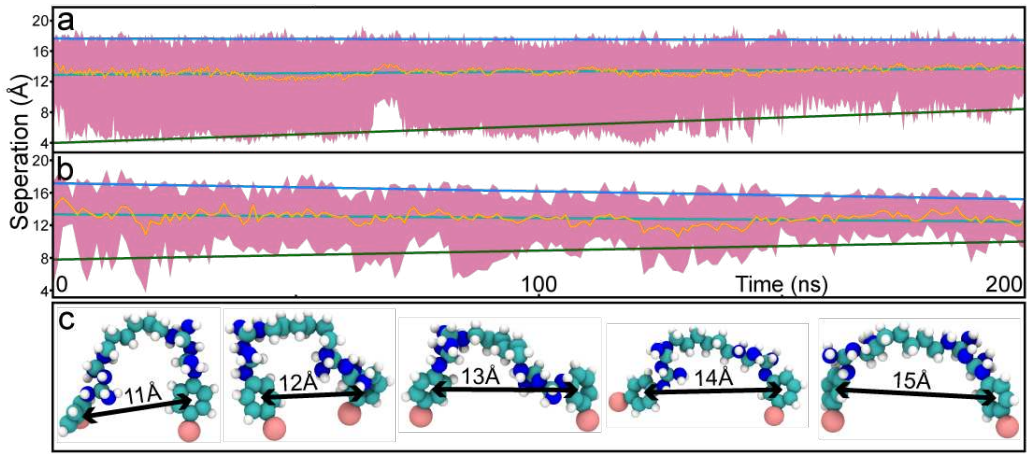


Figure 4.8: Conformational analysis of CHX in the SaCM and EcIM membranes exposed to CHX in water. (a) Separation of CPL regions in the SaCM system, showing (yellow) average, (red) error, (blue) trend in the maximum, (green) trend in the minimum and (cyan) trend in average. (b) Separation of the CPL regions in the EcIM system, coloured as in (a). (c) Visual representation of CHX at integer separations between 11 and 15 Å, showing (cyan) carbon, (white) hydrogen, (blue) nitrogen and (pink) chlorine.

difference was due to CHX intercalating with the lipid headgroups, increasing the APL. As CHX was not restrained in the equilibration steps to allow for mixing, the binding had already occurred at $t = 0ns$. The density of the membrane components as a function of distance from the membrane centre was compared at the start and end of the simulations, revealing no apparent change in any repeat (Figure 4.10). Thus, the binding of CHX to the membranes studied here did not cause any notable displacement of the membrane components. The EcOM

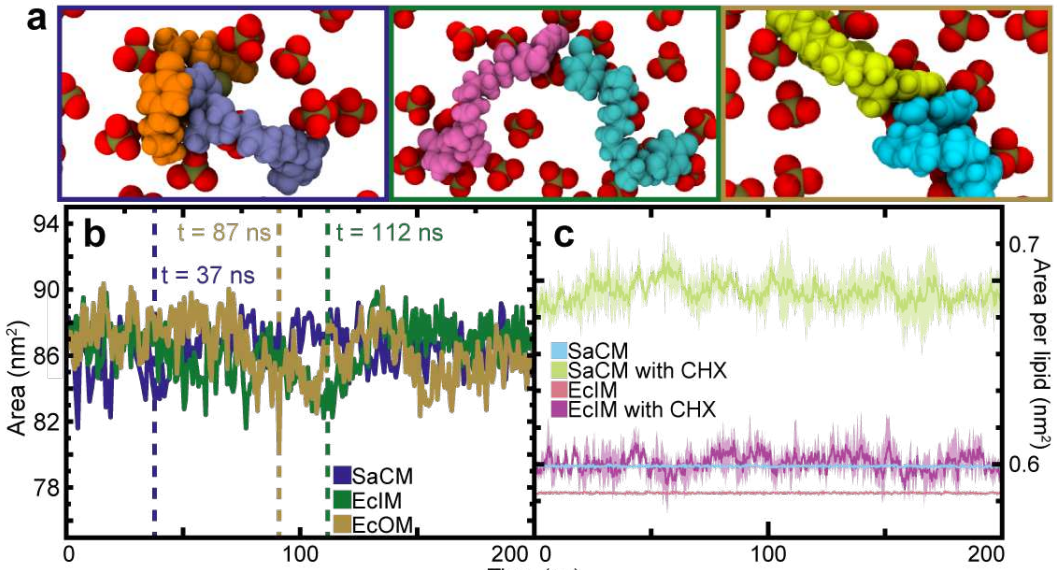


Figure 4.9: Clustering and APL in systems exposed to CHX without alcohol. (a) Visual snapshots depict transient CHX interactions at 37 ns (SaCM), 112 ns (EcIM) and 87 ns (EcOM). Snapshots show PL headgroup phosphate (gold) phosphorous, (red) oxygen and (coloured by residue number) CHX. (b) SASA of all CHX molecules combined in (purple) SaCM, (green) EcIM and (tan) EcOM systems. (c) APL of EcIM and SaCM with and without 0.5% w/v CHX throughout 200 ns of production in triplicate.

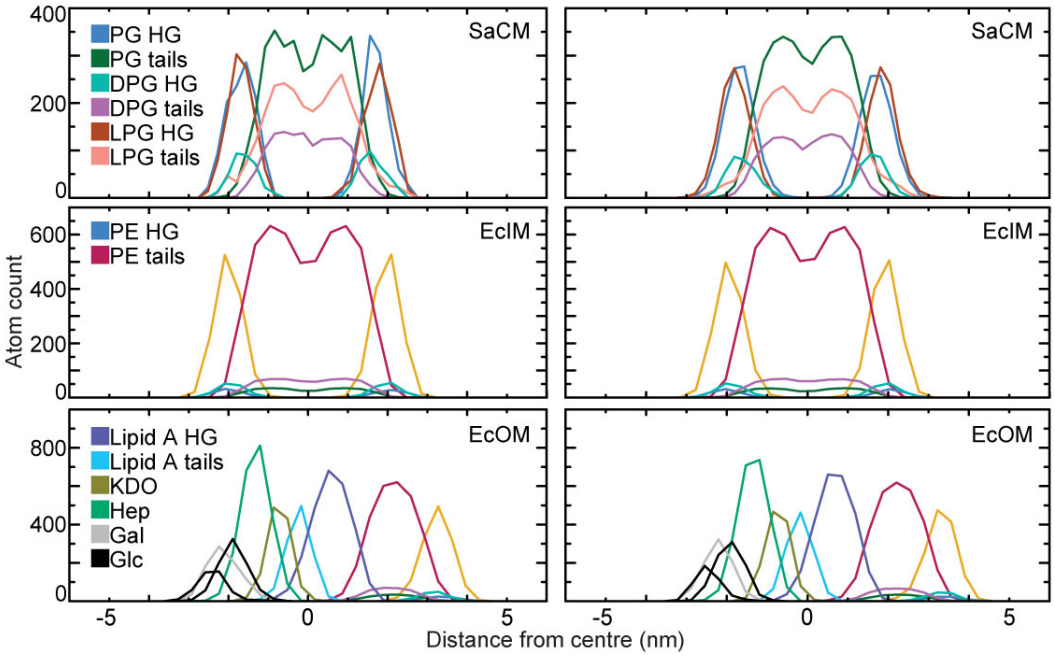


Figure 4.10: The density of membrane components along the z-axis of the SaCM (top), EcIM (middle) and EcOM (bottom) systems exposed to CHX in 0.15 M KCl solution, at (left) 0 and (right) 200 ns. Densities were averaged over 20 ns.

showed very little difference between systems with or without CHX.

4.4.2 Chlorhexidine in aqueous alcohol: Action on phospholipid membranes

To determine the effect of chaotropic sanitisers on PL membranes, alcohol and CHX were applied to the SaCM and EcIM. Given the intrinsic antibacterial activity of alcohol[177, 136, 150], determining the additional impact of CHX on bacterial membranes was a point of interest. We therefore performed 200 ns comparative simulations of two PL bacterial membranes (SaCM, EcIM) in the presence of 20% w/v PROH and ISOP solutions with and without CHX, in triplicate. This concentration of alcohol was used because higher concentrations risk membrane deformation to the point that a semi-isotropic system can no longer be maintained. These works are focused on understanding the process involved in membrane dissolution rather than destroying the membrane.

Simulated PL membranes were surrounded by 0.5% w/v CHX in 0.15 M KCl with either 20% PROH or ISOP in bulk water on either side of the membrane. Visual inspection revealed lipid dispersal in the SaCM occurring rapidly ($t = 45$ and 50 ns for PROH and ISOP, respectively), becoming even more deformed throughout the 200 ns of production. This deformation was much faster in the EcIM system ($t = 10$ and $t = 30$ ns for PROH and ISOP, respectively). The timeframe of lipid dispersal was extrapolated from the rate of change of the simulation box width (Figure 4.11). Throughout simulations, the lipids in these systems reoriented in a manner that disrupted the canonical bilayer structure of the membrane. Figure 4.12(a) shows that there is a much faster initial increase in APL for the EcIM when alcohol is applied, especially for PROH. This is due to hydrogen bonding between PG and LPG in the SaCM; these hydrogen bonds hold the headgroups together and provide an initial resistance to the dispersion caused by alcohol. However, this value plateaus at a lower value in the EcIM systems than in the SaCM ones, owing to a higher membrane charge in the SaCM (-0.3 e per lipid) compared to the EcIM (-0.15 e per lipid). Thus, the EcIM was impacted less by the alcohol/alcohol+CHX compared to the SaCM. This was evident from a gradual increase in APL, which plateaued

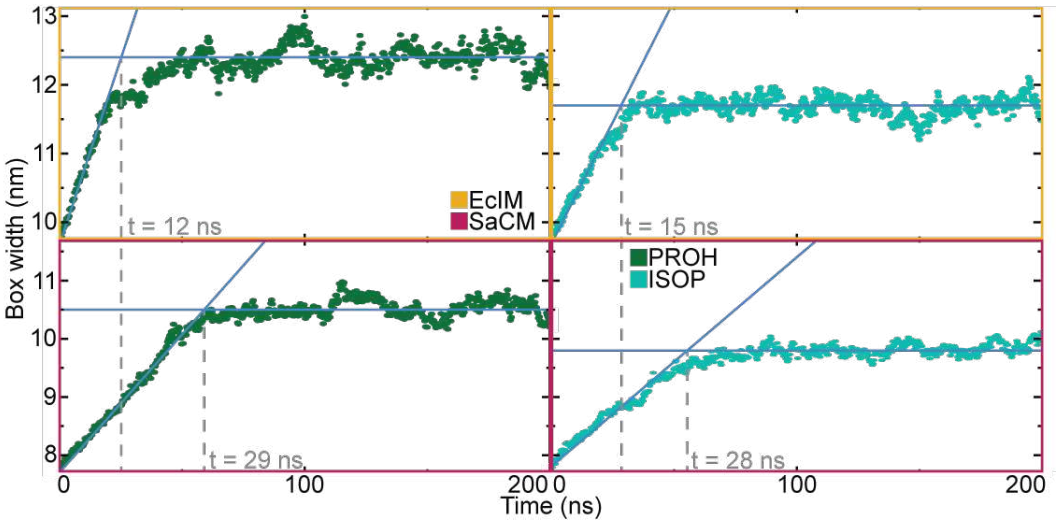


Figure 4.11: Extrapolation of the lipid dispersal time in (yellow) EcIM and (red) SaCM when exposed to (green) PROH and (cyan) ISOP, showing the best fit for the rate of dispersal and plateau in blue and the time extrapolated in grey.

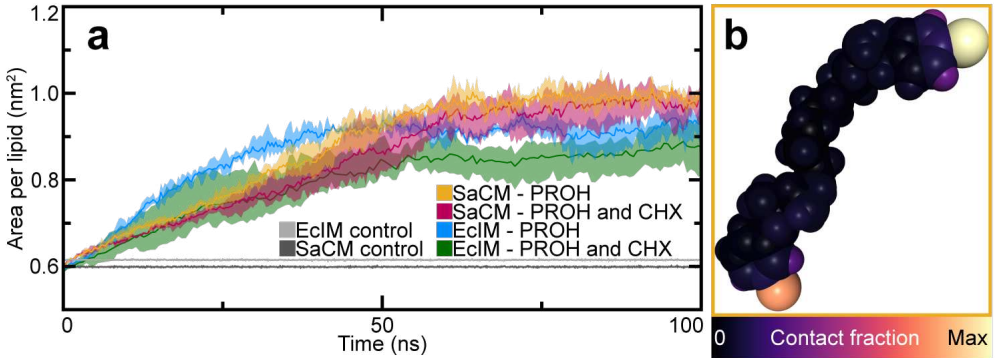


Figure 4.12: Analysis of the SaCM and EcIM when exposed to PROH solutions with and without CHX. (a) APL measurements of the SaCM and EcIM over the first 100 ns of production show the average value after the experiment in triplicate as a full line and error between the repeats as a shaded region. APL measurements of the membrane in 0.15 M KCl are shown in grey. (b) The structure of CHX heat is mapped by the frequency of contacts from each atom to the headgroups of PE over the final 10 ns of production.

at a value approximately 80% greater than the equilibrated membrane for the SaCM systems and approximately 60 and 70% more for the EcIM systems both with and without CHX, respectively. The lower final APL in the EcIM system, which also contains CHX, is due to CHX interactions with NH_3^+ functional groups in the headgroup of PE, which composes 90% of the bilayer. These interactions bind together lipid headgroups like a molecular staple. This results in a high proportion of contacts to chlorine in the CPL region of CHX while the rest of CHX experiences almost no interaction with the headgroups (Figure 4.12(b)). An analysis of the frequency of contacts of CHX with different lipids in the membrane was performed (Figure 4.13). This analysis showed far more likely interaction with DPG in both PL membranes due to the size of the lipid and the fact that it contains far more interface regions.

Alcohol molecules became distributed in the lipid core regions of the now distorted bilayers, allowing the CPL regions of CHX to bind beneath the membrane surface at the headgroup-tail interfaces, as seen in Figure 4.14(b and d). This binding occurred regardless but was more rapid and partitioned deeper when alcohol was present. The membrane deformation was characterised quantitatively by plotting the membrane density along the simulation cell

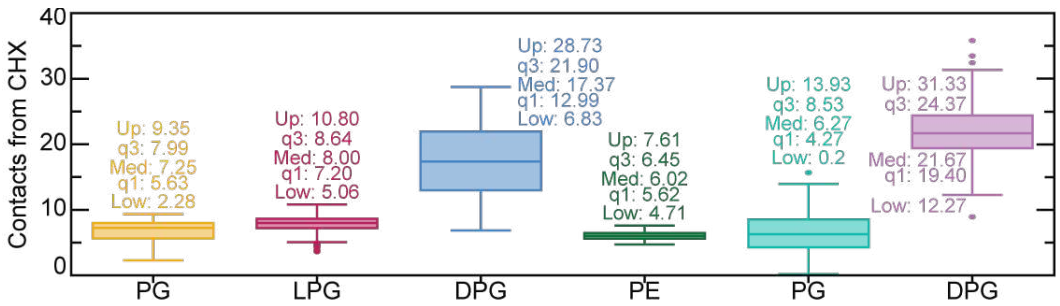


Figure 4.13: The total number of contacts from CHX to each lipid type in the SaCM and EcIM systems between 100 and 200 ns across all repeats normalised by their percentage contribution to their respective membrane.

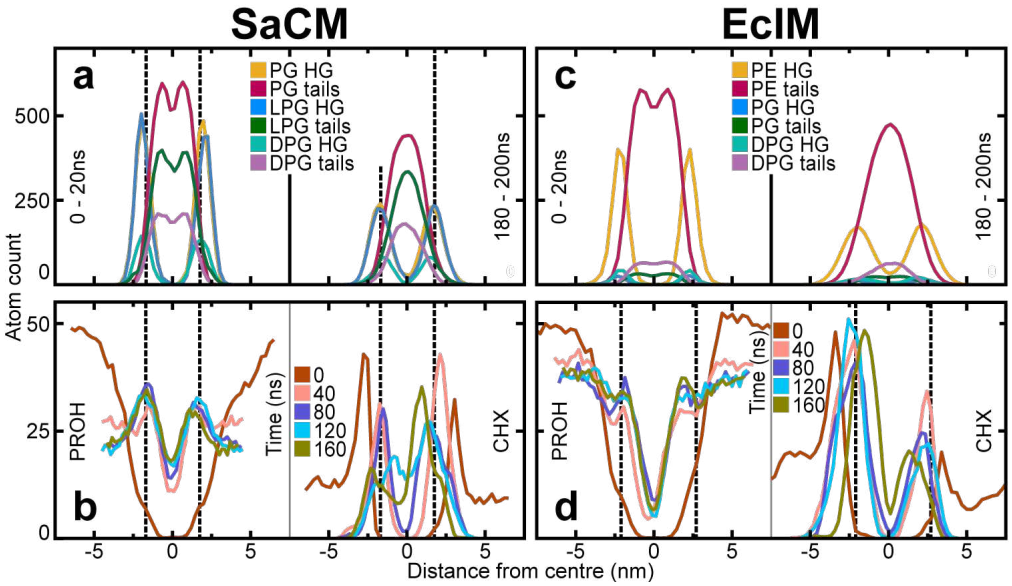


Figure 4.14: Measurements of the densities of membrane and sanitiser components along the z-axis of the simulation box in systems exposed to both PROH and CHX together. (a) Densities of the SaCM system membrane components exposed to PROH and CHX averaged over the first and last 20 ns of production. (b) Densities of PROH and CHX in the SaCM system over time. (c) Densities of the EcIM system membrane components exposed to PROH and CHX averaged over the first and last 20 ns of production. (d) Densities of PROH and CHX in the EcIM system over time.

z-axis (parallel to the membrane normal, Figure 4.14). System density plots also showed that alcohol distributed within the membrane by adopting energetically favourable positions at the headgroup/tail interface where it could minimise unfavourable interactions with water (Figure 4.14(b and d)); this positioning was seen in both PL systems. This positioning of PROH allowed the alcohol functional group to remain in the phosphate/ester region, where it maintained polar interactions. At the same time, its hydrocarbon tail could form lipophilic interactions with the lipid tail regions, minimising interactions with water.

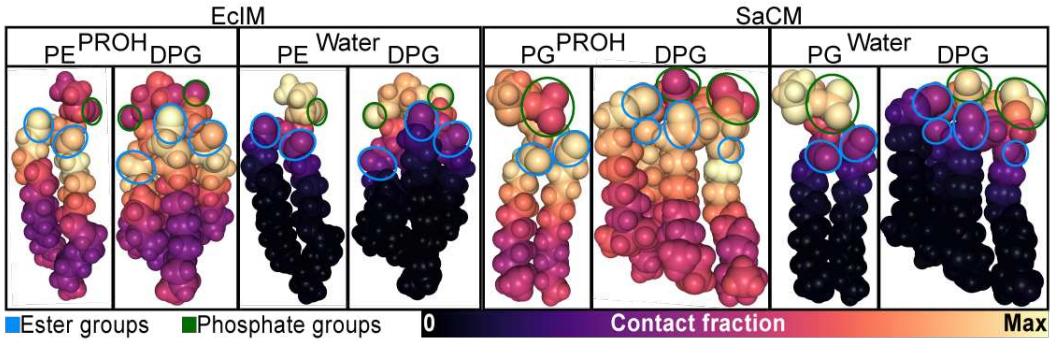


Figure 4.15: Contact mapping of PE and DPG within the EcIM system and PG and DPG in the SaCM system at 180 and 200 ns with PROH and water mapped linearly from (black) least to (yellow) most contacts per atom in the lipid, showing ester groups circled in blue and phosphate groups circled in green.

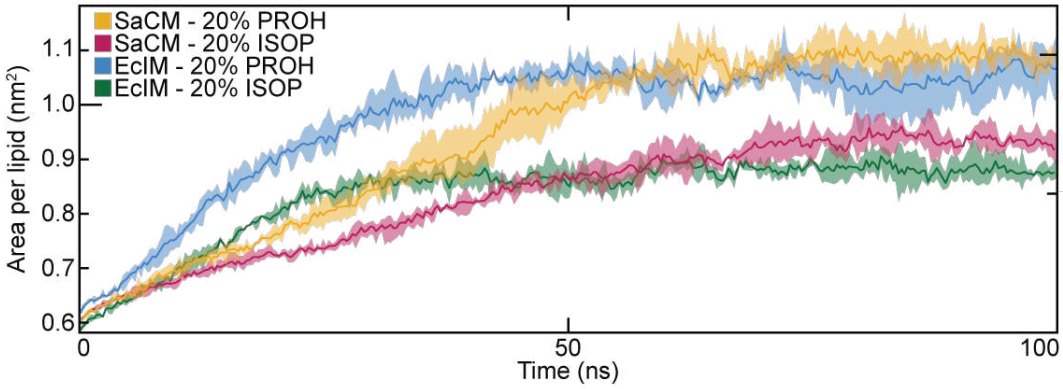


Figure 4.16: Comparison of the area per lipid over the first 100 ns of production for the SaCM and EcIM systems exposed to 20% w/v PROH or ISOP.

To illustrate this, I mapped the lipids in the SaCM and EcIM models through their contact with PROH and water using the heat map module described in Section 3.1. This clearly shows that alcohol accumulated much more around the headgroup/tail interface, where I propose to take energetically stable conformations (Figure 4.15). The absence of water interactions with the tails aligns with the observation that although severely deformed, neither membrane allowed pore formation. This observation is crucial as it shows that water did not enter the membrane during the early stages of deformation. A similar effect was observed for systems containing ISOP. However, the change was less dramatic since the branching in the secondary alcohol did not align as effectively with the lipid tails (Figure 4.16).

In Figure 4.17(b), the location of PROH partitioning in the membranes was determined by plotting the time that 100 PROH molecules in the SaCM and EcIM system spend in contact

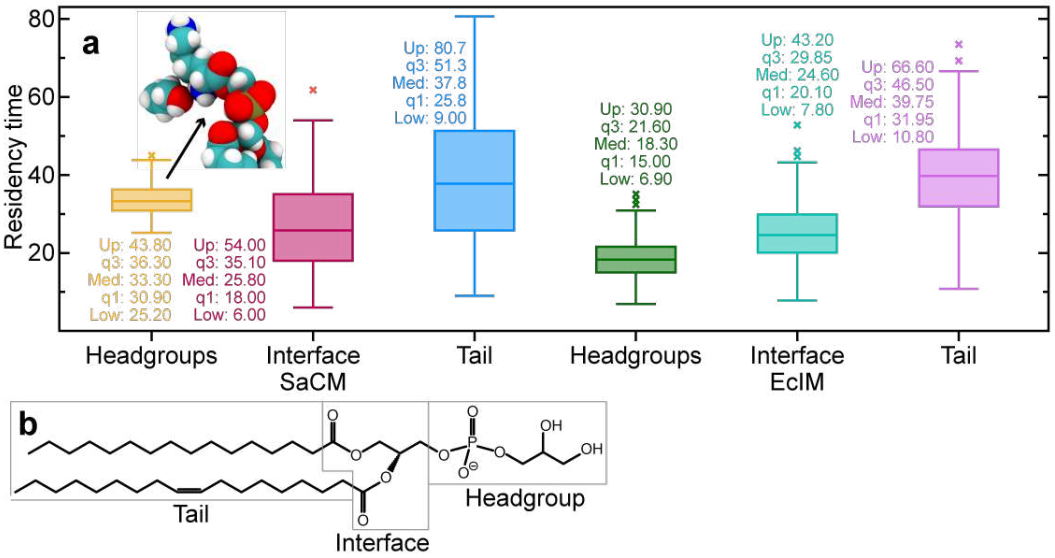


Figure 4.17: Analysis of the SaCM and EcIM when exposed to 20% w/v PROH and 0.5% w/v CHX in 0.15 M KCl. (a) A box plot of the residency times of 100 PROH molecules in each repeat of the SaCM and EcIM systems over 200 ns of production to defined regions in the SaCM. (b) Using PE as an example, a definition of the tails region (below the esters and phosphate groups) and the headgroup (above the phosphate).

with defined regions of the bilayer. The increased average residency time of PROH with the headgroups in the SaCM system is due to interactions with the large charge headgroup of LPG. The relatively well-defined residency time suggests that this is a strong interaction where PROH molecules occasionally get stuck. These results agree with similar computational work by Ghorbani *et al.*, which found ethanol accumulating at the headgroup/tail interface[135]. Analysis of system densities showed that CHX aggregated around the bilayer headgroups (Figure 4.14) where it interacted via the same c-shape interactions observed in the absence of alcohol. This binding mode is especially favourable as its polar termini can form interactions with the phosphate/ester region of the lipids at the headgroup-tail interface. Although the 'wedge' conformation proposed by Komljenović *et al.*[93] was observed occasionally, it was far less common than the c-shape conformation.

The results presented here for PL membranes are strongly supported by the findings of Rzycki *et al.*[178], who also observed that CHX molecules might bind individually rather than in clusters, and those that did bind as an aggregate would rapidly disperse. Evidence of this dispersal is shown in Figure 4.18. Similarly, they also found CHX to bind preferentially in the c-shape conformation and that the effect of CHX was minimal. Notably, density measurements of the

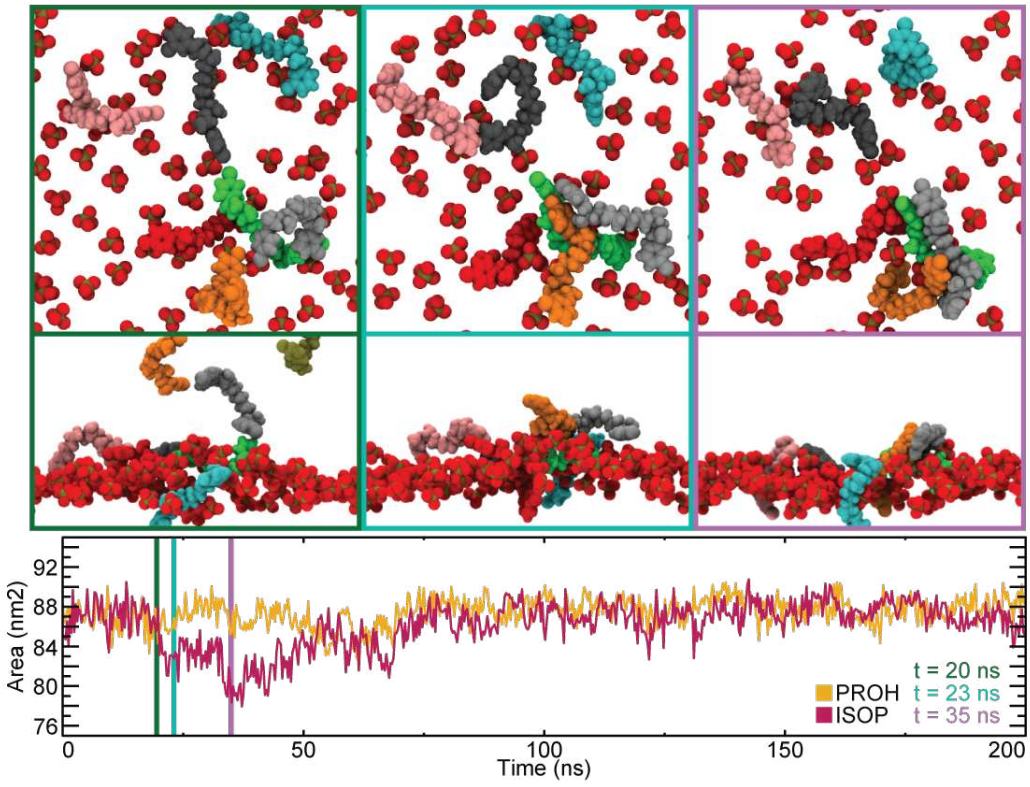


Figure 4.18: The SASA over time of SaCM model lipid membrane systems containing (yellow) PROH and (red) ISOP with CHX in aqueous 0.15 M KCl over 200 ns of production simulation. Visual snapshots of a CHX surface cluster forming in the ISOP system at (green) 20 ns, (cyan) 23 ns, and (pink) 35 ns show carbon CHX residues coloured by residue number, with lipid phosphate oxygen in red and phosphate phosphorous in gold. The top row of snapshots looks down on the membrane surface along the membrane normal, while the lower row looks perpendicular to the bilayer normal.

system showed less aggregation of CHX on the membrane surface of the EcIM relative to the SaCM, presumably due to the lower membrane charge and the fact that CHX is oppositely charged. Throughout the 200 ns production run, CHX molecules became bound by both of their BGU/CPL moieties, which then became embedded within the headgroup-tail interface of the membrane (Figure 4.18).

CHX 'bridging' was observed in the SaCM system; i.e. PROH caused deformation and thinning of the membrane such that CHX molecules bound to separate leaflets could contact each other across the bilayer, interacting via their BGU/CPL regions. However, there appeared to be no additional deformation compared to the alcohol systems, which did not contain CHX. This phenomenon was not observed in the EcIM systems. This is reminiscent of the 'handshake' binding mode proposed by Komljenović *et al*[93]. Their findings suggest CHX could operate via a similar mechanism to alcohol by penetrating the membrane and aggregating in the membrane centre. However, this was not seen in simulations of CHX without alcohol. The SASA of CHX in these systems was tracked to determine whether CHX molecules had any tendency to cluster. However, no suggestion was seen other than a dip for the SaCM system corresponding to a binding cluster of 4 CHX molecules, which dissipated within 30 ns of forming on the bilayer surface (Figure 4.18). To ensure that clustering was rare, I measured the average CHX separation for systems with and without alcohol. However, it did not indicate any significance. This was not seen in repeats but suggested the absence of clusters was due to CHX forming interactions with the membrane in preference to those it would form with itself. Contacts to CHX from the headgroups, interface and tails were mapped. The CPL regions and, to an extent, the BGU show most contact with the lipid tails as they are embedded in the membrane, while the interface sees most of CHX. The only atom in CHX to contact the headgroups regularly was chlorine due to a phenomenon where multiple NH_3^+ groups would coordinate with the halide due to its large size and electronegativity. This is likely instrumental in the initial phases of CHX-membrane interaction and why CHX remains tightly bound throughout production. This interaction is likely responsible for seriously reducing lipid dispersal when CHX was added to the EcIM system without water.

4.4.3 Electroporation driven insertion of chlorhexidine and effect of gluconate: MD study of the action on the *E. coli* inner membrane

To eliminate the possibility that membrane permeation of CHX into the EcIM was not observed due to insufficient simulation sampling, we initiated simulations with CHX already placed in the core of the membrane. An external, constant electric field (0.125 V nm^{-1}) was applied across the EcIM to create a sizeable water-filled pore. This method was adapted from the technique applied by Piggot *et al.* to induce pore formation[99]. This was performed so that a CHX molecule could be manually placed within this pore, and the electric field was then turned off to determine whether CHX would remain in the membrane core as the pore closed. Throughout a 200 ns production run, CHX shifted from the membrane core to the exact location as observed in the equilibrium simulations of CHX presented above (Figure 4.19). This behaviour was observed across all three of the independent simulations of this system, providing conclusive support that the headgroup-lipid interface is the preferred location for CHX.

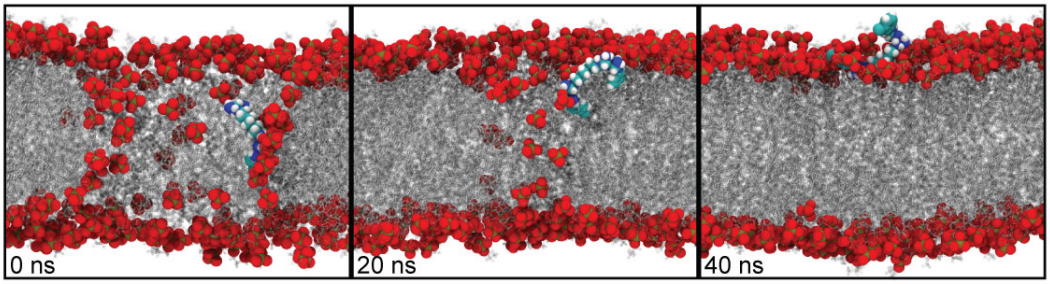


Figure 4.19: Snapshots of the CHX molecule placed in the membrane centre of a pore induced by a 0.15 V nm-1 field, which was turned off at $t = 0$ ns at (top left) 0 ns, (top centre) 10 ns, (top right) 20 ns, (bottom left) 30 ns, (bottom centre) 40 ns and (bottom right) 50 ns. Images show phosphorous in gold, oxygen in red, carbon in cyan, hydrogen in white, nitrogen in blue and lipids in translucent grey.

CHX is usually combined with gluconate (GLUC) in sanitising agents for increased solubility. To eliminate the possibility that GLUC is required to observe deep membrane penetration of CHX, we performed simulations including both molecules. 20 GLUC molecules were thus added to the bulk water (0.5% w/v) phase surrounding the EcIM. Following 100 ns of production simulations across three repeats, GLUC was not observed to penetrate the membrane and only briefly interacted with the surface ($<$ 1 ns events). Additional simulations with 10 CHX molecules were added to the same system and showed the same CHX membrane binding mode observed in simulations without GLUC. GLUC molecules were observed to bind to the CPL regions of CHX regularly, but this did not change the binding mode (Figure 4.20(a)). A

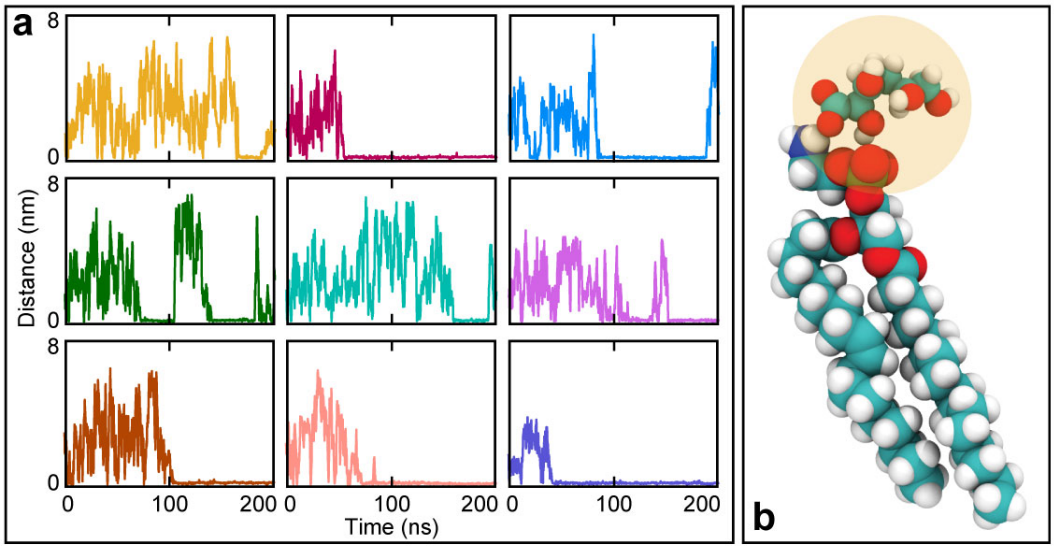


Figure 4.20: Analysis of the interaction of GLUC with the EcIM. (a) The minimum distance between 9 separate GLUC and all CHX molecules in a simulation system containing 20 GLUC and 10 CHX in 0.15 M KCl with an EcIM model. (b) Snapshots of GLUC interacting with the PE headgroups in the EcIM, where phosphorous is shown in gold, oxygen in red, carbon in cyan, hydrogen in white, nitrogen in blue and GLUC is highlighted by yellow circles

snapshot of a characteristic interaction between GLUC and PLs in the EcIM is shown in Figure

4.20(b).

4.4.4 Chlorhexidine in alcohol solutions: action on the *E. coli* outer membrane (Ra-LPS)

The action of the CHX/alcohol solutions on the EcOM was next studied in the same way as for the PL membranes. The EcOM was highly resistant to the deformation caused by alcohol in simulations of the PL membranes, despite the alcohol distributing within the membrane in the same way as for PL bilayers[178]. This is due to the strong interactions between LPS molecules and its stabilising calcium counter ions. Throughout the first 60 ns (200 ns simulations in triplicate), there was no discernible change in the membrane structure in any EcOM simulations. By this time, both PL membranes (SaCM and EcIM) had shown substantial deformation due to alcohol partitioning in the headgroup/tail interface. Due to the retained integrity of the EcOM, there was little change when analysing the bilayer density, other than a small shoulder of budding lipids (Figure 4.21). By 80 ns, the PL leaflet of the EcOM began to form a small bundle of approximately 50 lipids, which started to separate from the membrane (Figure 4.22). The budding phenomenon was observed in two of the three repeats with PROH, but none of the repeats performed in ISOP. PROH distributes into both leaflets' lipid headgroup/tail interface, whereas ISOP did not distribute into the interface as effectively because of its branched structure. Partitioning of alcohol into the membrane disperses the bilayer components due to crowding. This was alleviated in the PL membranes by lateral dispersal; however, due to the calcium binding LPS molecules, lateral movement was prevented here. This resulted in the PL leaflet budding into the periplasmic space to relieve the crowding, forming a shell of polar lipid headgroups sheltering their hydrophobic tails. As a small amount of both alcohols passed through the LPS leaflet regularly, this showed that deformation is possible with alcohol entering from only the extra-cellular leaflet, as expected physiologically. Given enough time to penetrate the membrane solely from the LPS leaflet

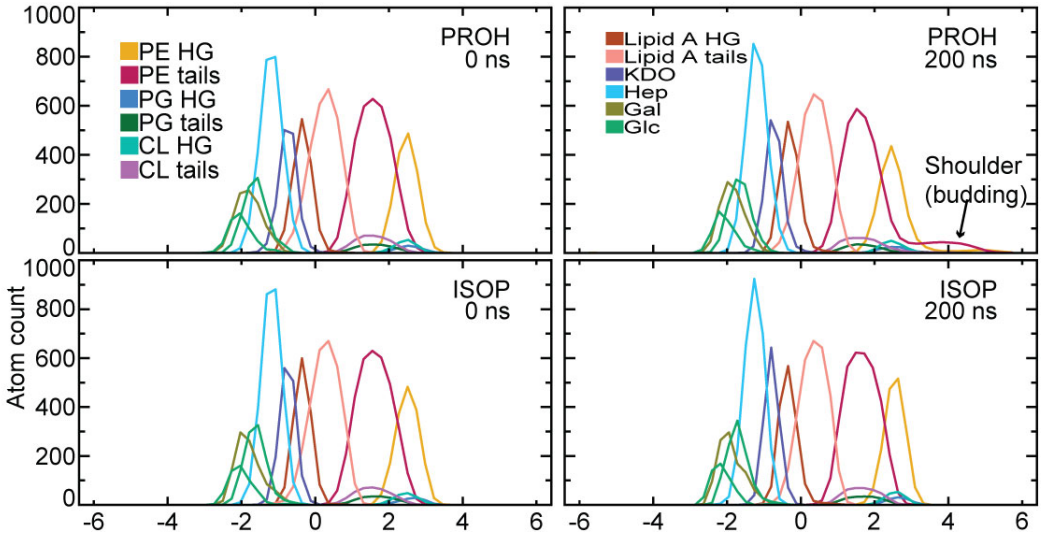


Figure 4.21: Density analysis along the z-axis of the EcOM system exposed to PROH and ISOP solutions. Analysed at 0-20 ns and 180-200 ns during a 200 ns production run, showing the density shoulder caused by the 'budding' effect.

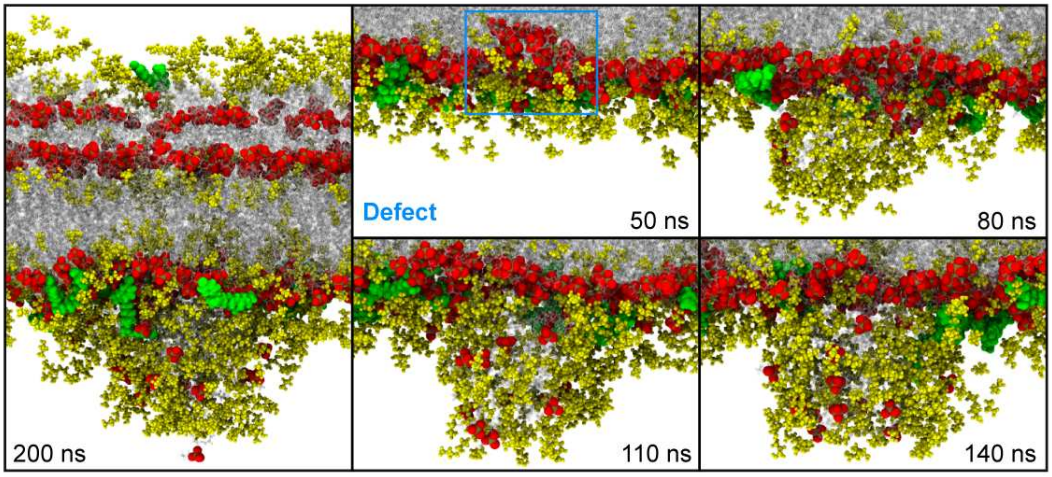


Figure 4.22: Snapshots of the EcOM exposed to 0.5% w/v CHX in 20% w/v PROH at 200 ns with further snapshots of the PL leaflet at 50, 80, 110, 140 and 200 ns showing phosphate oxygen as spheres in red, phosphate phosphorous in gold, CHX in green, PROH within 10 Å of the membrane in yellow and lipids as translucent.

side, the build-up of lateral pressure may result in budding to such a degree that it entirely deforms the EcOM. As commercial sanitisers contain more alcohol (70% w/v), this could be expected to occur faster than observed here.

Visual analysis of the production runs showed that only one CHX molecule ever became bound to the LPS leaflet in any of our simulations, and in 3 of our 6 simulations with alcohol and CHX present, none became bound to the LPS leaflet (2/3 PROH repeats and 1/3 ISOP repeat). The lack of binding to the LPS leaflet was due to the fewer favourable interactions between the HEX region of CHX and LPS sugars. Generally, CHX became bound to the Hep moiety in the LPS sugar region and did not translocate further (Figure 4.23).

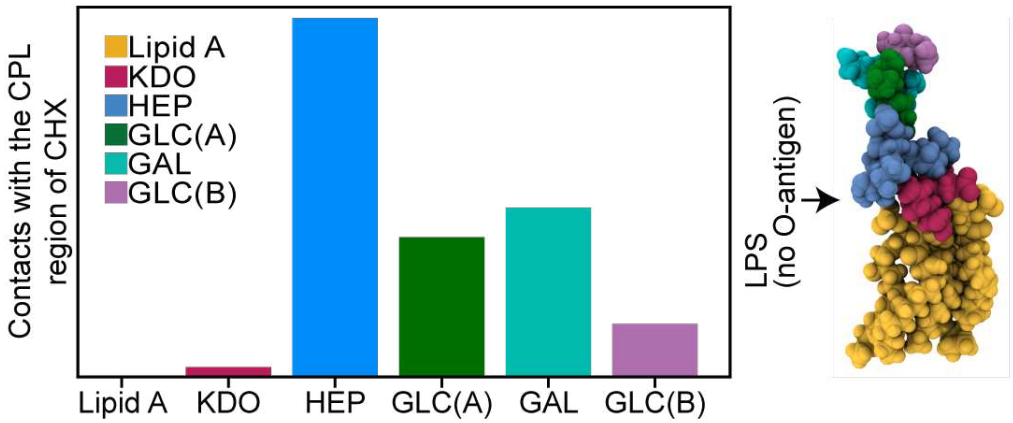


Figure 4.23: The proportions of contacts between the CPL region of CHX and sugars in LPS tracked over the full 200 ns of production by checking at each frame how many contacts there were within 2.5 Å of the CPL region and adding them to the total count per sugar, showing lipid A in yellow, Kdo in red, Hep in blue, Glc A in green, Gal in cyan and Glc B in pink. Glcs are labelled A and B at two different positions in the structure.

4.5 Conclusions and further work

The simulation studies here provide insights into the mode of action of the common sanitising agent CHX on *S. aureus* and *E. coli* membranes. We find that in the aqueous solution (no alcohol), CHX binds and inserts into the EcIM and SaCM, maintaining a position at the headgroup/tail interface of the former while moving slightly deeper into the latter. Intriguingly, the insertion of CHX does not induce disruption of the membranes during the timescales of the MD simulations. Nevertheless, CHX was found to reduce the MSDif of lipids as it binds and holds headgroups together, thereby acting as a 'molecular staple'. This result also supports a previously suggested mechanism.

No insertion of CHX into the LPS leaflet of the EcOM was observed in the MD simulations. This is likely due to limitations in timescale exacerbated by the slow-moving nature of LPS. It should be noted that a similar lack of penetration of other cationic antimicrobial agents, such as polymyxins, has previously been reported from atomistic MD simulations[179]. The comparative simulations show a more substantial kinetic barrier to CHX entry into the outer membrane than the PL membranes.

The inclusion of alcohol (PROH or ISOP) has a dramatic impact on the PL membranes. Both alcohols were seen in multiple simulation replicas to cause notable deformation of the EcIM and SaCM. Although alcohol deformation of PL membranes is not a novel finding, contact tracking has shown that deformation is the product of many transient residencies of alcohol in energetically beneficial positions within the bilayer. Large quantities of alcohol partitioning into this region resulted in lateral dispersion of lipids, which caused substantial deformation; at larger alcohol concentrations, this leads to the dissolution of the membrane. The same effect for PL membranes was not seen for the EcOM due to the lateral interactions between components of the LPS leaflet and calcium, which are more substantial than the PL membrane. Restrictions imposed by the LPS leaflet meant the bilayer could not disperse to alleviate steric crowding. This caused the PL leaflet to 'bud' into the solution, as this was the only way to reduce the crowding created by alcohol partitioning.

Although the work presented in this chapter is well-founded, it possesses limitations which must be addressed. These works' first and most obvious limitation is that the membrane models employ lipid bilayers with no proteins included. Although these simulations were explicitly designed to monitor the lipid bilayer component of bacterial membranes, not including proteins could limit their insight into bacterial membranes in vivo. Because proteins are a crucial component of bacterial membranes, lipid bilayers do not necessarily reflect the actual behaviour of bacterial membranes. These works aimed to characterise specifically the effect of lipids on membrane stability. With these caveats in mind, I have drawn conclusions that can be applied to more complex membranes. Considering that added complexity in vivo will also affect this, these findings contribute to understanding how sanitisers affect membrane stability.

The lipid bilayers in these simulations comprise a relatively simple mixture of lipids; as with the decision to not include proteins in these systems, this results in a lack of complexity. The lipids contained in a bacterial membrane are diverse, and not accounting for this diversity may limit

our insight into the dynamics of bacterial membranes. Although the membrane compositions I have applied here contain relatively simple lipid ratios, this is beneficial for two reasons. Firstly, this allowed me to focus on comparing my results to similar studies with these lipid ratios. Secondly, it makes it relatively simple to reproduce these membranes in vitro in the following section (Section 5) for direct comparison. The benefit of this is that it made the systems more straightforward to produce in vitro, and for experiments such as 2D NOESY NMR, this made the assignment and, by extension, resolving cross-peak relaxations far more reliable.

Further work in this area should focus on improving the model of these membranes so that we can make direct comparisons to bacterial membranes with more complexity. By furthering the complexity of the bacterial model membranes here, it could be assessed how including aspects such as proteins or more variation in lipid composition changes the process of membrane deformation. Making these membranes more elaborate could provide insight into how these features alter the mechanisms by which sanitisers deform bacterial membranes. Other suggestions for future work are primarily centred around techniques that could be used to enhance sampling or the production timeframe. CHX was seen not to affect the EcOM, and PROH may further deform the bilayer given more time. The EcOM was only seen to deform in half of the alcohol-based systems, implying that these deformations had not reached an endpoint. By exploring enhanced sampling or techniques that could dramatically extend the production timeframe, events beyond the timeframe of these simulations could be reported.

Chapter 5

NMR analysis of the sanitiser application to a model bilayer of the *E. coli* inner membrane

The work presented in this chapter is published in the Journal of Molecular Biology[126].

A note on the effect of the SARS-CoV-2 pandemic

This work was performed between October 2019 and October 2021. Unfortunately, this coincided with the SARS-CoV-2 pandemic and the restrictions on in-person interactions. As this work was experimental and relied upon lab preparation and experimentation, the scope of this work was limited by what we could do outside of restrictions.

5.1 Abstract

It is well known at this point that the sanitising agents we currently rely upon are effective. However, we must consider how we would engineer new ones if the ones we currently rely upon were to become ineffective. For these antimicrobials to work effectively as preventative agents, we must thoroughly understand how they bind to the envelopes of pathogens and trigger their bactericidal mechanisms. These antimicrobials' specificity makes them critically important in fighting off harmful bacteria. A key feature in their effectiveness is generally membrane charge, a property of some bacterial membranes that is exploited in drug design. To understand the effect of this binding mechanism, we need first to know the antimicrobial's exact locality and binding conformation.

I utilise various NMR techniques in 1D and 2D in this thesis. Here ^{1}H NMR is used to report on lipid structure and interactions of sanitiser molecules while ^{31}P NMR provides insight into the bilayer's dynamics and integrity. This allows me to determine the effect of these sanitisers on scales outside of the scope and timescale of MD simulations. The works performed here provide validation for those discussed in Chapter 4, while the work in the previous chapter

also provides visual context for the interactions described here. In knowing how sanitisers or therapeutics bind, scientists know where to look for the responsible mutations if a bacterial mutation renders them ineffective. These works provide an exploratory approach to sanitiser binding in the hope that it may better the engineering of small-molecule sanitisers in the future.

5.2 Introduction

5.2.1 NMR as a means of assessing liposome dynamics

NMR is a powerful and versatile biophysical technique that reports on the local structure and environment of molecules at an atomistic resolution. As mentioned in the introduction to this thesis, this technique is often used to study intricate properties of biological systems, both in vivo and vitro[180, 181]; SS NMR is often used to study bacterial and lipid vesicles. We must understand these systems well, as they are crucial to important research areas such as ecology, medicine and industrial applications. Various NMR techniques are used regularly in researching bacterial and lipid vesicles[113, 182].

A prominent technique used to study lipid vesicles in vitro is static ^{31}P NMR. This technique exploits the magnetic properties of ^{31}P to assess the dynamics of lipid vesicles (Section 2.2.7); ^{31}P is abundant in the headgroups of lipids. Static ^{31}P NMR provides researchers with a facile means of assessing the integrity and morphology of lipid vesicles in real time[183, 134]. Lipid vesicles are the basis of bacterial cells and viral particles, so this research is also invaluable in furthering our knowledge of these areas. Static ^{31}P NMR is commonly used to assess antimicrobial agents' effect on lipid vesicles. By creating liposomes that mimic those of bacterial cells or viral particles and applying sanitisers, the changes in the lineshape can be used to infer the dynamics of the liposome[184]. The change in dynamics of the liposome can be monitored and used to infer the effect of the sanitiser.

An important property of ^{31}P NMR is the chemical shielding anisotropy (CSA). The CSA is the variation in the magnetic shielding effect experienced by a nucleus within a molecule due to its local electronic environment. This effect is not uniform in all directions, depending on the molecule's orientation relative to the applied magnetic field. When the electron cloud surrounding the nucleus is distributed anisotropically, it leads to variations in the shielding effect, resulting in different NMR resonant frequencies, depending on the orientation in the applied magnetic field. CSA is a crucial factor in understanding the structural and dynamic properties of lipid vesicles; this is especially well suited to solid-state systems as it is performed statically. Whereas MAS is often applied in ^{1}H NMR, ^{31}P NMR relies upon the additive signal of many anisotropic environments. As ^{31}P is an anisotropic nucleus, the orientation of these in lipid bilayers contributes to the overall ^{31}P lineshape; thus, changes in the vesicle are evident from the lineshape.

1D proton NMR is a commonly applied technique. In the context of liposomes and their interactions with small molecules, 1D NMR provides a method of assigning environments to peaks in the spectra. This is often used to determine the structure of products in synthetic chemistry; however, here, it is applied to ease the assignment of 2D NOESY NMR peaks.

2D NMR has revolutionised the study of complex molecular systems by providing detailed insight into their structural organisation, dynamics, and interactions at an atomistic level of detail. This method has been used commonly to investigate the interactions between liposomes and small molecules of antimicrobials. 2D NMR provides further insight into 1D NMR as cross-peaks in 2D spectra can be used to infer spatial or bonded relationships between environments in the 1D spectra. In NOESY NMR, cross-peak intensity provides insight into spatially common interactions between different environments, unlike other methods that report the bonded relation between peaks. As the intensity of cross-peaks is NOESY NMR, which is directly proportional to the rate of magnetisation exchange between environments, it can be used to determine the locality of chemicals in the system[185]. The volume of cross-peaks in NOESY NMR can be integrated and, as described in Section 2.2.8, used to calculate the magnetisation exchange rate. Due to the level of detail the NOE effect gives us, 2D NOESY NMR is a technique well suited to studying liposome-based systems. This technique provides well-resolved peaks for lipid-lipid, lipid-sanitiser and sanitiser-sanitiser interactions in complex systems.

5.2.2 Studying small molecule binding and partitioning via NMR

NMR is a widely used technique in pharmaceutical studies due to its ability to report on molecular interactions with intricate detail. It is instrumental in binding and partitioning studies, which are crucial in developing effective antimicrobials. Recent research by Doyen *et al.* explored the application of NMR to peptide-liposome formulations, using 31P, 1D, and 2D proton NMR[186]. In this publication they comment on the dynamics of peptide-liposome interactions using the 2D NOESY NMR spectra produced through relaxation measurements. Similarly, Scheidt *et al.* used NMR to study the positions within lipid bilayers where aromatic drugs partition[187]. By monitoring cross-peak decay in 2D NOESY NMR, they were able to draw detailed insight into the subtle nuances of binding and partitioning mechanisms.

CHX is a widely used antiseptic and disinfectant. As mentioned previously, this antimicrobial is theorised to bind at the headgroup region of membranes via the CPL and BGU functional groups as a "wedge", parting the headgroups and creating gaps that make the membrane "leaky"[93]. It has been demonstrated via cell culture experiments and verified by Franz diffusion cell studies that, unlike other antiseptics, CHX can remain active due to epidermis binding even after the bulk of the sanitiser solution has been washed away[95]. In similarly themed work, performed by Karpanen *et al.*, they find that as long as 30 minutes after skin exposure to CHX, it can remain present in concentrations high enough to stay bactericidal[96]. This is a crucial finding as it presents a further benefit to using CHX. It not only acts as a potent bactericidal agent with high efficacy but remains present and continues to prevent potential infection long after use. This is likely due to the same binding as mentioned by Van Oosten in epithelial membranes. As such, it is noted by Karpanen that the presence of fatty acids may reduce said efficacy.

It has been shown by Feller *et al.* that short-chain alcohols such as ethanol will distribute themselves into specific regions of a lipid bilayer. In their work, they use cross-relaxation rates

(similar to techniques utilised by Doyen and Scheidt) to determine the location of ethanol inside a lipid membrane. This showed that ethanol molecules exchange magnetisation more rapidly with the headgroup-tail interface; this suggests that ethanol partitions into this region. The findings of Feller *et al.* shed light on the complex interactions between lipids and small molecules and exemplify the importance of NMR in future studies in this area. Further, the work of Huster *et al.* details similar spin diffusion interactions between lipid headgroups and their tails, demonstrating further use of NOESY NMR to study bilayers and their dynamics. The area of research has been further explored by Gawrisch *et al.* and Feller *et al.*, who provide an extensive background on methods to study these systems[188, 123]. In these works, a plethora of techniques are demonstrated as a means of extracting information about biomembranes as well as detailed descriptions of the interpretation of these techniques. These works provide detailed methods upon which the work reported here was based.

5.2.3 Lipid model of the *E. coli* inner membrane

The lipid composition of the EcIM is essential to the structure and fluidity of the bilayer, as mentioned in Section 1.4. In the introduction, it is stated that the generally accepted lipid ratio of the EcIM is approximately 75:20:5 PE:PG:DPG[81]; however, here, a lipid ratio of 90:5:5 PE:PG:DPG is used. This lipid ratio was used instead as it maintains PE and PG as the main contributors to the lipid bilayer while directly allowing us to compare the previous simulation section. This ratio was used in the simulation counterpart section to this work as it provides a reliable lipid ratio, which has been studied previously. This provided me with a lipid ratio that had already been validated and was reproducible[99].

5.3 Materials and methods

5.3.1 Construction

Lipid samples were prepared from a PE, PG and DPG stock solution in a ratio of 90:5:5, respectively. Stocks were placed in ethanol at a lipid concentration of 10 mg mL^{-1}. Samples of 5mg lipid mixtures were left under a high vacuum to remove residual solvent. The lipids were rehydrated in 20 μL of D_2O (25% w/v) containing alcohol or CHX at the appropriate concentration. Samples containing 5 mg, 25% w/v lipid samples were freeze-thawed five times. Each cycle involved placing the sample in liquid nitrogen, thawing and mixing with a vortex mixer. These samples were then transferred to 3.2 mm rotors for analysis. CHG was applied to the membrane at a concentration of 4 % w/v. Although the percentage volume of PROH applied was significantly larger than that of CHG, due to the size of CHG, the molar ratios of PROH and CHG to lipids are similar (1:4.67 and 4.12, respectively).

5.3.2 NMR parameters

Solid-state NMR spectra were recorded on a 600 MHz Agilent DD2 NMR spectrometer (Yarnton, UK) equipped with a 3.2 mm triple resonance MAS NMR probe. All spectra were acquired at 25°C unless specified. Static ^{31}P spectra were recorded with a Hahn echo pulse

sequence[189] with a 3.5 μs excitation pulse, 70kHz proton decoupling during acquisition, and a 50 μs echo time. A 2.5 s recycle delay was employed to minimise sample heating. All ^{31}P spectra were externally referenced to H_3PO_4 (85%). Before a fast Fourier transformation, data was left shifted to the top of the echo, zero-filled to 2048 points and 100Hz line broadening was applied. All proton spectra were recorded with 12.5 kHz MAS, a 2.75 μs pulse for excitation, and a 2.5s recycle delay. All proton spectra were acquired with the probe with the low gamma channel tuned to deuterium and connected to the spectrometer's lock, stabilising the field and significantly reducing t_1 noise in 2D spectra. Proton spectra were referenced to the residual water peak at 4.65 ppm. 2D ^{1}H ^{1}H MAS-NOESY spectra were recorded using a standard exchange sequence[190] with States-TPPI in the indirect dimension and 256 t_1 increments[191]. 2D NMR spectra at mixing times of 31.25, 62.5, 125 and 250 ms were acquired for cross-relaxation calculations. Data was processed with NMRpipe[192] or Matlab[193] using matNMR[194].

5.3.3 Cross-relaxation analysis

Cross-relaxation analysis was performed by integrating the cross-peak volume, representing these integrations in matrix format and then converting these values to magnetisation exchange rate. Integration of these peaks was performed using NMRpipe[192]. Conversion from cross-peak volume to the magnetisation exchange rate was performed using a matrix-based version of the Solomon equations[125]. This process was performed as described in Section 2.2.8. The mathematics of this process were performed in Matlab[193]. Peaks were assigned using spectra processed using matNMR[194].

5.4 Results and discussion

5.4.1 Static ^{31}P studies of the effect of increasing chlorhexidine and propanol concentrations on *E. coli* inner membrane model vesicle dynamics

The action of sanitiser components on the structure of the EcIM model was assessed experimentally using ^{31}P solid-state NMR. This method produces a lineshape characteristic of the dynamics and phase of a lipid vesicle. Spectra were acquired of multilamellar vesicles (MLV) composed of PE:PG:CL (90:5:5 mol %) exposed to PROH, GLUC and CHG to assess their influence on the bilayer structure. CHG is CHX with two GLUC counter ions; this is used in experiments as it dramatically increases the solubility of CHX with no effect on its ability to act as a sanitising agent. CHX is also generally added to sanitiser solutions as CHG for the same reason[195].

As expected in the absence of any reagents, the ^{31}P spectrum of the model EcIM exhibited an axially symmetrical lineshape with a CSA of 43 ppm, consistent with the lipids existing in the liquid crystalline phase. The method used for all experiments was consistent; thus, any change in lineshape must directly result from adding sanitisers. Increasing the concentration of PROH in the sample resulted in significant perturbations of the bilayer structure (Figure 5.1(b-f)). Up to 40 %, the lipids retained their bilayer character, with a slight reduction in the

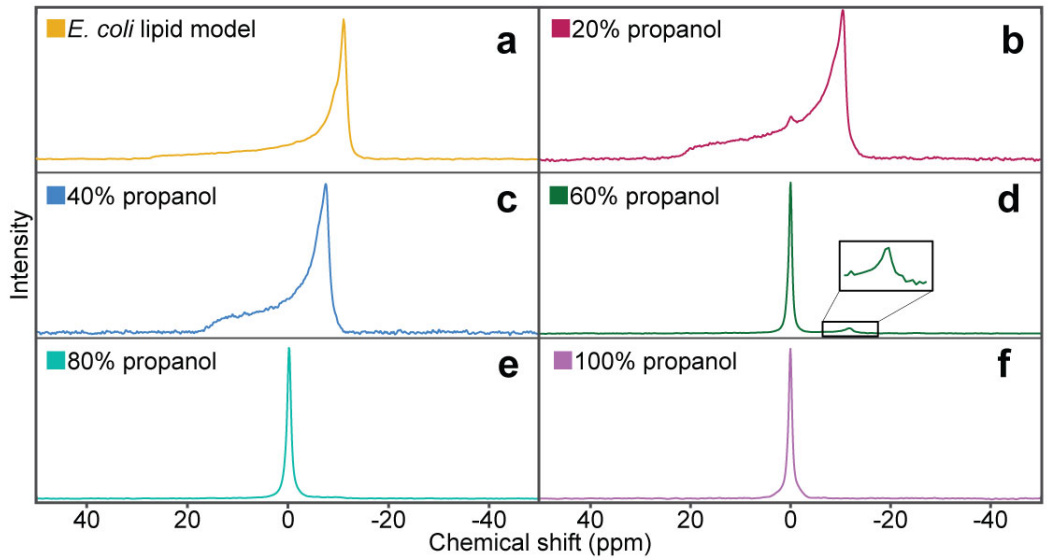


Figure 5.1: Static ^{31}P NMR spectra of the EcIM model alone, with PROH at (b-f) 20% w/v intervals

CSA indicative of increased headgroup mobility. This was consistent with the partitioning of the PROH at the top of the lipid chains, resulting in a lower lateral pressure in the headgroup region. This allowed the vesicle shape to fluctuate more quickly, resulting in changes to the observed change in lineshape as faster dynamics lead to more averaging of the CSA. Between 40% and 60% w/v PROH, most of the bilayer structures were solubilised with the spectra dominated by an isotropic signal at -0.1 ppm, with only a minor bilayer-like component still present at 60% PROH. This isotropic signal indicates that the lipids present in the vesicle bilayer are fully dissolved in the solution. This results in an isotropic environment as, once dissolved, all lipids exist in equivalent environments; hence, they all contribute to one sharp peak around 0 ppm.

Static ^{31}P NMR experiments were performed to obtain characteristic line shapes of systems containing between 0 and 4 % w/v CHG at 1 % intervals (Figure 5.2(a-e)). Concentrations up to 4 % w/v were used as this is the highest concentration commonly used in commercial sanitisers[196]. Increasing concentrations of CHG had limited influence on the ^{31}P lineshape, with all spectra showing a characteristic axially symmetric lineshape indicative of a bilayer-like phase. Notably, there was a slight increase in the CSA as the concentration increased, suggesting that the CHG may limit headgroup mobility. This is directly opposed to PROP's effect on the vesicle dynamics. These findings suggested that the decreased lipid mobility observed upon adding CHX in Chapter 4 represents the same binding as reported in experiments performed in vitro. Increases in the intensity on the downfield edge and overall broadening of the spectral features were similar to those observed in the gel phase. These are typically attributed to a reduction in T_2 of the phosphates in the headgroup. This suggests that the bilayer became more rigid and less elastic due to CHX holding lipid headgroups together like a molecular staple[197, 183, 198, 199, 200].

Similar observations were seen upon the addition of GLUC with the spectra (Figure 5.2(g))

exhibiting similar broadening due to a reduction in T_2. However, no changes in the relative intensity across the lineshape were observed, suggesting that GLUC had minimal influence on bilayer elasticity. These effects appeared to be controlled by CHX, as the corresponding spectrum of 4 % GLUC exhibits a classical axially symmetric lineshape expected from MLVs.

5.4.2 1D ^{1}H NMR assignment

To investigate the location of the CHG within the model EcIM bilayer, 1D ^{1}H-MAS and 2D ^{1}H/^{1}H MAS-NOESY experiments were recorded. Due to the high mobility within the lipid bilayer, the ^{1}H spectra were well resolved, as previously reported[123, 133]. High mobility within the lipid bilayer results in an averaging of anisotropic interactions. This decreases the line broadening, increases spectral intensities and generally results in better-resolved spectra.

1D ^{1}H-MAS experiments were performed on these systems first. Before assigning peaks that arose due to the addition of sanitiser, those belonging to lipids making up the bilayer were first assigned. The ^{1}H spectrum of the model EcIM was dominated by resonances from PE, the major lipid species in this system. Although this made studying PG and DPG significantly harder, it allowed us to assign and analyse PE quickly. The resonances from the lipids were assigned based on published assignments of PC and PE (Figure 5.3)[133]. Similarly to Section 4, PROH was added as 20% to avoid destroying the bilayer. This also allowed me to make assignments with PROH partitioned into the bilayer, making processing 2D data significantly easier later on.

The addition of 20% PROH gave rise to three significant resonances at 3.61, 1.61 and 0.96 ppm, which were assigned based upon published values (Figure 5.4(a))[177]. Expectedly, these peaks dwarfed the lipid as even at low concentrations of PROH, the PROH environments were significantly more numerous than those of the lipid. Adding 4% w/v GLUC gave rise to a large group of resonances superimposed on the lipid signals assigned to G1, G3, α and β

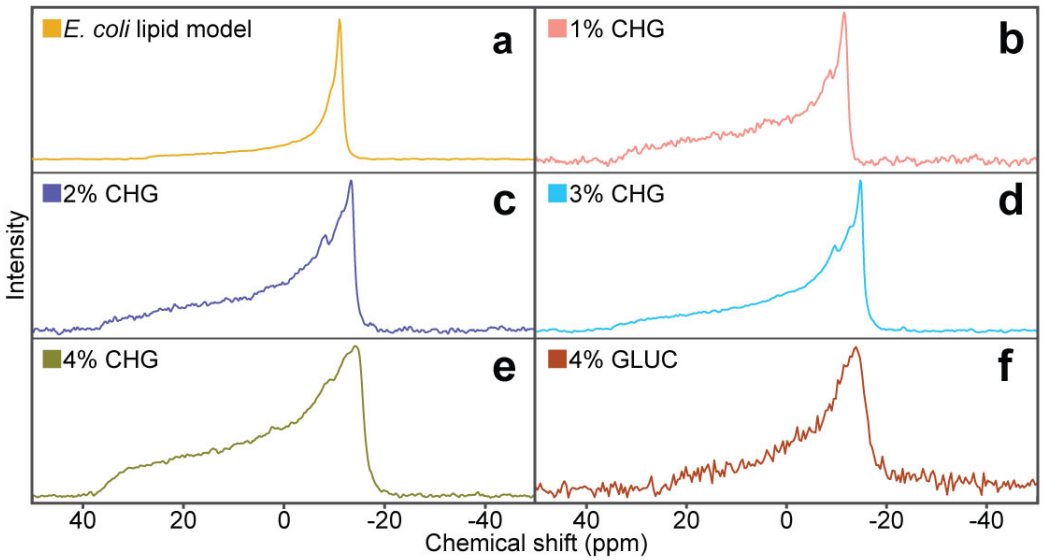


Figure 5.2: Static ^{31}P NMR spectra of the EcIM model alone, with CHG at (b-e) 1% w/v intervals and (f) 4 % w/v GLUC

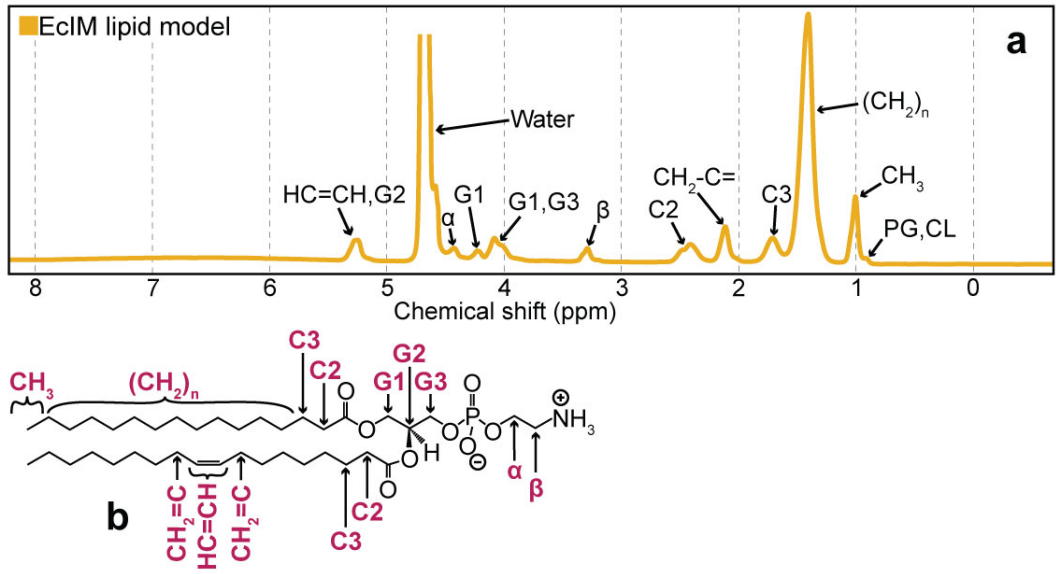


Figure 5.3: Assignment of proton environments in PE to peaks in the ^{1}H NMR spectra of the model EcIM vesicle system in water. (a) Assignment of peaks in the ^{1}H NMR spectra. (b) The skeletal structure of PE with proton environments labelled in red.

(Figure 5.4(b)). This was problematic because GLUC resonances may make resolving others challenging.

The addition of CHG resulted in the appearance of a further six resonances assigned to CHX. The two protons in the aromatic ring structures of CHX were significantly better resolved than those in the rest of the structure (Figure 5.5(a) - proton assignments 1 and 2). Protons in the CPL functional groups are significantly more deshielded due to strongly electronegative functional groups within the CPL group; chlorine and aromatics are strongly deshielding. This

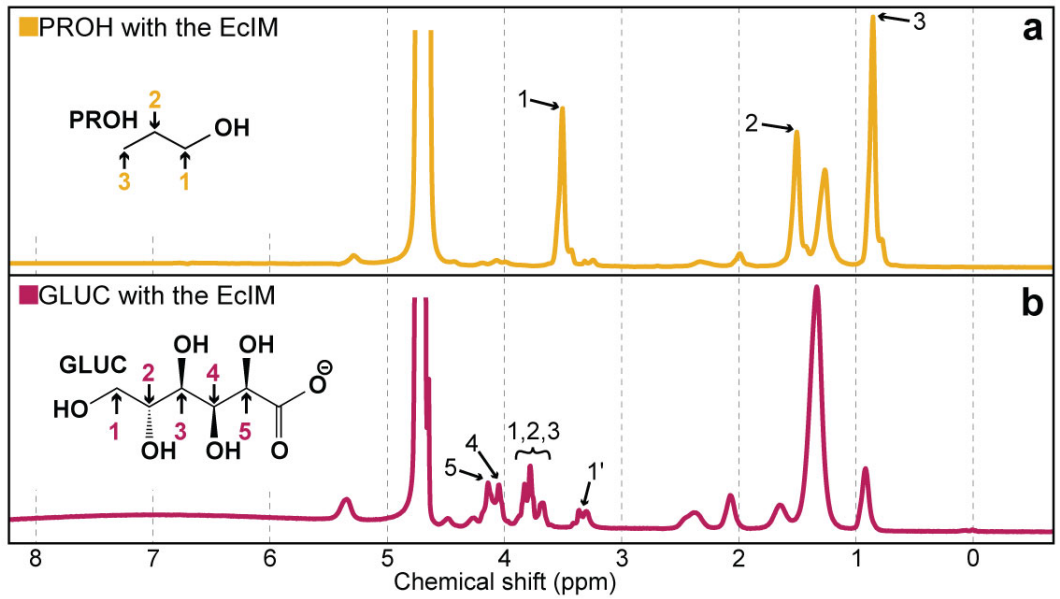


Figure 5.4: Assignment of proton environments in PROH to peaks in the ^{1}H NMR spectra of the model EcIM vesicle system in aqueous PROH.

is furthered by generating a ring current in the CPL group's aromatic ring, which generates its magnetic field and further deshields nearby protons. CHX protons in the BGU/HEX region are significantly better shielded by electron density due to the relative absence of electronegative functional groups (Figure 5.5(a) - proton assignments 3, 4 and 5). The shift of these protons moves further upfield from the BGU functional group to the HEX group. Close inspection of the CPL region of the spectrum (Figure 5.5(b)) revealed that CHX exhibited three resonances in the aromatic region. This was unexpected, given the presence of only two resonances for the aromatic protons in the CPL groups in CHX when CHX was examined without GLUC. The aromatic protons are expected to be magnetically equivalent. I suggest that the additional resonance arose due to a stable complex of CHX and GLUC, formed by the interaction of the CPL region of CHX with the carboxylic acid group of GLUC (Figure 5.5(c)); this would break the magnetic equivalence. This was supported by the complexes seen in MD simulations of CHX with GLUC (Figure 5.5(d)). Closer inspection showed that resonances from the C2, C3 and G2 sites within the lipid shifted upon adding CHG. It is plausible that the presence of the CPL groups at the headgroup-tail boundary may have generated ring-current effects resulting in the observed perturbations like that reported for ibuprofen by Kremkow *et al.*[125] When an aromatic ring is exposed to a magnetic field, such as that in an NMR experiment, the applied field will induce the movement of electrons in the ring. This cyclic movement will then, in turn, generate its magnetic field, which can increase or decrease the effective magnetic field experienced by nearby protons. To make peaks that are shifted due to ring currents in the CPL group obvious, an underlay of the EcIM 1D ^{1}H spectrum is shown in yellow in Figure 5.5(a).

5.4.3 1D ^{1}H NMR T_1 relaxation

To assess the effect that sanitisers had on the mixing of proton environments in the EcIM model bilayer, T_1 relaxation analysis was performed on the peaks presented for PE in the EcIM, shown in Figure 5.3. Figure 5.6 shows that for the α, β, G1/G3 and β environments, the T_1 relaxation time was substantially lower when either CHX or PROH were added to the lipid vesicles. This same effect can also be seen in the C2, CH_2=C and $(CH_2)_n$ environments; however, it is less substantial. This suggests these environments were more mobile, especially the α, G1 and β environments. This creates the largest decrease of T_1 in experiments that contained PROH due to the same membrane thinning effects present in Chapter 4. In some proton environments (α, G1, β), T_1 decreases less when CHX is added. The environments that show this are also generally in the headgroup. This lesser decrease in T_1 is due to the same headgroup binding of CHX presented in Chapter 4. The T_1 measurement of the CH_3 environment remains the same in the presence of PROH. This suggests that the tail ends are packing together in the membrane centre as PROH does not penetrate the bilayer to this depth. Interestingly, the relaxation rate of CH_3 does change in the presence of CHX, indicating there was potentially some significant penetration or even CHX bridging, as described in Chapter 4. These findings suggest that PROH partitions in the bilayer and increases lipid mobility in the region it partitions toward the lipid headgroup-tail interface. CHX is theorised to bind at the lipid headgroups, and in systems where both sanitisers are applied, it counteracts the effect of PROH on lipid mobility. The data presented here is shown in Supplementary Table 9.1.

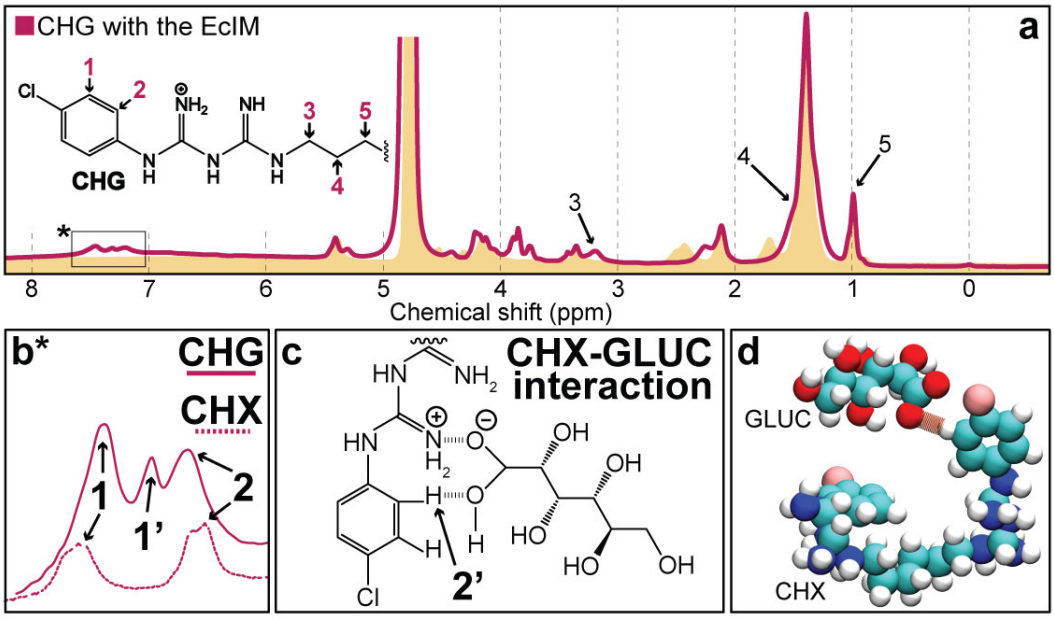


Figure 5.5: CHX dynamics and conformation analysis when added to a model EcIM vesicle system. (a) Assigning proton environments in CHX to peaks in the ^{1}H NMR spectra of the model EcIM vesicle system in aqueous CHG. (b) Resonances arising from CHX and CHG. (c) A visual representation of a common interaction between CHX and GLUC in an MD simulation of 10 CHX and 20 neutralising GLUC. Atoms are shown as spheres: hydrogen in white, carbon in cyan, nitrogen in blue, chlorine in pink, oxygen in red and hydrogen bonding in brown. (d) A mechanistic depiction of the interaction between CHX and GLUC is theorised to account for the additional resonance.

5.4.4 ^{1}H-^{1}H NOESY NMR studies of the locality and binding conformations of chlorhexidine in model *E. coli* inner membrane vesicles

To establish the location of the CHX within the membrane, 2D ^{1}H/^{1}H NOESY MAS-NMR experiments were performed. The dynamics present in the membrane give rise to the well-resolved ^{1}H spectra at moderate MAS frequencies, allowing the analysis of NOE-based magnetisation exchange. The cross-peak intensity in the 2D ^{1}H/^{1}H NOESY MAS-NMR reflects the residency time and proximity of protons relative to one another[124]. As shown in previous research[201, 202], an analysis of the inter-molecular cross-relaxation rates between resonances arising from the small molecule and the lipids can provide a valuable insight into the partitioning and location of small molecules within the lipid bilayer.

This experiment was first performed with the EcIM model vesicles and no added sanitiser. This yielded 2D spectra with cross-peak intensities reflecting how lipids in the bilayer interact. Figure 5.7 shows significant interactions between HC=CH, G2 and $(CH_2)_n$ resulting from the lipid tails interacting in close proximity. Similar interactions can be seen with other lipid regions in the bilayer centre. It should be noted, as seen in Figure 5.7, that the volume of cross-peaks also relies on the number of protons in that environment. For instance, although there is a relatively large cross-peak between $(CH_2)_n$ and β it is comparatively smaller than other cross-peaks between regions in the lipid tails. Interaction between these two regions is not necessarily unexpected; however, it appears far more prevalent as the $(CH_2)_n$ protons are

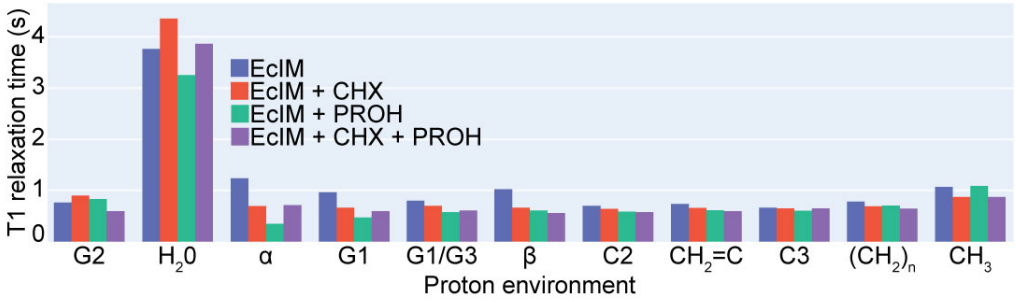


Figure 5.6: The T_1 measurement of different proton environments in the structure of PE in the EcIM systems in water and exposed to sanitisers.

much more numerous. Peaks along the diagonal will possess a larger volume if there are more protons in that environment, as there are more contributing to the same signal and as such, the intensity is much higher. If this is not considered, it can significantly skew our interpretation

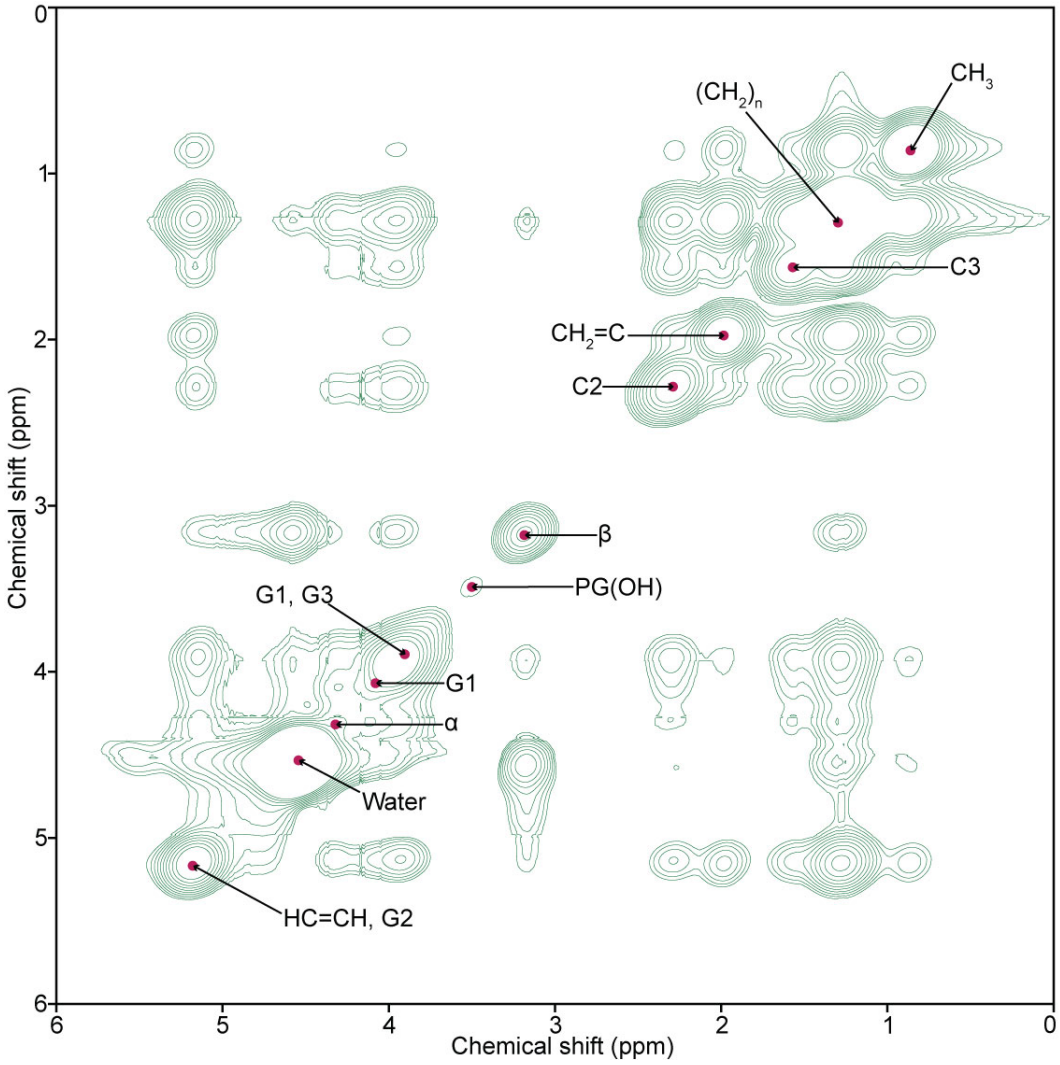


Figure 5.7: The 2D NOESY NMR spectrum of the EcIM model in water with a 250 ms mixing time. Positive contours are shown in green, and peaks are indicated by a mark in red.

of the results.

There were significant intra-molecular cross-peaks between the resonances arising from the CHG, most clearly defined with sites in the CPL rings and less clearly between the HEX resonances poorly resolved from lipid resonances. Other inter-molecular cross-peaks were also apparent; these were most clearly observed between the CPL protons and the intense signal from the CH_2 moieties in the lipid chains. Lower intensity peaks were also evident between the CPL protons and those from the protons at C2 and the glycerol backbone. This data again indicates that CHG has a preferred localisation at the level of the glycerol backbone in the lipid bilayer. Weak correlations were also seen between the C3 and the lipid resonances, indicating that the backbone of the CHG was lying in the plane of the membrane; this finding is in agreement with the MD simulations discussed in Chapter 4.

To ascertain the relative orientation of CHG within the EcIM, a quantitative analysis of the

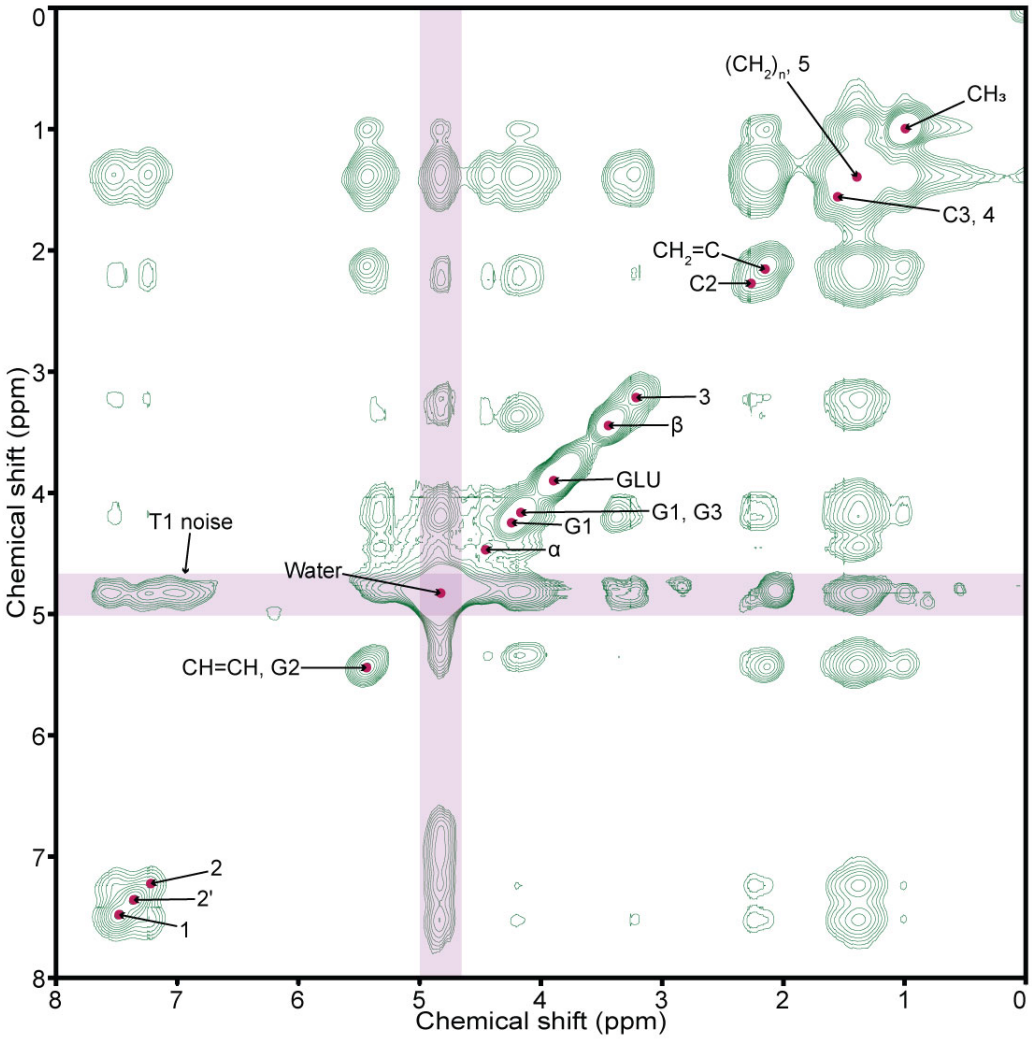


Figure 5.8: The 2D NOESY NMR spectrum of the EcIM model exposed to CHG at a concentration of 4 % w/v with a 250 ms mixing time. Positive contours are shown in green, and peaks are indicated by a mark in red. Areas that are distorted due to t_1 noise are highlighted in pink.

cross-relaxation decay rates, shown in Figure 5.9, was performed. Interestingly, the cross-relaxation rates suggest that resonances 1 and 2 of CHX interacted favourably with sites spanning the region from the unsaturated site within the lipid chains (HC=CH) up to the part of the lipid backbone. It should be noted that the calculation of the magnetisation exchange is standardised per proton. Because of this, high intensity peaks (such as that of $(CH_2)_n$) will not skew the results. In contrast, cross-relaxation of resonance three was dominated by exchange to the α proton of the PE headgroup, suggesting that the BGU/HEX groups of CHG prefer to localise within the region of the lipid headgroups, in keeping with the membrane-bound position observed during simulation.

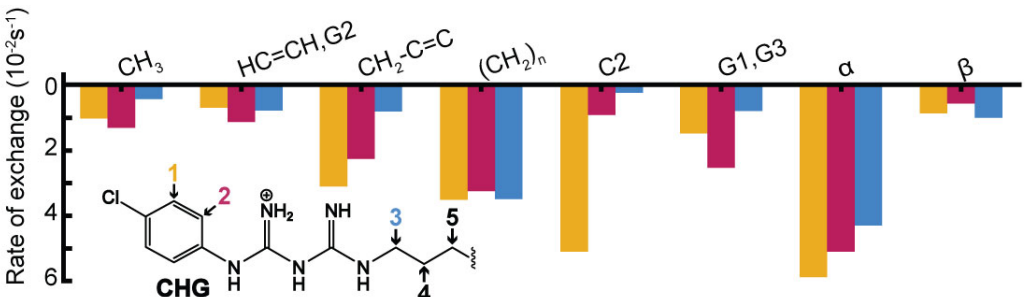


Figure 5.9: A bar chart of the magnetisation exchange rate of protons 1, 2 and 3 with protons in lipids in the EcIM lipid bilayer model, ordered from left to right to represent groups ascending from the bilayer centre toward the top of the lipid headgroups.

5.5 Conclusions and further work

Static ^{31}P NMR studies of the EcIM model vesicles revealed significant changes in the bilayer structure in response to the addition of sanitisers. Increasing the concentration of PROH led to a solubilisation of the bilayer, resulting in a large isotropic environment appearing. The solubilisation of the bilayer took place at approximately 60%. CHX had limited influence on the ^{31}P lineshape, suggesting it did not substantially deform the vesicle structure, and by all accounts, the vesicles remained intact. However, CHX did induce changes in the bilayer dynamics, making it more rigid and less elastic. A decrease in T_2 visible from minor changes in the lineshape in ^{31}P experiments suggests that CHX binds to the lipid bilayer and reduces lipid mobility. The findings presented in this section corroborate the findings of the simulations performed in Chapter 4.

1D ^{1}H NMR assignments provided a facile means of assigning peaks to proton environments before 2D experiments, where this information would be invaluable. This also allows me to assign the PROH, GLUC and CHX peaks. CHG resulted in six additional resonances attributed to CHX, with evidence of potential CHX-GLUC complexes signified by an extra resonance, which was only present when CHX and GLUC were present. Ring currents generated in the rings of CPL in CHX perturbed lipid environments, particularly at the headgroup-tail boundary, indicating its binding effects. This suggests binding of the CPL functional groups just below the lipid headgroups.

T_1 relaxation analysis revealed that PROH and CHX influenced lipid mobility. PROH increased lipid mobility, especially near the headgroup-tail interface, while CHX bound to lipid headgroups and counteracted PROH's effects, leading to decreased lipid mobility in the headgroup region.

Finally, 2D ^{1}H/^{1}H NOESY MAS-NMR experiments shed light on the localisation and binding conformations of CHX in the model EcIM vesicles. The data suggested that CHX predominantly localised in the glycerol backbone region of the lipid bilayer, with specific interactions between CHX and protein in the interface region between the lipid headgroups and tails. Cross-relaxation rates suggest a binding mode of CHX, which is very similar to that described in Chapter 4.

A noticeable limitation of the works performed here is that these vesicles are idealised systems containing only simple lipid bilayers, which only partially represent a bacterial vesicle. The experiments potentially miss nuanced information about how these structures may affect the binding process by not including protein or other biological structures associated with the bacterial membrane. This work was performed in the context of comparing work directly to equivalent MD simulations, which has been done successfully. Although this simplification results in a loss of some detail, it has elucidated how sanitiser molecules bind in bacterial lipid vesicles. A logical extension of the work performed here would be to include membrane proteins in the lipid vesicle. This was an area of interest during the experimental phase; however, unfortunately, this was not explored due to the effect of the SARS-CoV-2 pandemic on in-person work. 2D NOESY NMR studies, similar to those performed here, are also standard practice with protein structures. Thus, further analysis of the binding conformations and

location of CHX could have been performed.

Further, fluorescence experiments were intended to be performed with EcIM model vesicle systems and gradually increasing CHX concentrations. This would have allowed me to track the concentration at which CHX definitively, if at all, began to cause the vesicle to break. This could have been explored with a range of salt concentrations to explore the proposed effect of CHX on bacterial membranes, that it ruptures bacterial membranes by creating a charge gradient across the bilayer. This would have allowed me to round out this work and relate the findings regarding sanitiser binding and partitioning to the effect of CHX directly on bacterial vesicles.

These NMR studies provided valuable insights into the interactions between sanitisers and lipid bilayers, highlighting the effects of PROH, GLUC, and CHX on the EcIM model vesicles. These findings contribute to understanding sanitiser-membrane interactions and their potential implications for sanitisation processes in various contexts, including healthcare and sanitation. The comprehensive approach involving static ^{31}P NMR, 1D ^{1}H NMR assignments, T_1 relaxation analysis, and 2D ^{1}H/^{1}H NOESY MAS-NMR experiments has enriched our understanding of how sanitisers interact with lipid bilayers and how these interactions affect membrane structure and dynamics.

Chapter 6

Alcohol-induced denaturing of the SARS-CoV-2 Spike protein: A molecular dynamics and machine learning approach to protein sanitisation

6.1 Abstract

The S protein of SARS-CoV-2 is directly responsible for the initiation and propagation of virion-host cell binding and, thus, fusion. Due to the bespoke binding site in the RBM of the S protein RBD, it is exceptionally effective at binding to ACE2 and initiating host infection. As a result, the protein is generally the target of antimicrobials, antibodies and designed vaccines, as inhibition of this region will make successful infection significantly less likely while also providing a suitable binding site. For these regions, it is of significant interest to understand how sanitiser binding may affect the protein's functionality. It is also desirable to know at what concentrations it is effective and the mechanism of initiation.

Alcohol-based hand sanitisers have, in recent years, provided a broad means of slowing or stopping the transmission of SARS-CoV-2 between individuals. This is due to the ease with which they can be distributed and applied by the public whilst performing with high efficacy. Although these are generally thought to kill pathogens by destroying viral membranes by solubilising the lipid bilayer, it is thought that there may also be a significant denaturing effect on membrane-bound proteins. This may have significant implications for an alternate mechanism of slowing transmission by disabling crucial proteins, even at lower sanitiser concentrations. The mechanism by which they do this and the rationale behind where they successfully partition may have implications for how scientists can engineer protein-targeting sanitisers in future.

To explore these aspects, I propose a method by which I perform simulations in the presence

of sanitising molecules that deform said protein and then probe the mechanism by applying a novel machine-learning approach. By doing this, I avoid applying a confirmation bias and simplify complex simulation systems by supplying properties to an RFT algorithm so that I can examine the importance of these properties in the algorithm's decision-making process.

6.2 Introduction

Although alcohol-based hand sanitisers are still known to have high efficacy when applied to SARS-CoV-2, more recent mutations such as D614G have been associated with a decrease in the effectiveness of some sanitisers alongside an increase in transmissibility[203]. Although the SARS-CoV-2 pandemic is now mainly under control, it is essential to consider the origin of these resistant mutations so that we are better equipped when new, more effective mutations emerge. The implications of research into this could prove invaluable in treating future outbreaks or protecting those with immunodeficiencies in high-risk areas such as care homes from antipathogen-resistant strains.

6.2.1 Antipathogen activity of alcohol

Short-chain alcohols are theorised to kill pathogens by dissolving lipid membranes and causing cell lysis, a phenomenon explored in Chapter 4 regarding *E. coli* and *S. aureus*[126]. This is the reason alcohol is such an effective antipathogen, as this mechanism applies to both bacterial and virus lipid membranes. However, the disabling and destructive effects of these sanitisers may also be attributed to alcohol's ability to denature proteins. Deformation of essential proteins would render a pathogenic cell unable to function or remove its ability to bind to host receptor proteins. However, the exact mechanism of how proteins are denatured via this method must be more coherent and detailed mechanistically. Further analysis of this could reveal whether there are predictable starting points, and if this is the case, scientists may be able to engineer more targeted sanitisers in future; this is imperative as antipathogen resistance is only expected to worsen.

6.2.1.1 Destruction of protein structure *via* short-chain alcohols

Short-chain alcohols are well established as an efficient means of disrupting lipid bilayers and proteins. These sanitisers are responsible for interfering with hydrophilic interactions present in the protein, which maintain the higher-order structure. Further, it is theorised that this denaturing effect may have severe implications for the ability of proteins in cell envelopes to function as intended[91]. This effect of short-chain alcohols on protein structures has been studied in depth to show that when added, there are significant expansions of the protein chain structure[204]. Sashi *et al.* later elaborates on this to detail the exact effects of water exclusion caused by the addition of alcohol and how this may further affect the protein structure[205].

The effect of alcohol on proteins has been previously explored in terms of the immediate effect on the secondary structure by Kinoshita *et al.*, who found that upon application of alcohol to a protein, there is a significant increase in hydrogen bonding, contrary to what may be

expected[206]. This was due to changes in the protein environment, which caused solvophilic amino acids to become solvophobic and reorient to face inwards toward the protein core. This happened as the solution was no longer hydrophilic due to the addition of alcohol, which is amphiphilic. In doing this, these residues formed hydrogen bonds that were previously less beneficial, forming coils that rigidify the protein. Following this, the protein structure became further disrupted as more residues reoriented to form previously improbable interactions.

6.2.2 Supervised machine learning

Supervised machine learning is a method by which a computer will generate a model based on a known, labelled dataset. This is referred to as supervised because it is up to the user to define the label for the algorithm. These algorithms build predictive models of an area without specific domain knowledge by mathematically defining trends in the dataset. The theory is that a computer can be trained to make pattern-based predictions by mimicking how a human brain learns trends mathematically. The aim of using machine learning here was to extrapolate subtle trends that were too convoluted for a human to find and determine what features of these trends were instrumental in making good predictions.

6.2.2.1 Decision tree and random forest regression

Decision tree regression (DTR) is a technique which breaks a large dataset down into progressively smaller subsets. In this approach, the algorithm will make an initial prediction based on the mean of input features. Following this, the features will be reweighted, another prediction will be made, and the loss will be assessed. This process occurs iteratively, and the properties are probed by asking binary questions about the data (Figure 6.1(a)). By probing the data this way, the model can gradually reassess its weightings or add more nodes to minimise the loss between the prediction and the label. The benefit of this method is that it is relatively easy to train and employ, so we can make quick predictions to evaluate if there is a trend that we wish to explore further. However, this simplicity comes at the cost of being computationally cumbersome, unstable with continuous datasets and prone to overfitting if not managed correctly.

An RFT model is an average approach to a DTR that removes noise from the predictions, assuming that our initial DTR model was well-founded. It is well understood that a single DTR is prone to overfitting and bias, so averaging an array of these trees is a good way of improving the reliability of the prediction. In this method, an array of DTRs are trained with different weightings on different subsets of the input data. Training is primarily the same as a DTR, except a majority vote makes decisions from all trees, which means outliers and bias are less influential. At the same time, noise is cancelled by averaging many predictions. Regression is a beneficial approach as it enables us to rapidly generate results while also allowing us to extract the importance of features in our input dataset. This means we can make informed decisions on applying our dataset to more complicated models.

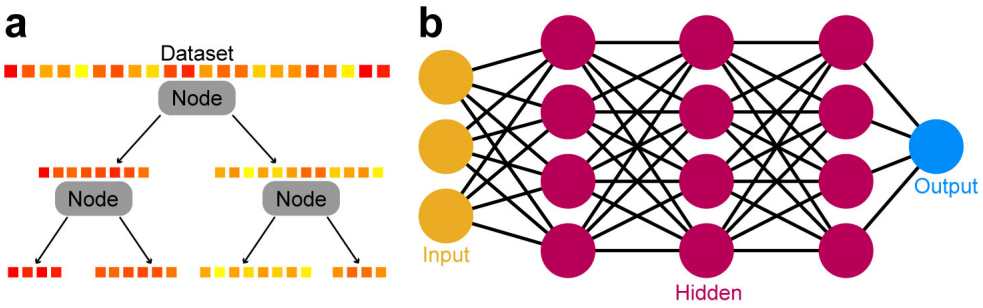


Figure 6.1: Computational graph representations of different machine learning techniques. (a) A DTR. (b) A simple neural network.

6.2.2.2 Neural networks

Figure 6.1(b) depicts a neural network. In this depiction, a node makes a transformation on data that is input from other connected nodes, weighted by the importance of the connection. In this sense, a node is similar to a neuron in a human brain. At the same time, lines between them are similar to synapses, similarly being enforced when positive reinforcement is given by increasing their weighting. Similarly to a DTR, the initial weighting is random and will give poor, uninformed results. The model will then back-propagate the loss and readjust the weighting of nodes to minimise loss. This is very similar to the process described for energetic minimisation of simulation systems, where we iteratively make small changes to the positions of atoms to minimise a multidimensional landscape, or in this case, we make minor changes to node weightings.

6.2.3 Validation and overtraining

The way validation is handled defines a supervised algorithm, as we check the predictions ourselves to ensure that the model is not doing anything unexpected. A model's training is assessed by measuring loss; this is how far off our model predictions are from what we expect them to be. This can be challenging as, ideally, we want to minimise the amount of loss; however, these models can sometimes be too good at predicting patterns to the point where they will learn the dataset rather than how to predict the underlying trend. This phenomenon

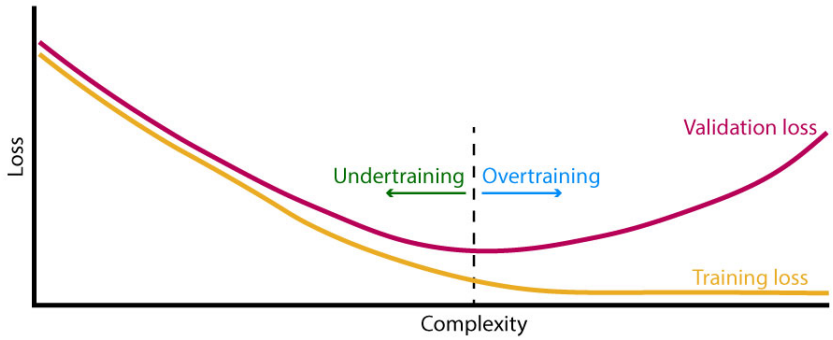


Figure 6.2: A graph showing an example of training and validation loss as the complexity of a model is increased to show the effect of under or overtraining on loss.

becomes clear when measuring the training error and the validation error. Training error is the difference between the model's prediction in training and the training labels, and validation error is the difference between the trained model's validation data predictions and the validation data labels, which the model did not see during training. The training error will continue to fall at a certain number of epochs, but the validation error will rise (Figure 6.2). This indicates the model learning the training data too well, to the point where it can predict the training labels almost perfectly and starts to learn stochastic relationships in the training data. These stochastic relationships exist in the training data purely by chance but are not representative of data outside the training data.

6.3 Methods

6.3.1 Simulation system setup and model parameters

The AA S protein ECD structure variant used to perform these simulations was the ancestral Wuhan strain supplied by the Pete Bond research group. The completed structure with attached glycans is described in their previous publication[207]. The glycan model used in this publication was based on liquid-chromatography-mass spectrometry (LC-MS) glycomics data[208]. Systems were solvated in 0.15 M NaCl using TIP3P water and were subjected to NVT and NPT equilibration steps before a 500 ns production run with 0%, 20%, 40% and 60% w/v PROH. PROH was parameterised via CHARMM-GUI with standard parameters for CGenFF. During equilibration, position restraints were imposed upon the protein backbone. Alcohol solutions in which the proteins were solvated were created in the same way as those in Chapter 4.

6.3.2 Simulation protocol

Simulations were set up and performed using the GROMACS[167, 115] MD software package (version 2020.2) with the CHARMM36 force field[168, 169]. Systems were maintained at 310 K, using the Nosé-Hoover[170, 171] thermostat with a time constant of 1 ps. The system's pressure was maintained at 1 atm, with a time constant of 2 ps, using isotropic pressure coupling with the Parrinello-Rahman[172, 121] barostat. All van der Waals interactions were cut off at 1.2 nm, and a smooth PME[173] algorithm was used to treat electrostatic interactions with a real space cut-off of 1.4 nm. Each system was subjected to 500 ps of NVT simulation for equilibration, followed by 2 ns of NPT. Positional restraints (1000 kJ mol^{-1} nm^2) were placed on the protein backbone during NVT and NPT equilibration. Production simulations were then performed without any positional restraints. The results were analysed using GROMACS tools and in-house scripts[115]. Visualisation was performed using the VMD[174] software package.

6.3.3 Contact analysis

Contact analysis of PROH to the S protein ECD was performed by running the contact analysis function from the heat mapper module with each chain as the analyte and PROH as the probe molecules.

6.3.4 Data preprocessing

Residue features were assigned to each residue using computational biology packages such as BioPython, MDTraj and MDAnalysis[209, 210, 176].

6.4 Results and discussion

To gain further insight into the process of alcohol based protein deformation, I have applied fully atomistic MD to systems containing the ECD of the wild type S protein in concentrations of 0%, 20%, 40% and 60% w/v PROH in neutralising 0.15 M KCl. In Table 6.1, a brief overview of the simulations discussed in this section is shown. Following this, I applied a DTR algorithm to discern properties or trends in amino acids which influence the interaction of proteins with them. In Table 6.1 the simulations run for this section are shown.

Alcohol concentration (% w/v)	Length (ns)	Repeats
0	500	2
20	500	2
40	500	2
60	500	2

Table 6.1: A brief overview of the simulations performed and analysed in this Chapter.

6.4.1 Structural degradation and rigidifying effect of propanol on secondary structure of the ancestral Wuhan variant ectodomain

500ns simulations of the wild type S protein ECD in PROH at concentrations of 0%, 20%, 40% and 60% w/v were performed in duplicate. To aid the reader, a visual representation of the trimer is shown in Figure 6.3. Viewing the protein in this way provides visual context for discussing these regions later in this chapter.

Visual inspection of these systems shows that when exposed to a higher concentration of PROH, glycans attached to the S protein would expand outward into the solution. To demonstrate this effect, the up-state monomer of the trimer is shown in a similar perspective to how it is displayed in Figure 6.3(d), in Figure 6.4. This shows that as alcohol concentration increased, the glycans attached to the monomer would spread out into the solution. For comparison, the glycans in the 0% PROH system are relatively compact even after 500ns of simulation and remain close to the protein structure. By 40-60% PROH concentration, the glycans appear to have fully extended. This is because PROH creates a shell around them. Figure 6.5 shows the shell of PROH created around the glycan, which stops regions in the glycans from interacting with one another or forming water bridges between them. This results in few glycan-glycan interactions as glycans are polar and will not interact well with the aliphatic tails of PROH. This may have substantial implications for the ability of PROH to reduce the stabilising effect of glycans. Zuzic *et al.* reported in their publication that when simulating the S protein with and without glycans, they were shown to stabilise the structure. Analysis of pocket structure and dynamics shows that, in some cases, there were major differences between the structure

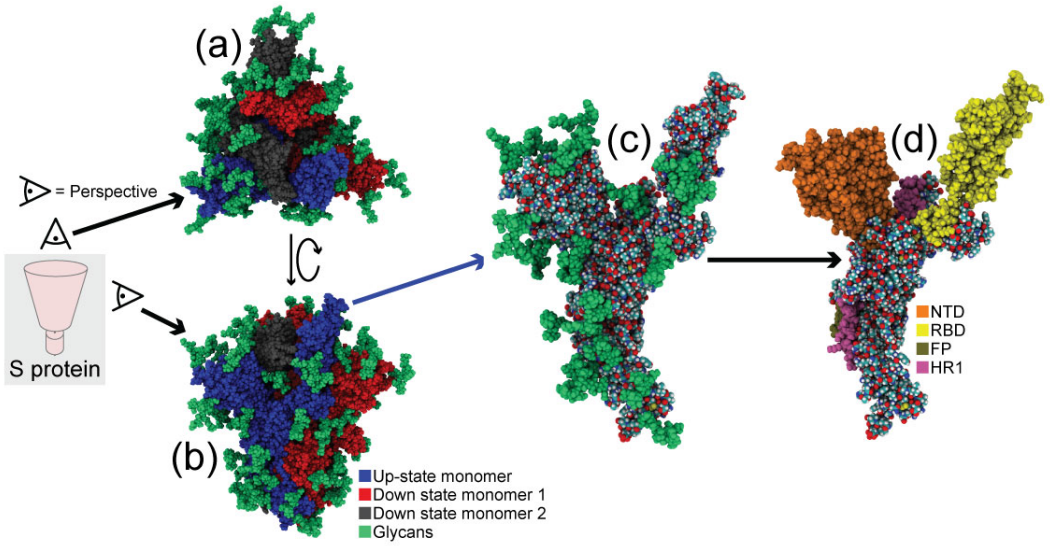


Figure 6.3: A visual representation of the S protein. The whole trimer from a (a) top-down (viewed from above the RBD) and (b) side-on perspective. (c) The up-state monomer viewed from a side-on perspective. (d) The up-state monomer viewed from a side-on perspective coloured by region.

with and without glycans. This branching effect of glycans may have further implications for protein stability. Further, this will affect glycans' ability to assist in protein binding, suggesting that sanitising alcohol could have a further effect on pathogens by disabling or reducing the function of glycans.

This same effect was not seen for protein residues, which reoriented differently when PROH was applied to the system. Polar side chains reoriented, changing conformation to face the protein core, as reported by Kinoshita *et al*[206]. Figure 6.6 shows a representation of the S protein RBD in the 0% and 60% PROH systems over time with sidechains highlighted. This shows that in the 60% system, the sidechains are substantially more compacted toward the protein backbone. This suggests that some residues are opposed to interacting with the solution and

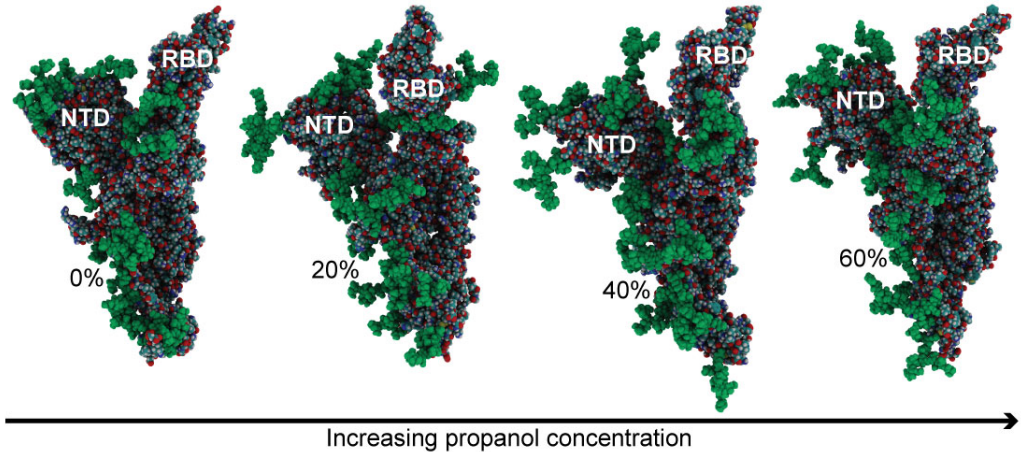


Figure 6.4: The up-state monomer of the S proteins ECD when exposed to PROH at increasing concentrations. Showing glycans in green and the S protein coloured by element.

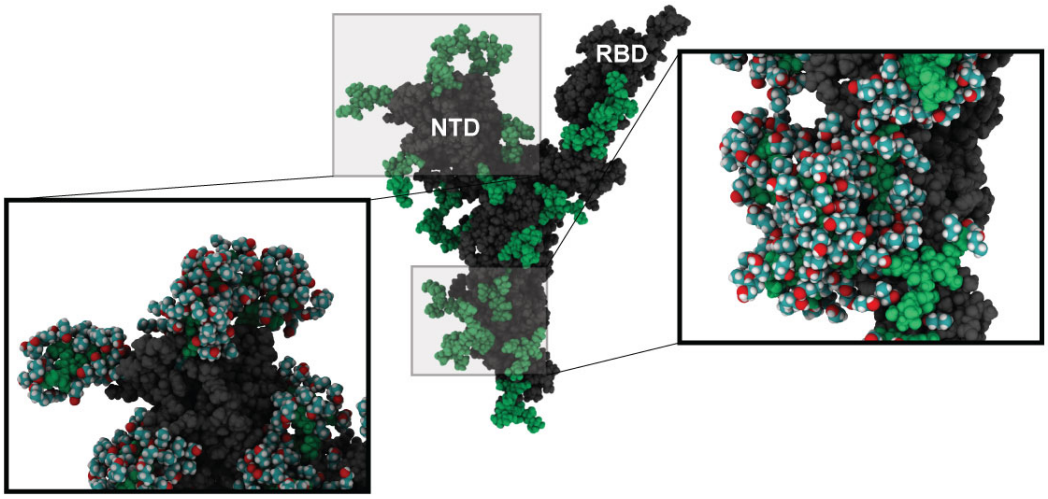


Figure 6.5: PROH shells formed around glycans in the 60% PROH system at 500ns, shown on the up-state monomer, showing the monomer in grey, glycans in green and PROH coloured by element.

are compressed to minimise their interactions. Inspecting the compressed residues suggested that these were often polar or aromatic. This is likely due to the large hydrocarbon portion of PROH, which will cap hydrophilic regions while the alcohol functional group interacts with them, similar to what was seen with the glycan chains. To investigate this further, intra-protein hydrogen bonding in the trimer was examined. Figure 6.7 shows that increasing the concentration of PROH in the solution had the counterintuitive effect of increasing hydrogen bonding rather than decreasing it. This value would be assumed to decrease, most likely due to the expectation of structural deformation; however, deformation was due to changing hydrogen bonding rather than simply reducing it. Due to significant changes in the solvent environment, when alcohol was added, interactions between polar residues and the solution became less beneficial. PROH capping of polar regions meant they were compressed into the protein core and less able to form interactions with the solution. This was because hydrogen bonding between these residues and PROH is significantly less beneficial than those previously formed with water due to the amphipathic structure of PROH. This resulted in an increase in

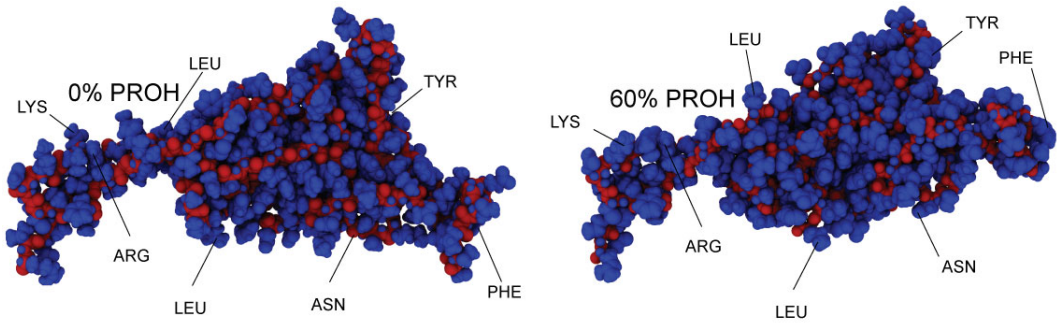


Figure 6.6: The RBD of the S protein in the up-state monomer after 500ns of simulation in 0% and 60% PROH. Showing the protein backbone in red and residues sidechains in blue.

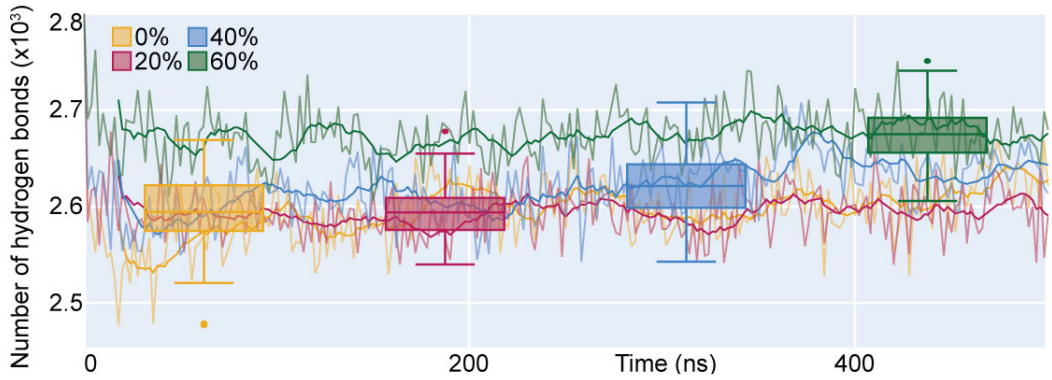


Figure 6.7: Intra-protein hydrogen bonding with a running average (lines) over 500 ns of production with a (box-plot) overlayed distribution.

hydrogen bonding in the protein as residues were forced toward the protein core.

Following this, I became interested in how this increased hydrogen bonding affected the secondary structure of the S protein. Figure 6.8 shows an analysis of the number of coils contributing to the ECD structure during production. Coils are irregular interactions in the protein structure and not considered as secondary structural terms. A coil refers to a disordered motif that lacks any form of repeating pattern. Although coils are disordered structures, they can have a rigidifying effect on the protein. After applying PROH at any percentage, there was an immediate increase in the contribution of coils to the secondary structure. This suggested that while polar residues in the protein reoriented themselves to face inwards, the protein backbone began to form intra-protein hydrogen bonds to minimise the system's energy. This contribution increased significantly from 0% to 20% PROH and remained similar in the 40% and 60% PROH systems. There was a marginal decrease in the 40% and 60% systems after 250ns; however, this could be stochastic. The resulting formation of a coil is depicted in Figure 6.9. In this figure, residues 330-350 (located in the RBD) of a monomer in the up-state were aligned for the 0% and 60% PROH systems to show an example of this coil formation. This shows the reorientation, which resulted in hydrogen bonding in the backbone and the formation of a coil. This effect was relatively common in positions where clusters of hydrophilic residues were reorienting as an abundance of residues could form interactions.

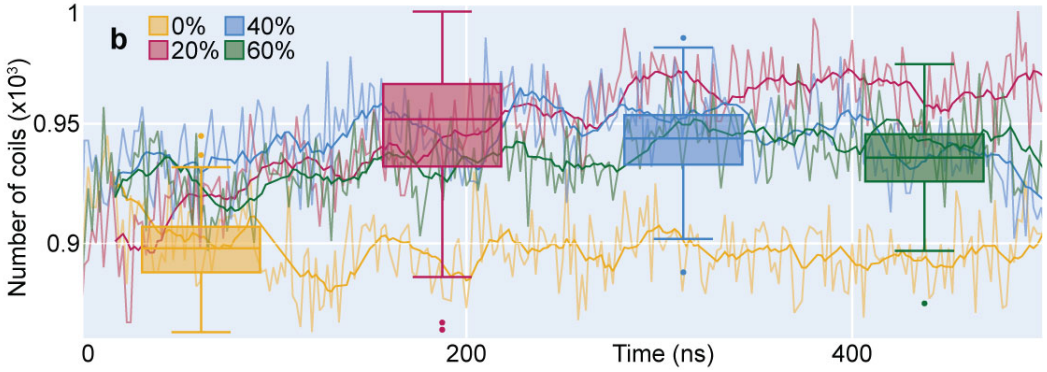


Figure 6.8: Number of coils in the secondary structure with a running average (lines) over 500 ns of production with a (box-plot) overlayed distribution.

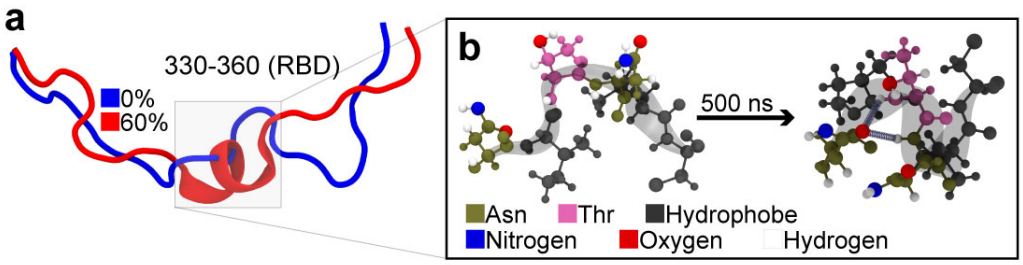


Figure 6.9: Analysis of structural changes in the protein when exposed to alcohol. (a-b) Representation of coil forming at residues 330-350 due to the addition of alcohol. (a) Cartoon representation of the (blue) control system and (red) protein exposed to 60% PROH, both at 100 ns into production. (b) A representation of the hydrogen bonding responsible for the coil formation.

Despite the contribution of coils to the proteins secondary structure increasing, the number of structural motifs (α-helicies, β-sheets and turns) decreased significantly (Figure 6.10). This was because although hydrogen bonding was increasing, it was doing so in an unstructured manner for which the protein structure was not equipped. Ordered structural motifs in proteins result from specific hydrogen bonding between different amino acids in the protein's primary structure. In this system, where high concentrations of alcohol increased hydrogen bonding, there was no increase in ordered secondary structure motifs as the hydrogen bonds formed were disordered.

To assess how the protein structure fluctuated when exposed to alcohol, the RMSF of both the protein's backbone (Figure 6.11) and sidechains (Figure 6.12) was analysed for an up-state monomer in the S protein. The backbone showed substantially greater fluctuation in the RBD region when alcohol was applied; however, this appeared to lessen as the concentration of alcohol was increased. It makes sense that residues in the RBD would show more fluctuation as the RBD is far more exposed to the solution and is not shielded by glycans in the same way as other regions, as it remains exposed to enable binding to ACE2. These fluctuations likely correspond to the formation of coils in this region due to its exposure to the PROH

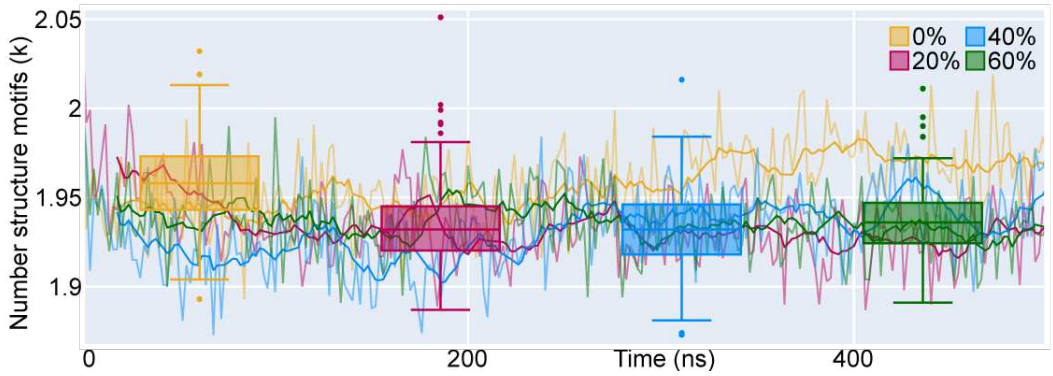


Figure 6.10: The sum of the number of α-helicies, β-sheets and turns contributing to the secondary structure with a running average (lines) over 500 ns of production with a (box-plot) overlayed distribution.

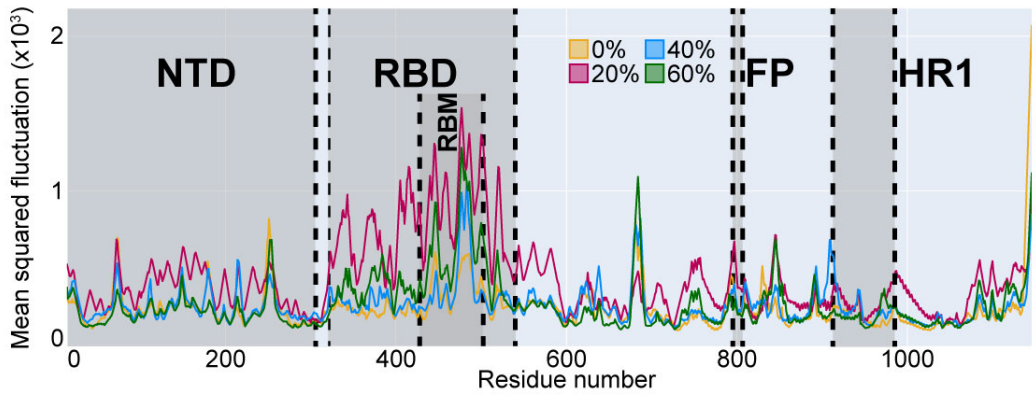


Figure 6.11: The RMSF of each residue in the backbone of the S protein when PROH is applied at concentrations of 0%, 20%, 40% and 60%.

solution. As mentioned, this results in the reorientation of the protein residues and forms coils in the secondary structure. The fluctuation most likely decreases when PROH concentration increases as this creates a more homogenous solution. While low concentrations of PROH mean residues may fluctuate between interacting with water or alcohol, higher concentrations of PROH mean fluctuations due to interactions with water become rare.

Similarly to measurements of the RMSF in the S protein backbone, there was also significant fluctuation corresponding to residues in the RBD (Figure 6.12). These fluctuations decrease again when the concentration of PROH exceeds 20% as the solution becomes more homogenous. Regions that show fluctuation are generally similar. The higher relative intensity of fluctuation in the RBD is likely due to a greater representation of polar or aromatic residues, as it contains the binding site. In agreement with this is the fact that the highest degree of fluctuation is, in both the backbone and sidechains, the receptor binding motif (RBM). This is the region of the RBD which binds to the ACE2 binding site directly and, as such, has a large amount of polar or aromatic residues.

RMSF measurements of the monomers in the down-state showed similar fluctuations when the backbone and sidechains (Figure 6.13 and 6.14, respectively) other than that of the RBD.

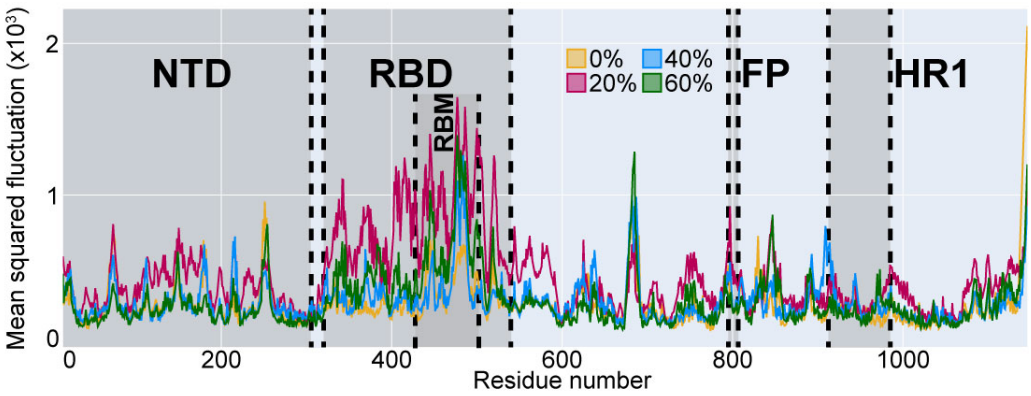


Figure 6.12: The RMSF of each residue of the S proteins sidechains in the up-state monomers in the up-state monomers when PROH is applied at concentrations of 0%, 20%, 40% and 60%.

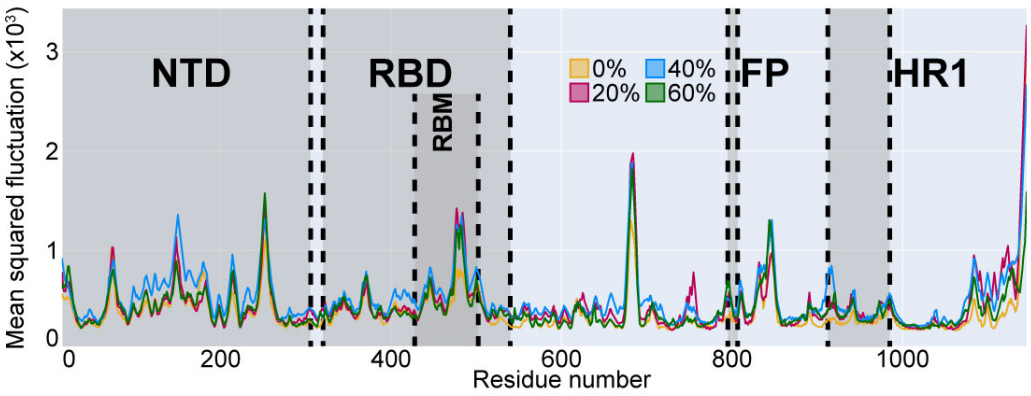


Figure 6.13: The RMSF of each residue of the S proteins backbone in the down-state monomers when PROH is applied at concentrations of 0%, 20%, 40% and 60%.

Because the RBD is buried in the ECD to avoid detection by the immune system, it does not interact with PROH as much as in the up-state monomers.

Finally, to thoroughly assess the structural stability of the ECD when exposed to PROH, the RMSD for the N-terminal domain (NTD, residue 13-305), RBD (residues 319-541), fusion peptide (FP, residues 788-806) and heptapeptide repeat sequence 1 (HR1, residues 912-984) domains were recorded over both 500ns production runs. This gave results that agreed with what was seen in the previous analysis. The NTD, FP and HR1 domains (Figure 6.15) all showed very large fluctuations in their RMSD when alcohol was added. When simulated in water without PROH, the fluctuation in these domains was negligible compared to that of the alcohol-soaked systems. In all three of these analyses, the average RMSD of all three concentrations that contained PROH was approximately the same. Error for the RMSD was exceptionally large in measurements for the NTD and FP measurements, which reflects previous findings that there was more fluctuation in systems that possessed a more inhomogeneous solution. In measurements for the HR1 domain, the error is exceptionally large at 60% PROH, likely due to how sheltered this region is. As the HR1 domain is buried in the protein structure, exceptionally large RMSD values were only recorded in systems with high concentrations; however, it was still rare.

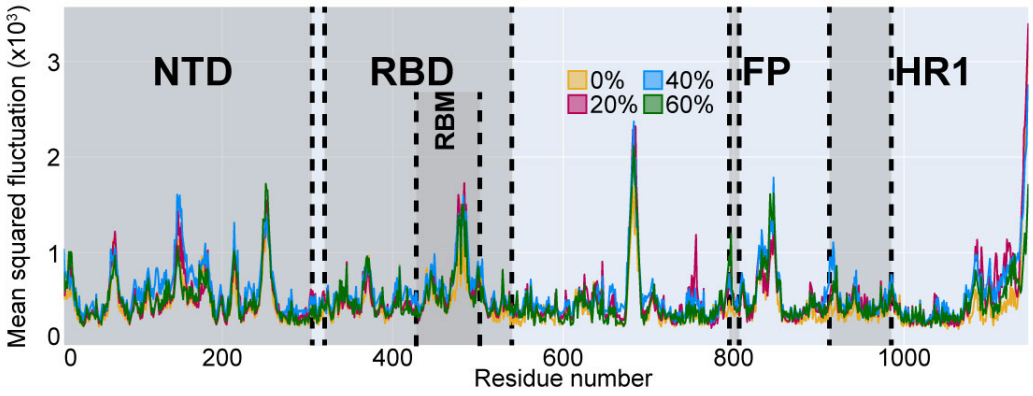


Figure 6.14: The RMSF of each residue of the S proteins sidechains in the down-state monomers when PROH is applied at concentrations of 0%, 20%, 40% and 60%.

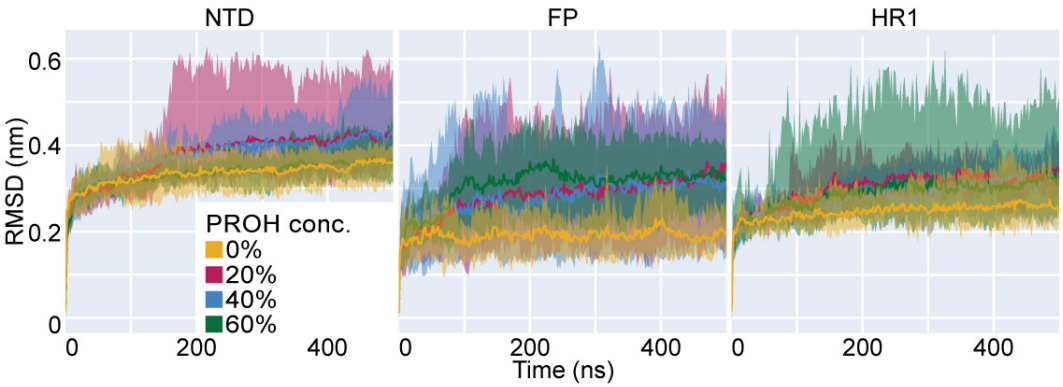


Figure 6.15: The average RMSD of the NTD, FP and HR1 domains across all monomers in both repeats, with the least squares fitting to the domain's respective protein backbone.

The RMSD of the RBD domain is, expectedly, higher in systems containing alcohol (Figure 6.16). The RMSD was also higher when in the up-state for the system containing no alcohol as the RBD is not buried in the ECD. Error was again large in the 20% system, suggesting that a less homogenous solution creates more fluctuation in the protein structure. This error reduces after 300ns of production, suggesting that the RBD reaches a stable state. The large error seen for RMSD measurements in the up-state RBD and NTD makes sense, as these regions are exposed to the solution. If any region were to experience a deformation of its structure due to the solution being inhomogeneous, it would make sense that it was in regions that are heavily exposed.

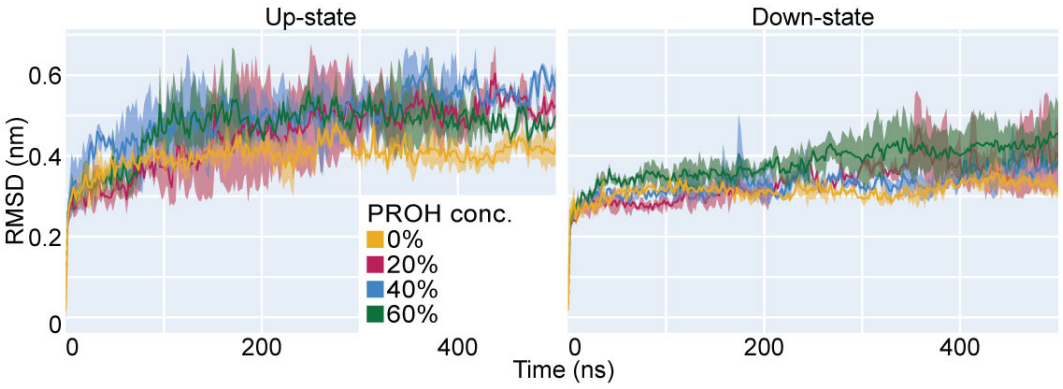


Figure 6.16: The average RMSD of the RBD domains in the up- and down-states across all monomers in both repeats, with the least squares fitting to the domain's protein backbone.

6.4.2 Contact analysis of propanol to the ectodomain

Having determined that the deformation of the trimer *via* alcohol was not uniform, my attention was turned to which points in the protein were most susceptible to these changes. Prominent regions such as the RBD, which showed a significant fluctuation in RMSF, were obvious targets; however, assessing where PROH interacts with the protein most may suggest less obvious points or trends. The heat mapper Python module described in Section 3.1 was developed to assess contacts between a probe molecule and each residue of the protein structure.

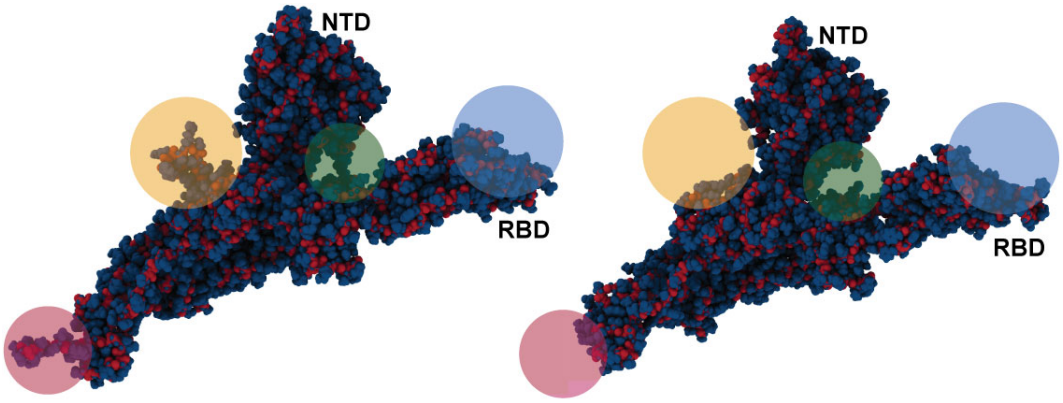


Figure 6.17: The monomer of the S protein in the up conformation when simulated in (left) 0% and (right) %60 PROH for 500 ns. Regions where noticeable changes are seen due to hydrophilic residues turning inwards are highlighted in yellow, red, blue and green.

As shown in Figure 6.17, regions highlighted show residues that moved toward the protein core to avoid the solution as it became soaked in alcohol. To further assess this, I monitored regions that experienced more contact with PROH and, using the heat mapper module scaled these onto a colour map. Figure 6.18(a) shows that the heat map of the S protein by contacts to PROH has a relatively even distribution of high contact residues. With this being said, when the

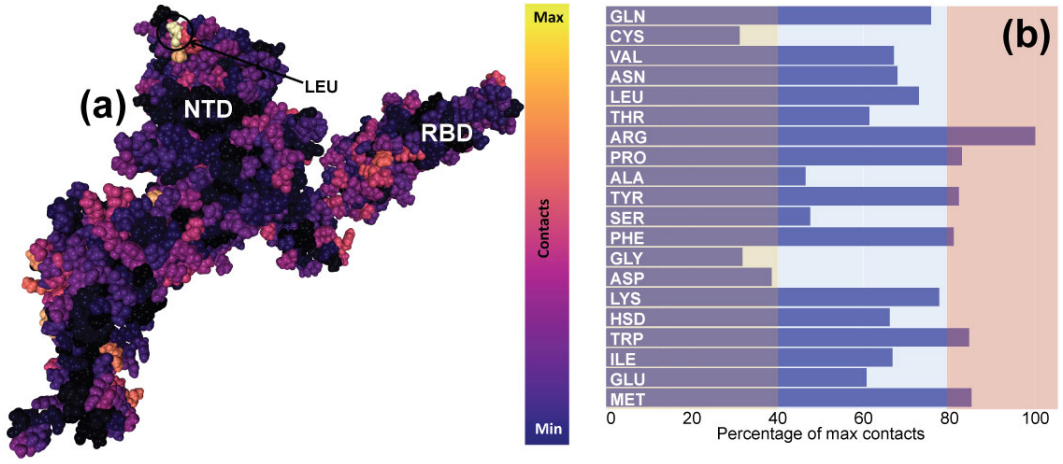


Figure 6.18: Contact analysis per residue. (a) Average contacts per residue of the up-state monomer in 60% PROH, with contacts scaled from (black) the least contacts to (yellow) the most contacts seen by any residues. A Leu residue with an uncharacteristically high number of contacts was highlighted in blue. (b) The average contacts are standardised by residue type and abundance on a scale of (least contacts) 0 to (most contacts) 1.

residues were analysed by their average number of contacts to PROH (Figure 6.18(b)), some of the residues with a large number of contacts had an over-representation relative to the amino acid types average number of contacts. Though this was far from definitive, it suggested that there may be structure characteristics other than residue type, which determined how much the protein was affected by the solution, such as structure, charge or hydrophobicity.

Considering hydrophobicity, it should be noted that there was an overarching similarity between the residue types, which tended to see more interactions with PROH. Figure 6.19 shows a trend between the contacts per residue and the residue size and hydrophobicity. These trends are weak; however, they potentially represent features that contribute toward determining how likely a residue is to interact with the solution. This suggests that the many properties of residues in a protein may contribute to their likelihood of interacting with PROH. When extracting the amino acids of the ECD that contacted PROH more it became clear that these were generally large, polar or aromatic amino acids (Figure 6.19(b)).

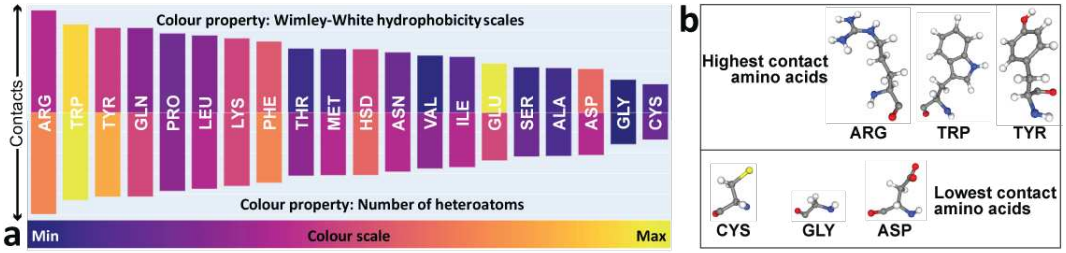


Figure 6.19: Analysis of the trend between amino acids contacts to PROH and hydrophobicity/size. (a) Contacts per residue type are coloured on a scale from (purple) low hydrophobicity/atom count to (yellow) high hydrophobicity/atom count. (b) The structure of the amino acids that saw the most and least contacts from PROH.

6.4.3 Predicting propanol interaction *via* feature engineering

In the previous sections, I have shown that the structural deformation caused to the ECD by PROH in different concentrations was less stochastic than it first appeared. Although a lot happens quickly, characteristic interaction points appear to be responsible for deformation. Of notable interest was whether there were also specific points in the structure where alcohol would partition into the protein and incite deformities. Further to this, regions that experience a considerable RMSF value may experience this due to local partitioning into known or cryptic pockets. Knowing where and how small chaotropes bind may have significant implications for designing sanitisers that target protein-binding sites rather than the pathogenic membrane.

Contact data detailing how many contacts each amino acid experiences with PROH in the surrounding solution was calculated using my in-house heat mapping library for Python. This library assesses contact distances to each amino acid from PROH in the solution per frame and returns this as contacts per amino acid per frame. Further details of this can be seen in Section 3.1. By providing this contact data and the characteristics of the local structure to an RFT machine learning algorithm, I obtained relatively good predictions of my validation data. It should be noted that I was not interested in predicting interaction points without

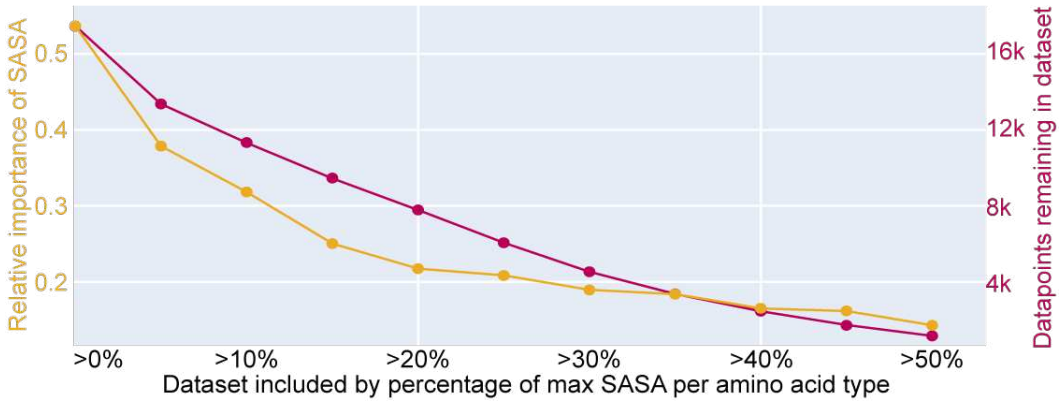


Figure 6.20: Analysis of feature importance after removing the lowest SASA amino acids using the Wimley-White hydrophobicity scale. Decrease in SASA importance relative to all feature importance as unavailable residues are removed.

running simulations but in determining which features were heavily weighted in making these predictions. The features supplied to the RFT are shown in Figure 6.20. In this figure, I show that when considering all amino acid contacts at each frame while dropping none, the SASA was extremely important in predicting contacts per amino acid. This made sense, as the number of contacts and the available surface area of a residue should be directly proportional, meaning regions of the protein sheltered by local folding or steric bulk were less susceptible to deformation. Logically, it follows that hiding the RBD may also increase the resistance of the protein to alcohol-based hand sanitisers. Although this result was meaningful, the importance of SASA dwarfed other features and made it difficult to determine the importance of more nuanced features. In Figure 6.20(a), I show that removing the lowest 20% of SASA amino acids from my dataset significantly decreases the importance of SASA in the decision-making process.

This shows that the RFT used SASA as a proxy to immediately attribute unavailable amino acids with a prediction of 0 or very few contacts, showing that there were regions of the protein that alcohol remained unable to access even after significant simulation time. When

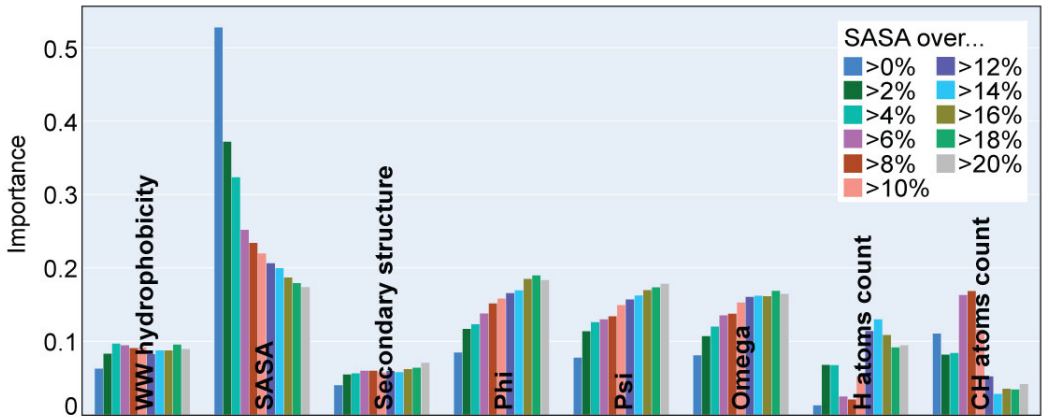


Figure 6.21: Relative change in feature importance as the lowest 20% SASA residues are removed.

unavailable amino acids were removed from the dataset, I immediately saw that the secondary structure, phi, psi, omega and the number of hydrogens in the amino acid became essential features in making predictions. In contrast, the sum of carbon and hydrogen atoms became less important (Figure 6.21). This suggested that in regions available to the sanitiser solution, the local structure of the protein was an essential factor in determining whether alcohol was likely to partition there. It is worth noting that the algorithm has no context for these decisions and reports the significance of a feature's values in correlation to an associated label. This means that although secondary structure was important, it was not clear whether this was simply a structural feature or some proxy related to the structure in that region. For instance, some structural motifs could be considered significant; however, this may be because certain amino acids that appear in them were predisposed to interacting with the alcohol more. From this, I show that the number of atoms in an amino acid became more important as SASA became less significant; however, as unavailable amino acids were removed, the critical feature switched from the number of carbon atoms to the number of hydrogen atoms. To ensure that the number of hydrogen atoms was not simply a proxy for larger amino acids, any feature related to the number of atoms in amino acids was removed from the dataset, and the RFT was trained on the remaining features. Those removed were added back in, one at a time, separately (Figure 6.22). This showed that when these features were added back in, and the RFT was trained over 1000 random seeds to remove bias, the number of hydrogens was consistently the most important feature relative to all features, followed closely by carbon+hydrogen, carbon+hydrogen ratio to all atoms and the atom count.

The significance of this was that the property this represents was amino acids with large, hydrophobic regions such as aliphatic chains. These would contain a relatively large amount of hydrogen and carbon. This would explain why hydrophobicity was not seen to increase or decrease in importance, as it was not necessarily the overall hydrophobicity that was important but the distribution of hydrophilic and hydrophobic regions within an amino acid. The mean number of contacts to amino acids with different numbers of total atoms, hydrogens and carbons are shown in Figure 6.23(a). This showed a general trend of increasing mean contacts as the number of atoms in a residue increased; however, the trend was very noisy. There

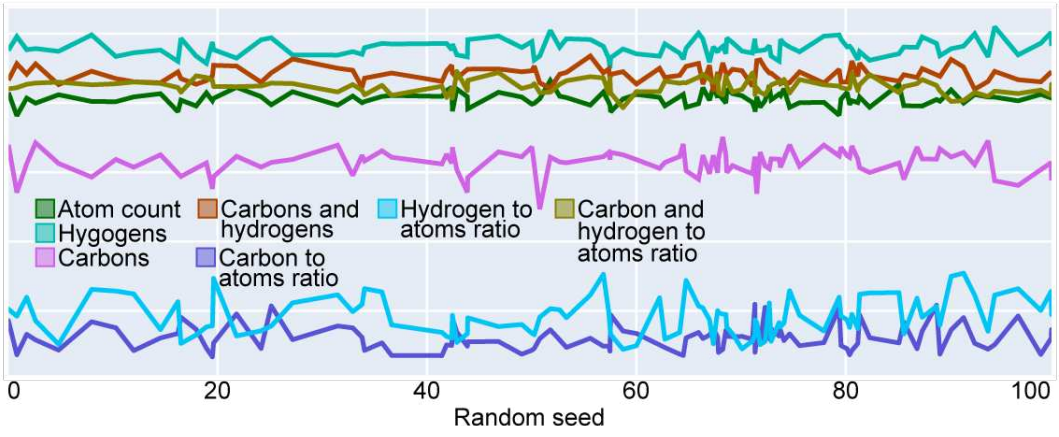


Figure 6.22: The importance of atom counts and ratios as a feature as they were included separately over 1000 random seeds.

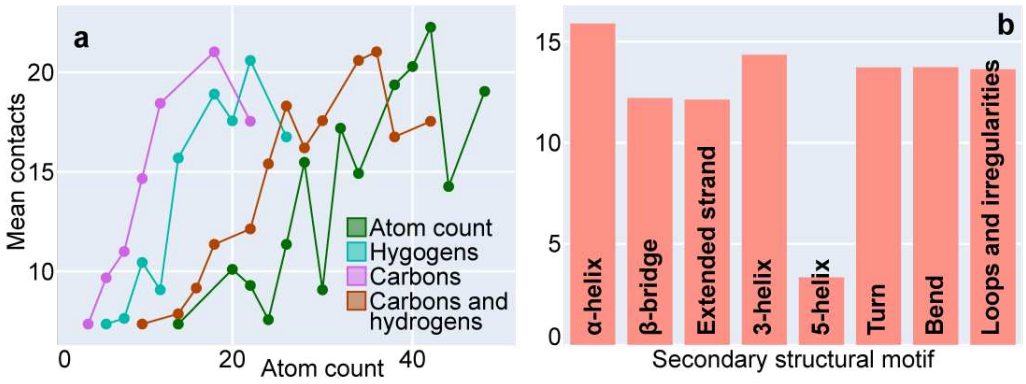


Figure 6.23: Analysis of the average contacts per residue. (a) The mean contacts per residue type per residue. (b) The mean contacts per secondary structure motif per residue.

was less noise when plotting to mean contacts against the hydrogen and carbon count, which decreased further when plotting only hydrogen or carbon. This trend was attributed to the carbon/hydrogen count acting as a proxy for amino acids with large, hydrophobic regions where the aliphatic chain of PROH can align itself and minimise its interaction with water. These amino acids would have a disproportionately large number of hydrogens and carbons relative to their overall size, as the aliphatic region was exclusively composed of hydrogen and carbon. Another feature that saw increased importance when unavailable amino acids were removed was the secondary structural feature (Figure 6.23(b)). Similarly, the mean contacts to amino acids in each secondary structural motif were plotted and showed little difference other than in 5-helical amino acids. However, these refer to amino acids buried in the core of the trimer, so this finding was analogous to that of SASA being used as a proxy for unavailable amino acids.

6.5 Conclusions and further work

In this section, I explored the nuances of how alcohol-induced deformation impacts protein structures, focusing specifically on the ECD of SARS-CoV-2. The first notable effect was that PROH would cap glycans attached to the ECD, potentially affecting their ability to contribute to protein and in agreement with findings of Zuzic *et al.* that glycans are important in stabilising the proteins structure[207].

Further, the deformation of the protein by short-chain alcohols presents a substantial change in the structural features contributing to the secondary structure. This change was primarily due to the dehydrating effect that the application of short-chain alcohols incites, causing reorientation of amino acid residues. I observed irregular reorientation of protein residue, resulting in the formation of new coils in previously unfavourable positions while side chains were compressed into the protein structure. The change in structural motifs heralds significant changes to the overall structure, which affect the protein, potentially impacting its ability to function adversely.

In-depth analysis uncovers a more nuanced relationship between PROH application and the deformation of proteins. Changes in the contribution of coils to the secondary structure and the effect that this has on RMSF suggest that this relationship is not simply linear. This analysis showed the contribution of coils to the secondary structure of the S protein at 20% PROH. Intra-protein hydrogen bonding was seen to increase linearly. This shows the transition of the protein from a stable state in water, past a fluxional structure in an inhomogeneous solution, into a relatively stable conformation as the solution becomes homogeneous again. This transition is supported by RMSF measurements, which show a spike at 20% PROH. Although the structure stabilises as the solution becomes more homogenous, there is a clear degradation in the secondary structure, and the protein is not necessarily functional.

Considering the predictability of the partitioning shown in Chapter 4, it seemed likely that similar interactions could determine the locality of alcohol-protein interactions. Characteristics such as the protein region's availability to the surrounding solution were significant in deformation; however, feature engineering allowed me to probe this further. Excluding the least available residues in the protein reduced the importance of SASA as a feature and increased the importance of structural features. This implied that for residues accessible to the solution, the local structure and amino acid properties were important. The relative importance of the hydrogen count over other metrics that measure the amino acid size suggested that this was, however, not the size that was necessarily important but the size and presence of a hydrophilic-hydrophobic boundary. These amino acids generally contained the largest number of hydrogens as they possessed aliphatic regions where hydrogen is abundant. This showed that the number of hydrophobic atoms in an amino acid can be a proxy for how large the hydrophobic portion of said amino acid is. This allowed me to conclude that alcohols similarly partition into proteins as they do in membranes, into a hydrophobic-hydrophilic interface.

Future work in this area would address limitations in this method, such as the timeframe of simulation, RFT or experimental corroboration. The timeframe of these simulations, although

extensive for AA simulations of this size, appears to have only captured the initial phases of deformation. With this as a consideration, repeating these works using techniques that allow sampling of longer timeframes could provide insight into how these interactions evolve. RF, although effective and relatively simple to apply, is not necessarily an exhaustive predictive ML method. Although better predictions could likely be achieved with a neural network, RFT was chosen as feature importance values are easily extracted. However, if this work were to be repeated, a more comprehensive approach using explainable AI could provide valuable insight. Experimental corroboration would be an invaluable means of improving these works as it could provide insight into how this process progresses without time constraints. Applying techniques such as hydrogen-deuterium exchange (HDX) could provide insight into how contacts to the solution are distributed in vitro. However, this would likely correspond to a protein in a significantly more denatured state.

Chapter 7

Assessing SARS-CoV-2 spike-receptor cross-linking in coarse grain simulations

7.1 Abstract

The successful infection of a host cell by SARS-CoV-2 depends on the initial interaction between the virion and a host cell receptor. The S protein of SARS-CoV-2 needs to form strong interactions with the binding site of the receptor to ensure efficient cell recognition and fusion. Although this process has been studied extensively, there is still a lot that we do not know about it. Understanding how S proteins bind will make future research of viruses much easier, as these proteins are not unique to SARS-CoV-2[211, 212]. One area of particular interest is the stoichiometry of S protein to ACE2 binding and how this affects the stability of the protein complex. As the S protein is a trimer, it possesses three RBD regions, each with the potential to bind to ACE2. Although the S protein trimer has the potential for 1:3 binding of S protein to the ACE2 complex, this does not mean it is necessary or sufficient.

Research in this area is vital because it provides valuable information that could be used to develop faster life-saving solutions for outbreaks similar to SARS-CoV-2. Understanding the binding mechanism of these viral proteins means we are better equipped to inhibit them. CG MD simulations were used to study systems containing 1, 2, or 3 S proteins cross-linked to between 0 and 3 ACE2 complexes. The simulations showed that higher-order cross-linking creates intrinsically unstable conformationally restrained complexes that are less likely to exist in nature.

7.2 Introduction

7.2.1 Cross-linking

Please note that in this section the use of the term "cross-linking" means the binding of proteins between two membranes, such as that seen in SARS-CoV-2 binding.

Cross-linking is necessary for many viruses to successfully merge the virion with the host cell that it will infect[213]. Cell surface associative fusion is necessary for SARS-CoV-2 to infect a host cell, and we must understand the binding mechanism well to develop sanitisers or therapeutics that can effectively inhibit it (Figure 7.1). Binding occurs as the RBM in the RBD of the S protein extends from the ECD and binds to receptor proteins such as ACE2[213]. Benton *et al.* found that by mixing the ECD of the 2P mutant of the S trimer with ACE2, they could analyse how the trimers bound[214]. This showed that 16% of all S trimers remained unbound with a single monomer in the up position. At the same time, 49% had a single up monomer bound to ACE2. From this, they only saw 14% and 3% bound to two or three ACE2 moieties, respectively. The conclusion is that binding from the S trimer has preferred coordination. This is why most bound complexes were in a 1:1 coordination (Figure 7.1). Research by Benton *et al.* shows that the RBD undergoes a rigid-body rotation during the receptor binding process. This is important as it suggests that the conformation and dynamics of the S trimer are well-defined and rely upon small, nuanced conformational changes. 1:1 binding may be energetically favourable and allow for small conformational changes such as the rigid body rotation of the RBD.

7.2.1.1 Structural biology and mechanisms of the SARS-CoV-2 S protein

The S protein is a trimeric fusion protein found utilised by SARS-CoV-2 for cell entry. The S protein is compose of two subunits referred to as S1 and S2; S1 contains the RBD and S2 contains the fusion peptide which is responsible for S protein-lipid membrane fusion before entry.

One of the more nuanced features of the RBD is that it can switch between two defined states

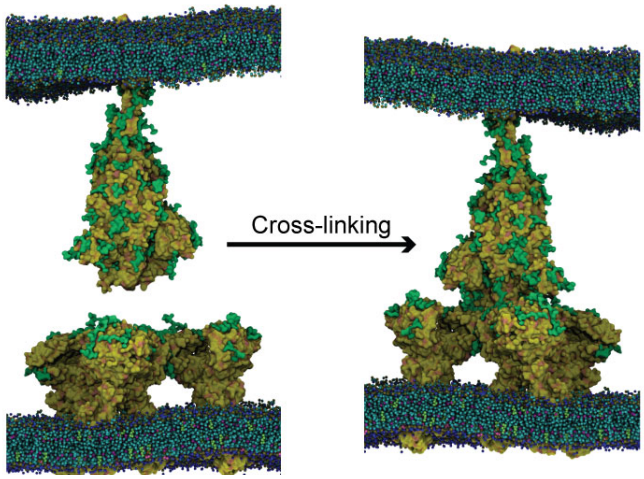


Figure 7.1: Visual representation of cross-linking.

referred to as the up and down states. In the up-state the RBD is exposed and able to bind to the receptor domain of ACE2 which is essential for entry of the virus into the host cell. In contrast, the down-state hides the RBD in the ectodomain of the S protein and means that the S protein is not able to initiate receptor binding. The down-state is necessary as it avoids detection of the RBD by the host immune system and allows the protein to avoid detection until binding. Where as the rest of the S protein is shielded by glycans to avoid detection, this is not the case in the RBD as this would make protein-protein interaction between the S protein and ACE2 difficult.

7.2.1.2 Monitoring conformational changes

Work performed by Turoňová *et al.* reports the presence of three hinges in the neck region of the S protein, which they term the "hip", "knee", and "ankle"[215]. In their study, they combine cryoelectron tomography (CET) and MD to report on the flexibility of the neck region in the S protein. They note that the hip region is the least flexible while the ankle and knee regions are more flexible. Similar work by Choi *et al.* showed that the neck region of the S protein is confined to energetically stable conformations by measuring the probability density of bending motions in the neck region of the S protein throughout production in AA MD simulations[216]. They detail the conformational dependence of the S proteins in terms of a series of angles and distances describing the protein's conformation during simulation. estimating the S protein's tilt angle as high as 90° but generally about 48°. They also theorise that this tilt is necessary for multiple S trimers to bind to a singular ACE2 receptor complex. The conformational changes in the S protein, as represented by angles defining changes in the neck region, are also detailed in work by Zuzic *et al.* Zuzic *et al.* report the distribution of the ankle and knee regions reported by Turonˇová *et al.* Therein, they reference similar works that also report the S protein's conformation by referring to angles measured for the S protein. They report that the tilt angle observed for the S protein in the model validation was as high as 90° and a product of the ankle and knee angles, in agreement with Turoňová *et al.* and Choi *et al.* Following this Zuzic *et al.* state that the flexibility exhibited by the S protein neck region is likely beneficial as it allows for greater sampling of the RBD at the host cell receptor region. This would potentially increase the probability of recognition and fusion. This observation is supported by the work of Raghuvamsi *et al.*, who further propose that the conformational freedom of the S protein is changed upon binding to ACE2[217].

7.2.2 The ACE2 receptor

ACE2 is a crucial receptor protein located in the membrane of cells in multiple areas of the human body and mammals[218]. The primary function of this receptor is to regulate the renin-angiotensin system. The chemical angiotensin is created when blood flow to the kidneys is reduced. This goes through further processing but eventually is recognised by the ACE2 receptor, which results in the narrowing of blood vessels to increase blood pressure[219].

The S protein exploits ACE2 to bind to and merge with mammalian cells to allow viral RNA insertion[110, 220]. This protein is a good target for the S protein as it is found on the

surface of various cell types throughout the human body, including those in the respiratory tract, lungs, heart, kidneys, and gastrointestinal tract. The benefit is that there are many potential infection points, and the virus does not have to rely on finding a specific cell type for successful infection. ACE2 has even been found to be expressed in areas as exposed as pores in the human skin[218].

Generally, ACE2 is found as [ACE2:B^0AT1]$_2$; this dimer-of-dimers is shown in Figure 7.2 for reference. B^0AT1 is a neutral amino acid transporter. However, for this work, it simply acts as a chaperone protein. The details of receptors that bind to the S protein are essential to help us understand this interaction better.

7.2.3 Cell infection

A series of crucial steps follow the successful binding of the S protein's RBD with ACE2 in a host cell membrane. The conclusion is a successful host cell infection by the SARS-CoV-2 virus. Once the RBD of the S protein binds to ACE2 (Figure 7.3(a)), the S1 domain of the bound S trimer will dissociate from the S2 domain, triggering a conformational change that stabilises the S2 domain (Figure 7.3(b))[221]. This exposes the primed S2 domain, which was previously masked by the S1 domain, the removal of which reveals the fusion peptide (FP). The FP domain is a short sequence of amino acids that facilitate the merge of the viral envelope with the host cell membrane. The FP will then embed itself in the host cell membrane (Figure 7.3(c)). This destabilises the host cell membrane, allowing the viral particle and host cell membranes to merge. The S2 domain contains two heptad repeat (HR) regions, HR1 and HR2. These regions interact with one another and form a pre-fusion hairpin conformation (Figure 7.3(d)). This brings the bacterial membranes together and allows the viral particle and host cell membranes to fuse (Figure 7.3(e))[222]. Following this, viral RNA translocates from the virion to the host cell[51]. To assist the reader, a visual representation of the S protein domains mentioned in the previous description is shown in Figure 7.4.

Understanding the mechanism involved in the virus binding to the host cell and successfully achieving infection is crucial to research. By understanding the process from start to finish, we can identify points at which targeted therapeutics could intervene. Research into points in

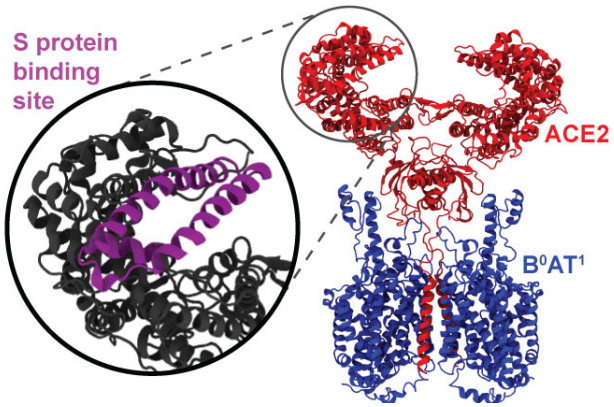


Figure 7.2: A visual representation of the ACE2-B^0AT1 heterodimer of dimers, showing the S protein binding site. This ACE2 structure is taken from PDB 6m17.

this process where the protein is strained or unstable may provide us with new, creative ways of destroying or disabling harmful pathogens.

7.3 Methods

7.3.1 Simulation system setup and model parameters

The epithelial membrane composition was based on the known lipidome[224]. The ERGIC membrane is less well-defined in terms of composition and is known to be dynamic. However, it was based on what is known of the lipid composition[225]. These membranes were then generated using CHARMM-GUI's membrane builder tool[162, 163, 164]. S proteins were aligned to ACE2 receptor binding sites in VMD using known PDB models of the binding in all coordinations (7a96, 7a97, 7a98)[174]. Membranes were added to the aligned protein complexes by aligning in VMD and remove clashing membrane lipids. The systems were then solvated in 0.15 M NaCl (Figure 7.5). The RBD and ACE2 binding domains were released from position restraints for a short period at the end of equilibration to ensure binding was well modelled.

7.3.2 Martinization and glycosylation

The ancestral Wuhan strain was the S protein variant used to perform these simulations. The Pete Bond research group provided the protein structure, which was converted to the CG resolution using the Martini martinize script[226, 227]. The S protein trimer with the required number of up-state RBDs was aligned with each system's receptor sites of the appropriate number of ACE2 complexes before membranes were aligned. Alignment was performed in VMD by overlaying the CG S protein and CG ACE2 receptor complex over a fully atomistic structure of the RBD bound to ACE2. Both the S protein and ACE2 structures were generated

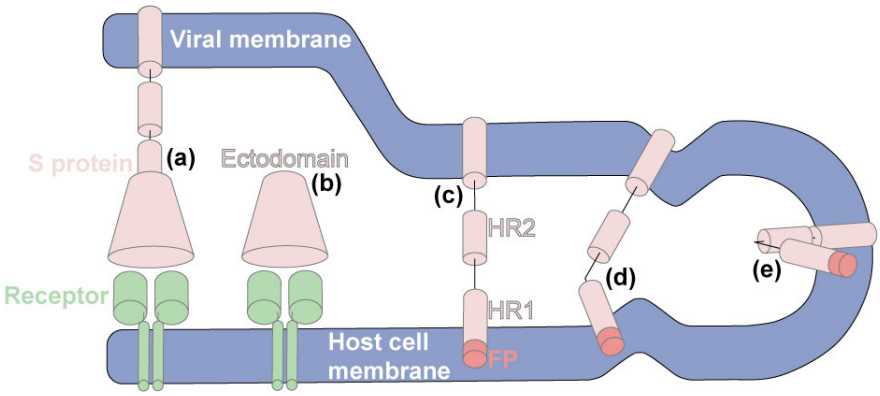


Figure 7.3: S protein binding and the resulting membrane fusion between the viral particle and the host cell. (a) The binding of the S protein to the receptor in the host cell membrane. (b) Cleavage of the S1 domain from the S2 domain of the S protein. (c) Embedding of the fusion peptide, previously masked by the S1 domain, into the host cell membrane. (d) Interaction of the HR1 and HR2 regions to form the pre-fusion hairpin conformation. (e) Fusion of the viral particle membrane and the host cell membrane. This diagram was inspired by work published by Heald-Sargent *et al*[223].

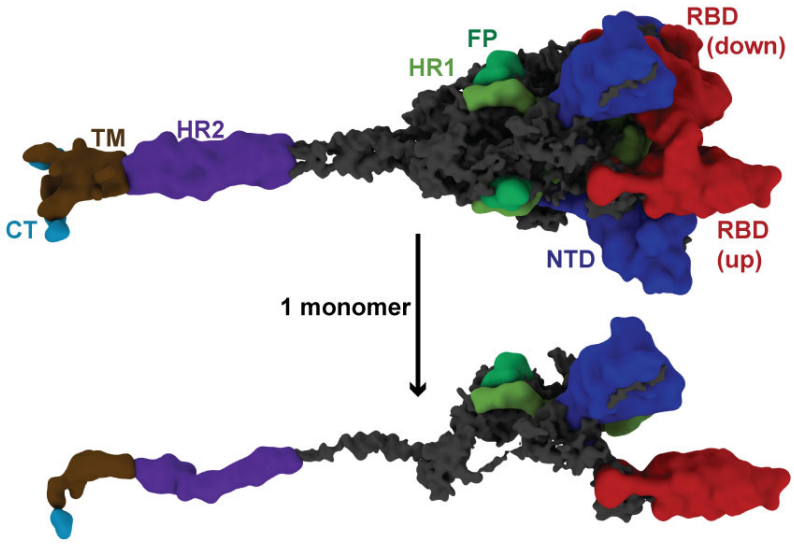


Figure 7.4: A visual representation of the S protein with all of the structure other than specified domains in grey.

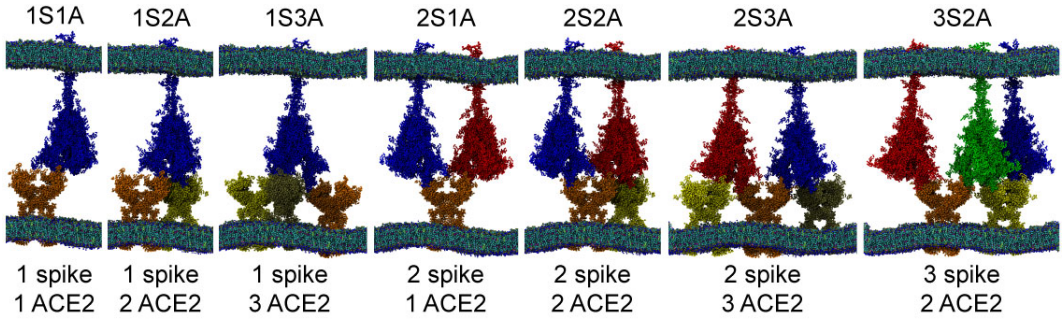


Figure 7.5: A representation of the systems simulated to explore the effects of cross-linking, showing the S protein in blue, red and green and ACE2 in orange, yellow and tan.

with an elastic network using the Martini 22 forcefield, an elastic bond force constant of 500, a lower cutoff of 0.5, an upper cutoff of 0.9, a decay factor of 0, a decay power of 0 and a disulphide bonding factor of 0.3. Second, structures of the single S systems were generated to examine the effect of the underrepresentation of the secondary structure in high-order complexes. Glycosylation of the CG structures was performed using the Martini glycosylator script. Glycans and the respective position at which they were added to the S protein and ACE2 are shown in Supplementary Table 9.2.

7.3.3 Simulation protocol

Simulations were set up and performed using the GROMACS[117, 115, 116] MD software package (version 2020.2) with the Martini version 2[227] force field. At the time of conceptualising this project, Martini 3 did not sufficiently support the components of these systems. Systems were maintained at a temperature of 310 K, using the V-rescale[228] thermostat with a time constant of 1 ps. The system's pressure was maintained at 1 atm, with a time constant of 5 ps, using semi-isotropic pressure coupling. The Parrinello-Rahman[172, 121] barostat was

used. All van der Waals interactions were cut off at 1.1 nm, and a reaction-field[229, 230, 231] algorithm was used to treat electrostatic interactions with a real space cut off of 1.1 nm. Systems were minimised with two consecutive steps using a steep algorithm, the first of which was a softcore minimisation. Each system was subjected to consecutive 1 ns NPT ensemble simulations for equilibration with gradually decreasing position restraints (Table 7.1). Restraints in the RBD and ACE2 binding domain were released in the final ns of equilibration to allow relative equilibration. Production simulations were then performed without any positional restraints. The results were analysed using GROMACS tools and in-house scripts. Visualisation was performed using the VMD[174] software package.

7.3.4 Conformational measurements of the S protein

To assess the evolution of the S protein conformation during production, similar measurements to those reported in various publications were measured *et al*[216, 207, 214, 215]. To assess general changes in the S protein conformation, the tilt angle of the S protein, RBD-NTD distance, RBD-NTD angle, ankle and knee angles were monitored. These measurements were reported by reading residue coordinates via MDAnalysis. To measure the tilt, the angle between the vector from the TM domain (residue 1213–1239) to the hinge region (residue 1208–1212) was

HR2 domain (residue 1157–1207), the hinge region (residue 1208–1212), and the TM domain (residue 1213–1239). The knee is measured as the angle between HR2 domain, the hinge region (residue 1142–1156) and the ECD (residue 27–1141)

Step	Time step (ps)	Total time (ns)	Restraints (kJ mol^{-1} nm^{-2})
1	200	1	200
2	500	1	100
3	500	1	50
4	500	1	20
5	500	1	0

Table 7.1: Equilibration steps used to set up cross-linked S protein-ACE2 systems

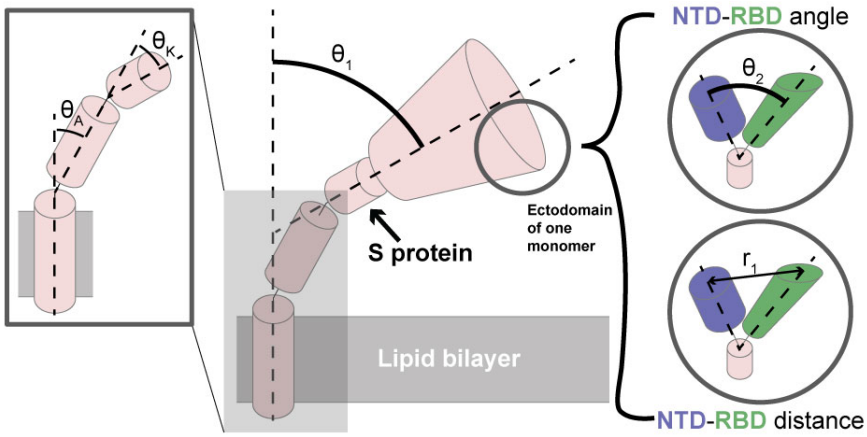


Figure 7.6: A visual representation of the S protein tilt angle (θ_1), the NTD-RBD angle (θ_2), the NTD-RBD distance (r_1), the ankle angle (θ_A and the ankle angle (θ_K.)

7.4 Results and discussion

In Table 7.2, a brief overview of the simulations discussed in this section is shown. Systems in this section are often referred to as n_1Sn_2A, where n_1 is the number of S proteins in the system, and n_2 is the number of ACE2 dimers present. Due to the size of the proteins involved and the increase in simulation box dimensions upon creating large cross-linked complexes, the work performed here was done in CG resolution. This meant the production runs could be significantly longer, allowing me to assess conformational changes over long timeframes.

Name	S proteins	$[ACE2:B^0AT1]_2$	Length (μs)
1S0A	1	0	3
1S1A	1	1	3
1S2A	1	2	3
1S3A	1	3	3
1S0A (NE)	1	0	3
1S3A (NE)	1	3	3
2S1A	2	2	3
2S2A	2	2	3
2S3A	2	3	3
3S2A	3	2	3

Table 7.2: A brief overview of the simulations performed and analysed in this section. NE denotes that a simulation had no elastic network applied.

7.4.1 Conformational requirements of the S protein

To assess the changes in conformation and stability when the S protein is cross-linked in high-order systems, I first simulated the S protein in an ERGIC membrane without ACE2; this allows the trimer to adopt the most energetically stable conformation. The most obvious result was that the protein's tilt gradually changed during production. As mentioned in Section 7.2.1, previous work by Choi *et al.* and Zuzic *et al.* estimate the tilt angle of the S protein as being as high as 90° but very flexible[216, 207]. It is theorised that this tilt was necessary for multiple

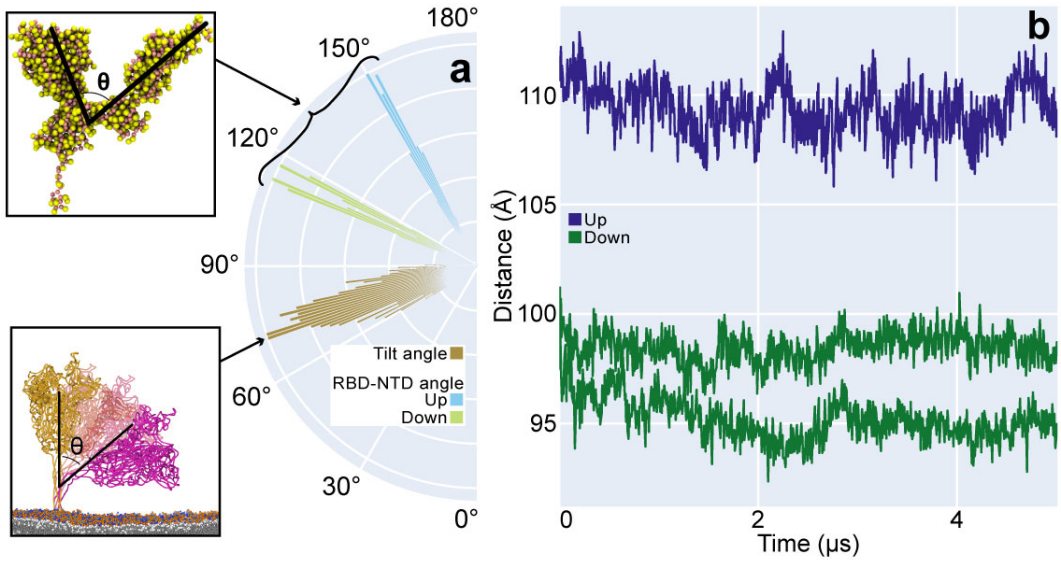


Figure 7.7: Analysis of the S trimers conformation alone. (a) Distribution of tilt angles of the S protein relative to the membrane normal and the angle of the RBD relative to the NTD about the core of the ECD per monomer between 2.5 and 5 μs in simulation. (b) The distance between the tips of the RBM and NTD in all three monomers.

S trimers to bind to a singular ACE2 receptor complex and to enable efficient binding. This could, however, impose particular conformational requirements on the system, which strain highly ordered systems.

In my simulations, the S protein tilt was seen to adopt an angle of approximately 70°, which is reasonable as it is exceptionally flexible in the neck region (Figure 7.7(a)). Following this, the angle between the RBM and NTD about the core was measured for both the up and down monomers to assess how well-defined the binding was. The simulated S protein model had one up-state RBD, and there was a distinct difference in the angle between the RBD and NTD in the up-state versus the down-state. The up-state RBD was not seen to deviate from this state. Similarly, when the distance between the RBD and NTD was measured, the two states also showed a distinct difference. Measuring this distance showed, as expected, that there was a distinguishably greater distance between the two when a monomer adopted the up-state (Figure 7.7(b)). In the unbound state, S monomers which were not coordinated were modelled in the down-state and remained that way throughout production, which buries the RBD in the ECD to avoid immunodetection. This resulted in the RBD and NTD being approximately 10 Å closer together in down-state monomers.

These measurements act as a control for further work, which will focus on the ACE2-bound protein. Binding various size S protein clusters with various size ACE2 clusters will restrict the dynamics, which can then be compared to the control. This allowed me to monitor whether the stress imposed by the S protein attempting to adopt the tilt angle had a visible effect on the system. Stress imposed in one region may affect aspects such as the RBD state or the rigid body rotation in the RBD. In theory, binding multiple ACE2 receptors could improve the chance of successful infection; thus, multivalency may be hypothesised for the S protein to bind at as many points as possible.

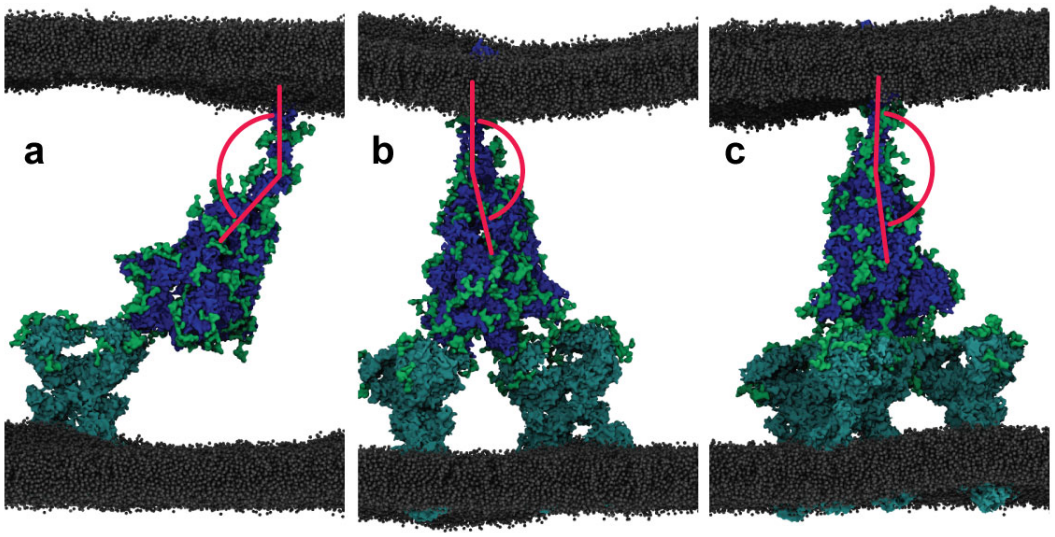


Figure 7.8: The (a) 1S1A, (b) 1S2A and (c) 1S3A systems at 3μs into production with the S protein shown in blue, [ACE2:B^{0}AT1]$_2$ in cyan, membranes in grey, glycans in green and tilt angle shown in red.

7.4.2 Effect of binding stoichiometry on tilt in various cross-linking states

Once I had determined which conformations of the S protein were stable and which measurable properties best defined this, I began to "confine" the protein by binding ACE2 at various RBDs. Upon increasing the number of ACE2 complexes bound to the S protein, the trimer adopted a noticeable change in the tilt angle during simulation. This is obvious from system inspection, as shown in Figure 7.8. This presented as a significant decrease in the adopted tilt angle, which became more evident as the number of ACE2 complexes bound to the trimer was increased (Figure 7.9). This happened because the increased binding of the S protein to ACE2 created multiple points at which the trimer was restrained. This made it significantly more challenging for the protein to tilt as it was restricted from tilting in specific directions due to these binding points. This implied that the S protein may experience stress when tilting as binding points restrict the neck region's mobility in more cross-linked systems. Once bound by two ACE2 receptors, the S protein's ability to tilt was severely impacted.

In the system containing a single S protein, this presented as a noticeable decrease in the average tilt per ACE2 complex added. Substantial decreases in the tilt angle for the single S protein occurred when the first and second ACE2 complexes were bound. The decrease resulting from adding a third ACE2 complex created only a marginal further decrease. The 1S3A system showed less fluctuation than the 1S1A or 1S2A systems, as it was heavily confined. With increased restriction, the S protein could not adopt the preferred conformation due to its inability to tilt easily in any direction. However, the S protein's tendency to adopt an energetically favourable conformation may affect the stability of the interaction between the S protein and ACE2.

The increasing number of ACE2 complexes did not have the same effect on systems containing two S proteins. Whereas increasing the number of ACE2 moieties in the single S system consistently decreased the tilt, in the double S systems, all three ACE2 cluster sizes resulted

in similar tilting. Figure 7.9 shows slight fluctuation in the tilt of the 2S2A system; however, this system saw both the highest and lowest tilt in any S protein for the double S systems. This suggests that the proteins depended on each other's tilt, and a stochastic high tilt in one resulted in a low tilt in the other to compensate. The ACE2 complex appeared also to tilt to accommodate this binding (Figure 7.10). This could, however, be an artefact created by exceptionally strong binding in the CG model, as no systems were seen to dissociate despite obviously extensive stress.

The triple S system was highly fluxional but experienced the highest tilt angle. Although this system was extremely crowded, it should also be noted that it contained two singly-bound S proteins and one double-bound S protein. From what was shown for these binding modes in the 1S and 2S systems, it should only be expected that one of these proteins experiences a substantially lower tilt. With this being said, however, the steric crowding involved in such an extensive system has considerable implications for conformation due to the available space. Early on in the simulation, one of the three S proteins is seen to take a substantially greater tilt angle, approximately 10-20° higher than the other two. Similarly to what was seen in the double S systems, this is likely some form of compensation. The 3S2A system is an extremely large protein complex with multiple facets that contribute to its overall energetics, and as such, there are predictable trade-offs. In this system, a large increase in the tilt of one S protein allows the complex to maintain better binding in the remaining two. A comparison of the triple S cluster is shown in Figure 7.11(a) with the S protein that shows a significantly larger tilt angle coloured in red. For comparison, the 1S1A system is also shown in Figure 7.11(b).

7.4.2.1 Changes in the ankle and knee hinges

As stated previously, work performed by Turoønová *et al.* describes the presence of hinge regions in the neck of the S protein. These same angles were measured by Zuzic *et al.* while simulating the S protein in benzene. To assess the effect of cross-linking on the conformation adopted by the neck region, these angles were measured here for single S systems bound to between 0 and 3 ACE2 receptors. Figure 7.12 shows that as the S protein became cross-linked

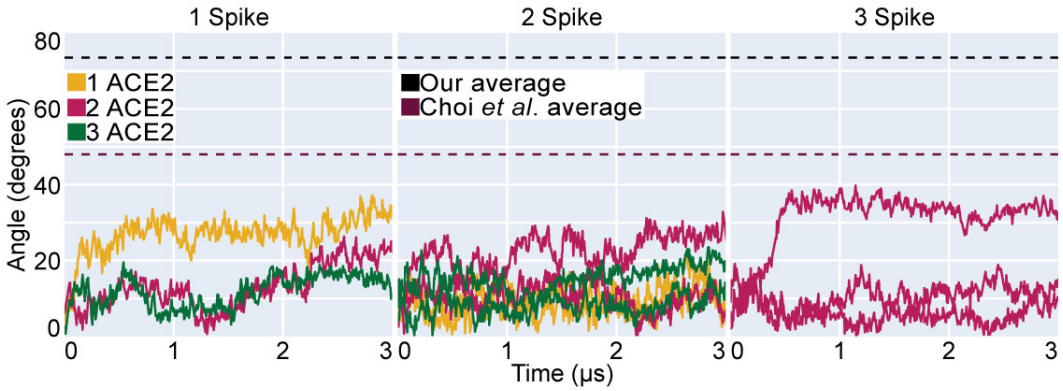


Figure 7.9: Analysis of an individual S protein's tilt angles when bound to 1, 2 and 3 ACE2 complexes in the single S, double S and triple S systems, compared to the average tilt in our unbound system and the average tilt reported by *Choi* et al.

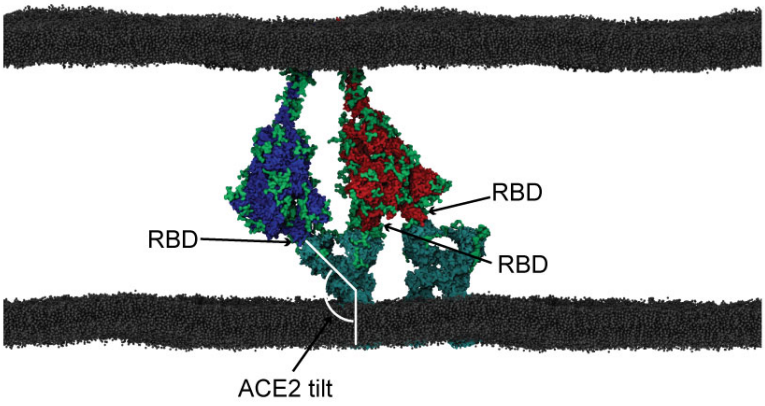


Figure 7.10: The 2S2A system at 3μs into production, showing the S proteins in blue and red, the ACE2 complex in cyan, both membranes in grey and the ACE2 tilt showing in white.

to more ACE2 dimers, the ankle region (Figure 7.12(a)) showed a substantial decrease in angle, dropping as low as 20° when the S protein was bound to 3 ACE2 dimers. This change in the ankle shows a far greater difference between the triply-bound S protein system and the other single S systems. This suggests that the major decrease in tilt of the S protein is largely attributed to conformational changes in the ankle region. Comparatively, the S protein knee remains relatively consistent, averaging out to a similar value in both the 1S0A and 1S3A systems (Figure 7.12(b)).

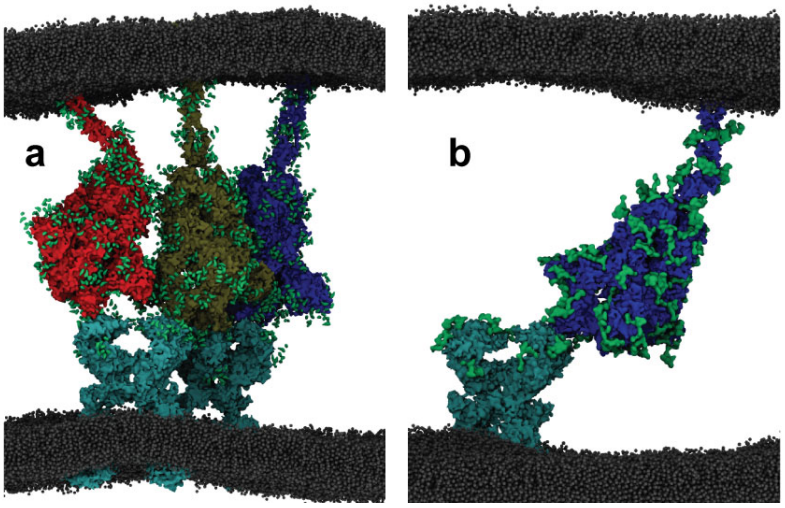


Figure 7.11: Representation of drastically different tilt angles seen. (a) The 3S2A system after the tilt angle in S2 spiked, showing S1 and S3 in blue, S2 in red and ACE2 in cyan. (b) The 1S1A system shows S1 in blue and ACE2 in cyan. Membranes in these systems are coloured by bead type.

7.4.3 Effect of binding stoichiometry on the angle and distance between the receptor binding domain and N-terminal domain in various cross-linking states

Here, I theorise that the stress imposed on the binding region by the tilt in the neck region can detrimentally affect binding. This could have significant implications for the S protein's ability to maintain binding between the RBD and receptor domain of ACE2.

When measuring the RBD-NTD angle, it was observed to deviate significantly from the angle measured for unbound or singly bound structure as coordination increased. These changes can be seen in Figure 7.13. The 1S1A system showed angles for the up- and down-state RBD, which reflect those of the unbound 1S0A system. This is important as the 1S0A system was simulated with no strain, and as such, the angles in that system theoretically reflected the most energetically stable states. This effect changed upon binding a second ACE2 moiety to the S protein, where the RBD angle of both up-state monomers decreased to a value substantially less than that of the unbound system. This was due to the stress induced by tilt in the neck region and obstruction created by ACE2, as was mentioned in Section 7.4.2. Increased coordination meant the degrees of freedom in the S protein were further reduced so that it could tilt in fewer directions. This decreased average tilt because monomers were further restrained. This effect was only furthered in the 1S3A system as the S protein had no direction in which it could tilt and would not clash with ACE2 binding, adversely effecting the RBD-NTD angle. It should be noted that the 1S3A system showed the highest and lowest RBD angles in any single S system, similar to what was seen for the tilt angle. This was a compensation effect, where tilt in one direction adversely affected the RBDs angle but allowed another to extend further. Figure 7.14 shows that while some RBDs were at a relatively acute angle, others took a much larger angle. Figure 7.14(a) shows a rotated view of reduced contact between the RBD and

Figure 7.12: The (a) ankle and (b) knee angles of the S protein neck for single S systems bound to between 0 and 3 ACE2 dimers. The ankle is measured as the angle between the HR2 domain (residue 1157–1207), the hinge region (residue 1208–1212), and the TM domain (residue 1213–1239). The knee is measured as the angle between the HR2 domain, the hinge region (residue 1142–1156) and the ECD (residue 27–1141).

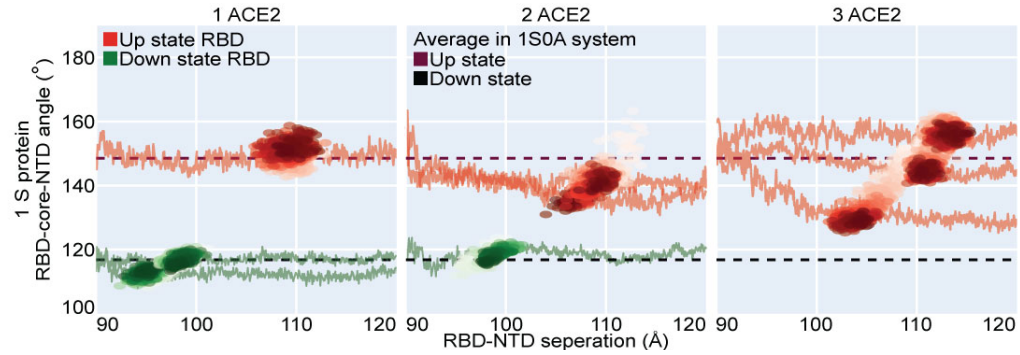


Figure 7.13: Analysis of individual monomers RBD-core-NTD angle vs RBD-NTD separation when S was bound to 1, 2 and 3 ACE2 complexes in the single S systems. A comparison of each monomer RBD-NTD angle throughout production is underplayed against the average angle for a (purple) up and (black) down monomer in the (dotted lines) 1S0A system.

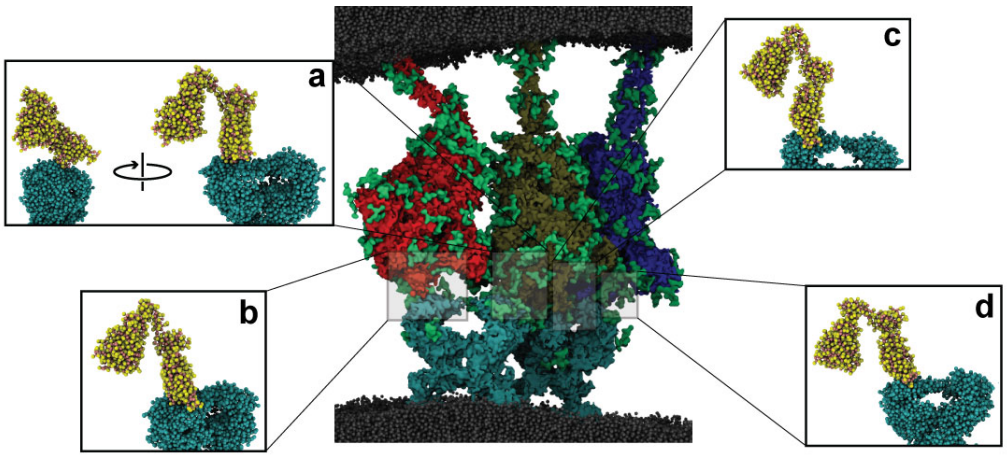


Figure 7.14: A visual representation of the binding in the 3S2A system, showing S proteins in blue, red and tan, ACE2 in cyan, membranes in grey and glycans in green. Zoomed-in representations of each RBD binding at 3μs are shown in a-d.

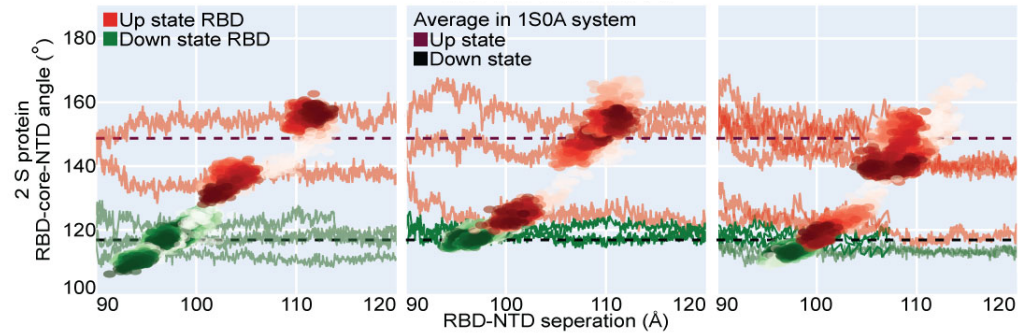


Figure 7.15: Analysis of individual monomers RBD-core-NTD angle vs RBD-NTD separation when S was bound to 1, 2 and 3 ACE2 complexes in the double S systems. A comparison of each monomer RBD-NTD angle throughout production is underlayed against the average angle for a (purple) up and (black) down monomer in the (dotted lines) 1S0A system.

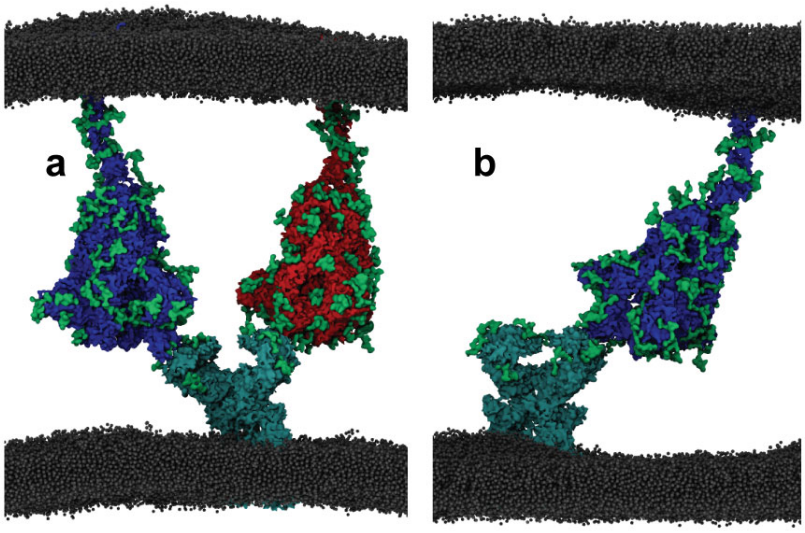


Figure 7.16: A visual representation of the (a) 2S1A system and the (b) 1S1A system, showing the S protein in blue and red, ACE2 in cyan, membranes in grey and glycans in green.

ACE2. This shows that the RBD and ACE2 may begin disrupting contacts to relieve stress in the S protein once significantly crowded.

In the double S systems (Figure 7.15), the RBD angles became less definitively up or down. Even in the 2S1A system, the up RBD angles on either trimer deviated from those of the unbound 1S0A system. This was unexpected, as neither S protein was coordinated to more than one ACE2. Because of this, it would be logical to assume that each S protein would have reflected the binding of the 1S1A system. This was not the case due to the dynamics of the receptor complex. The ACE2 complex was seen to tilt in the single S systems toward the point of contact (Figure 7.16). In double S systems, it could not tilt toward one trimer without adversely affecting its binding to the other. This reorientation of ACE2 was also observed in the 1S1A system (Figure 7.16(b)). The 2S1A system showed that one trimer had an increased angle as the ACE2 complex tilted beneficially, while the other decreased for the same reason. The effect of increasing the size of the ACE2 cluster then began to push RBDs out of the up-state and into a down-state. In the 2S2A and 2S3A systems, one RBD was seen to move to an angle that reflected the down-state. This was due to the increased strain resulting from many S proteins trying to tilt while also restrained at multiple points. Adding a second S protein further affected this as it created further steric hindrance, and the receptors to which the S proteins were bound could not reorient themselves. This effect in the 2S2A system showed an initial compensation that increased the angle in one bound RBD before the two bound RBDs evened out to a mutually beneficial state. This was also seen in the 2S3A system, but it happened much faster due to the increased restrictions.

The 3S2A system (Figure 7.17) again showed similar effects to the 2S2A and 2S3A systems. The degree of restriction in the system resulted in two RBDs in a beneficial position while the other two entered a pseudo-down-state across all S proteins in the system.

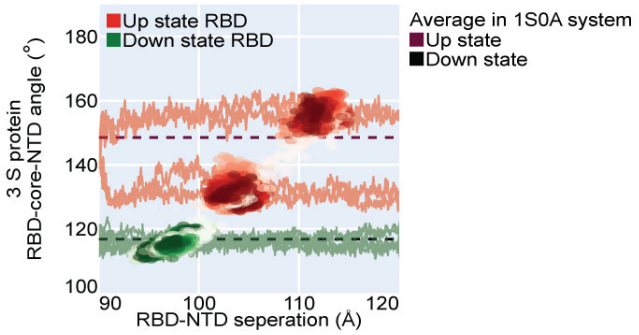


Figure 7.17: Analysis of individual monomers RBD-core-NTD angle vs RBD-NTD separation when S was bound to 2 ACE2 complexes in the triple S system. A comparison of each monomer RBD-NTD angle throughout production is underlayed against the average angle for a (purple) up and (black) down monomer in the (dotted lines) 1S0A system.

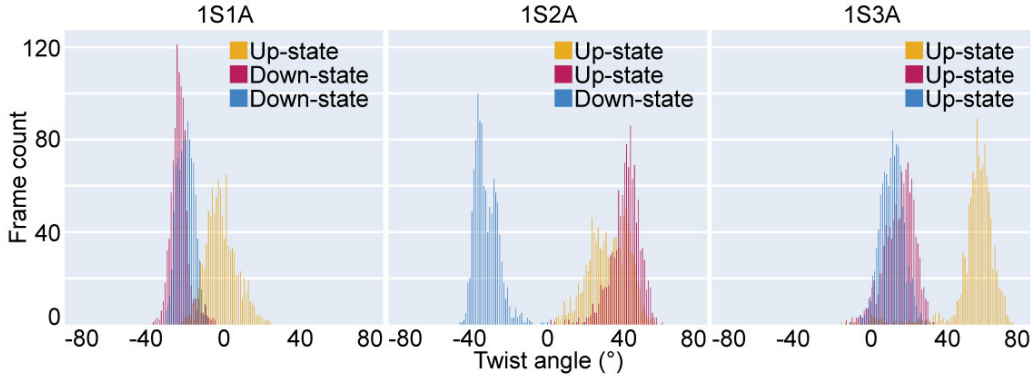


Figure 7.18: Count of frames for which the RBD was seen to take twist angles between -80° and 80° for the single S systems.

7.4.4 Effect of binding stoichiometry on the rigid body rotation of the receptor binding domain

Previously, in Section 7.4.1, it was mentioned that the RBD of the S protein is known to undergo a rigid body rotation upon binding to ACE2, as described by Choi *et al*. In my simulations, the single S systems showed a dramatic change in the twist of the RBD as the coordination of ACE2 increased from one to three.

Strain-induced in the protein complex can be seen to create large deviations in the twist angle of the RBD for a single S system, as shown in Figure 7.18. The 1S1A showed that down-state RBDs had a twist of approximately -30°, while the up-state monomer had an average twist of 0°. This difference is accentuated in the 1S2A system, where the increased restriction causes a larger discrepancy in the twist between these monomers; however, the twist in the down-state monomer does not show significant deviation from the 1S1A system. Binding 3 ACE2s caused all monomers to take a positive twist angle and one to take a significantly larger twist angle. This was likely due to the twist acting as a stabilising factor, allowing the S protein to move while bound without becoming unbound from ACE2. As ACE2 binding increased, this twisting became more significant as it was necessary to maintain binding.

7.4.5 Structural analysis of the S protein

To assess the effect of stress in the cross-linked protein complex, the 1S0A and 1S3A system was rerun without an elastic network. This creates an underrepresentation of the secondary structure and can lead to deformations in regions under stress. In simulations where an elastic network and disulphide bonds were removed, the RBD of the S protein was seen to buckle and deform entirely in high-order simulations (Figure 7.19). The stress mentioned above is imparted upon the protein complex by conformational changes clashing with restrictions caused by multidentate binding. This emphasises the stress imposed upon the RBD regions as the underrepresentation of the secondary structure resulted in the entire deformation of the RBD. For comparison, the 1S0A system was also simulated without an elastic network and showed minimal change in the ECD (Figure 7.19(a)).

RMSD analysis was performed for the NTD, RBD, FP, HR1, HR2 and TM domains in systems with an elastic netowrk to determine the effect of high-order binding on the structure of the S protein. Figure 7.20 shows that all domains showed significant deviation throughout production, especially the HR2 and TM domains. The NTD and HR1 domains showed the slightest deviation as the NTD is generally not involved in binding, and the HR1 domain is relatively removed from the structure, being capped by the ECD. The RBD shows apparent deviation in the doubly and triply bound state because of the strain mentioned above, which is transferred to the RBD by tilting in the neck region. Although the FP domain is not responsible for S protein-ACE2 binding, it is located nearby and may be partially deformed when the S proteins strain to accommodate high-order binding. The most significant deviation is the HR2 and TM domains, as the HR2 domain is located in the neck, while the TM domain is adjacent and liable to move while the protein tilts. Many domains show an apparent step in deviation from the initial structure when binding increases to two ACE2s, with the RMSD of the FP domain appearing to step between the two states at 2-3μs.

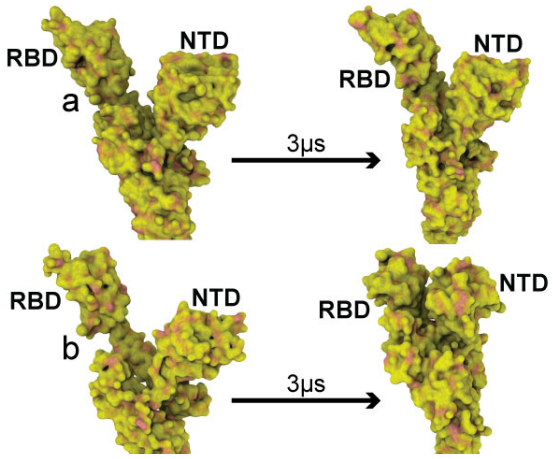


Figure 7.19: The ECD of an up-state monomer in the (a) 1S0A system and (b) 1S3A system after 3μ s of production without an elastic network. Showing the protein backbone in red and amino acid side chains in yellow.

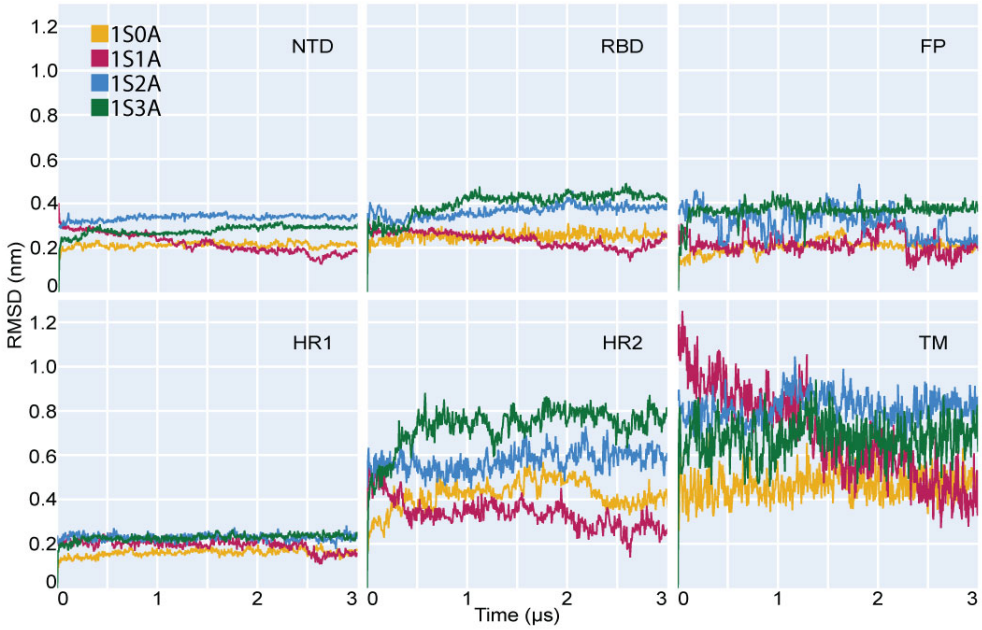


Figure 7.20: The RMSD of domains in the S protein when bound to 0 to 3 membrane-bound ACE2 receptor proteins. RMSD was calculated with the least squares fitted to the respective protein backbone of said domain.

7.5 Conclusions and future perspective

In this study, I examined the conformational requirements of the S protein when cross-linked in high-order systems. Our simulations showed that increased coordination of ACE2 resulted in a decrease in the adopted tilt angle of the S protein due to increased binding points that restricted the protein from tilting in specific directions. This effect was most significant in the highly crowded triple S system, where one S protein would take a significantly greater tilt angle.

Furthermore, I observed that the stress imposed on the binding region by the neck region may detrimentally affect the binding and stability of the RBD region. As binding increased in the single S systems, the RBD angle deviated significantly from the unbound system. In contrast, RBD angles became less definitively up or down in the double and triple S systems. This showed that deviations in the optimal conformation were not limited to the trimer's neck region. Generally, there was little change to RBDs in the down position. This may have severe implications for the S protein's ability to maintain binding to the ACE2 receptor when bound at multiple points.

Overall, this study provides insights into the conformational requirements of the S protein and its response to high-order receptor binding, which could aid in developing effective treatments for SARS-CoV-2 or similar S protein-dependent viral pathogens. In particular, our findings provide a foundation for further research to explore and develop innovative approaches for exploiting weak points in the SARS-CoV-2 binding mechanism.

The future perspective of this work would be to explore more intricate aspects of this cross-linking. For instance, AA simulations of the 1S1A and 1S2A systems may provide valuable

insight into the exact conformational requirements of high-order binding. This would also explain why higher coordination is generally less favourable. Further exploring the difference between the ancestral wild type and omicron variants may provide insight into why some variants are significantly more virulent. In future work, replication of all wild type systems with the omicron variant would provide insight into the differences between the two variants; however, this was not performed due to time constraints. Further, adding sanitisers would provide insight into whether stress created by high-order cross-linking makes binding liable to disruption.

Chapter 8

Concluding remarks and future perspective

This thesis has undertaken a multifaceted exploration of bacterial and viral sanitization, combining computational simulations and experimental NMR studies to gain insights into the complex mechanisms governing the interactions between sanitizing agents and microbial structures. The investigation covers multiple scales, shedding light on the intricate details of the sanitization process and its implications for infection control and public health. These studies broadly aim to further our understanding of antimicrobial interactions with pathogens and provide insight into potential ways in which we can engineer new antimicrobials.

The first two results chapters focus on bacterial membranes and their interactions with the sanitizing agent CHX. The simulations reveal CHX's interactions with bacterial lipid bilayers, where it binds in a c-shape conformation. Although this is integral to its ability to act as an effective antimicrobial, it is also seen to reinforce the stability of *S. aureus* and *E. coli* membranes. This work underscores the importance of specific, initial interactions of sanitisers and aims to quantify their effects on the lipid bilayer. Analysis of CHX's binding to bacterial membranes through solid-state NMR techniques echoes the computational results, validating my findings and furthering our understanding of the binding conformation of CHX. This and the associated simulation section provide a detailed study of CHX as an antimicrobial. By applying MD and NMR together, I provide visual context to the NMR studies and validate MD results.

The third section studies the impact of short-chain alcohols on the S protein. Simulations reveal how alcohol-induced deformation transforms protein structures and identifies critical structural features exploited by short-chained alcohols in partitioning and deforming said protein. This research elucidates the dynamic interplay between protein conformations and alcohol concentrations by applying machine learning to predict areas that contact the sanitiser environment more. By analysing the feature importance of this prediction, I reveal complex trends involving multiple properties that determine the partitioning points of alcohol into the protein structure.

The final section focuses on high-order receptor binding between the S protein and ACE2.

Simulations unveil the nuanced adjustments in S protein tilt angles with increased ACE2 receptor coordination, bearing significant consequences for binding and stability. These studies aim to detail the S protein-ACE2 binding mechanism. Exploration of stress imparted upon the larger protein complex by restrictions in binding provides insight into potential flaws in the cross-linking interaction which the infection mechanism of SARS-CoV-2 is reliant upon.

Overall, this thesis provides work detailing the multifaceted nature of sanitisation, spanning bacterial membranes and viral structures. While each section offers valuable insights into specific facets of sanitization, overarching themes emerge. Specific hydrophilic-hydrophobic interfaces are pivotal in the binding of sanitisers, be it in membrane composition or viral protein structure. However, further exploration into these studies may provide further insight critical to the design of effective sanitising agents. Incorporating more complex membrane models, extended simulation timeframes, and experimental validation of findings stand as logical next steps. These studies would refine our understanding of sanitization processes, enhancing our ability to combat evolving pathogens and antimicrobial resistance.

In conclusion, this thesis embodies a comprehensive study of the intricacies of bacterial and viral sanitization. By combining insights from diverse scales and disciplines, I provide exploratory works that may provide a basis for creating effective sanitisers in the future of infection control and public health.

Bibliography

[1] J. Sapp. "The Prokaryote-Eukaryote Dichotomy: Meanings and Mythology". In: *Microbiology and Molecular Biology Reviews* 69 (2 2005), pp. 292–305. ISSN: 1092-2172. DOI: 10.1128/mmbr.69.2.292-305.2005.

[2] A. Oren. *Prokaryote diversity and taxonomy: Current status and future challenges*. Apr. 2004. DOI: 10.1098/rstb.2003.1458.

[3] W. Lucas. "Viral Capsids and Envelopes: Structure and Function". In: *eLS* (Apr. 2010). DOI: 10.1002/9780470015902.a0001091.pub2. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/9780470015902.a0001091.pub2.

[4] H. Nikaido. "Structure and Functions of the Cell Envelope of Gram-Negative Bacteria". In: *Reviews of Infectious Diseases* 10 (1988), S279–S281. ISSN: 01620886.

[5] M. J. Osborn et al. "Structure and Biogenesis of the Cell Envelope of Gram-Negative Bacteria". In: *Ann N Y Acad Sci* 10 Suppl 2 (1988), pp. 52–65. ISSN: 01620886. DOI: 10.1111/j.1749-6632.1974.tb43256.x.

[6] M. Rajagopal and S. Walker. "Envelope Structures of Gram-Positive Bacteria". In: *Curr Top Microbiol Immunol* 404 (2017), pp. 1–44. DOI: 10.1007/82_2015_5021.

[7] O. G. Mouritsen and M. J. Zuckermann. "What's So Special About Cholesterol?" In: *Lipids* 39 (11 2004), pp. 1101–1113. DOI: 10.1007/s11745-004-1336-x.

[8] J. E. Vance. "Cellular itinerary of LDL cholesterol". In: *Proceedings of the National Academy of Sciences of the United States of America* 119 (6 2022), pp. 1–3. ISSN: 10916490. DOI: 10.1073/pnas.2122584119.

[9] T. Silhavy, D. Kahne, and S. Walker. *The Bacterial Cell Envelope*. 2010. DOI: 10.1101/cshperspect.a000414.

[10] D. Segré et al. "The Lipid World". In: *Orig Life Evol Biosph* 31 (1-2 2001), pp. 119–145. ISSN: 01696149. DOI: 10.1023/A:1006746807104.

[11] C. Sohlenkamp and O. Geiger. "Bacterial membrane lipids: Diversity in structures and pathways". In: *FEMS Microbiol Rev* 40 (1 2016), pp. 133–159. ISSN: 15746976. DOI: 10.1093/femsre/fuv008.

[12] P. H. Rampelotto. "Extremophiles and extreme environments". In: *Life* 3 (2013), pp. 482–485. ISSN: 20751729. DOI: 10.3390/life3030482.

[13] F. Balloux and L. van Dorp. "Q and A: What are pathogens, and what have they done to and for us?" In: *BMC Biology* 15 (91 2017). ISSN: 17417007. DOI: 10.1186/s12915-017-0433-z.

[14] K. Young. “Bacterial morphology: why have different shapes?” In: *Curr Opin Microbiol* 10 (6 2007), pp. 596–600. ISSN: 13695274. DOI: 10.1016/j.mib.2007.09.009.

[15] M. R. Salton. “The bacterial cell envelope - a historical perspective”. In: *New Comprehensive Biochemistry* 27 (1994), pp. 1–22. ISSN: 01677306. DOI: 10.1016/S0167-7306(08)60404-4.

[16] K. Reglinski et al. “Fluidity and Lipid Composition of Membranes of Peroxisomes, Mitochondria and the ER From Oleic Acid-Induced Saccharomyces cerevisiae”. In: *Front Cell Dev Biol* 8 (574363 2020). ISSN: 2296634X. DOI: 10.3389/fcell.2020.574363.

[17] R. Koebnik, K. P. Locher, and P. V. Gelder. “Structure and function of bacterial outer membrane proteins: Barrels in a nutshell”. In: *Mol Microbiol* 37 (2 2000), pp. 239–253. ISSN: 0950382X. DOI: 10.1046/j.1365-2958.2000.01983.x.

[18] S. Soto. “Role of efflux pumps in the antibiotic resistance of bacteria embedded in a biofilm”. In: *Virulence* 4 (3 2013), pp. 223–229.

[19] P. Gilbert, D. G. Allison, and A. J. McBain. “Biofilms in vitro and in vivo: Do singular mechanisms imply cross-resistance?” In: *J Appl Microbiol* 92 (2002), 98S–110S. ISSN: 02674440. DOI: 10.1046/j.1365-2672.92.5s1.5.x.

[20] J. Zhao et al. “Assembly and Functional Role of PACE Transporter PA2880 from Pseudomonas aeruginosa”. In: *Microbiology Spectrum* 10 (2 2022), pp. 1–12. ISSN: 21650497. DOI: 10.1128/spectrum.01453-21.

[21] P. M and P. K. “Peptidoglycan”. In: *Subcell Biochem* 92 (2019), pp. 127–168.

[22] C. Gram. “The differential staining of Schizomycetes in tissue sections and in dried preparations”. In: *Gen Microbiol* 2 (1884), pp. 216–218.

[23] R. Coico. “Gram staining”. In: *Curr Protoc Microbiol* (2005). DOI: 10.1002/9780471729259.mca03cs00.

[24] S. Baron. “Structure”. In: ed. by S. Baron. 4th ed. Vol. 4. University of Texas Medical Branch at Galveston, 1996.

[25] L. Pasquina-Lemonche et al. “The Architecture of the Gram Positive Bacterial Cell Wall L.” In: *Nature* 582 (7811 2020), pp. 294–297. DOI: 10.1038/s41586-020-2236-6.The.

[26] T. N and S. A. “Gram Staining”. In: *Gram Staining* (2021), pp. 1–5. URL: https://www.ncbi.nlm.nih.gov/books/NBK562156/.

[27] S. Miller and N. Salama. “The gram-negative bacterial periplasm: Size matters”. In: *PLoS Biology* 16 (1 2018). ISSN: 15457885. DOI: 10.1371/journal.pbio.2004935.

[28] M. Papanastasiou et al. “The escherichia coli peripheral inner membrane proteome”. In: *Mol Cell Proteomics* 12 (3 Mar. 2013), pp. 599–610. ISSN: 15359476. DOI: 10.1074/mcp.M112.024711.

[29] M. Bogdanov et al. “Phospholipid distribution in the cytoplasmic membrane of Gram-negative bacteria is highly asymmetric, dynamic, and cell shape-dependent”. In: *Sci Adv* 6 (23 2020). ISSN: 23752548. DOI: 10.1126/sciadv.aaz6333.

[30] J. Gebhardt et al. “Combining molecular dynamics and machine learning to predict self-solvation free energies and limiting activity coefficients”. In: *J Chem Inf Model* 60 (11 2020), pp. 5319–5330. ISSN: 1549960X. DOI: 10.1021/acs.jcim.0c00479.

[31] P. A. Levin and E. R. Angert. “Small but mighty: Cell size and bacteria”. In: *Cold Spring Harbor Perspectives in Biology* 7 (7 2015). ISSN: 19430264. DOI: 10.1101/cshperspect.a019216.

[32] J. Louten. “Virus Structure and Classification”. In: *Essential Human Virology* (2016), pp. 19–29. DOI: 10.1016/B978-0-12-800947-5.00002-8.

[33] H. R. Gelderblom. “Structure and Classification of Viruses”. In: ed. by S. Baron. 4th ed. University of Texas Medical Branch at Galveston, 1996.

[34] V. C.-C. Cheng et al. “Viral Infections, an Overview with a Focus on Prevention of Transmission”. In: *Reference Module in Biomedical Sciences* (2016). DOI: 10.1016/B978-0-12-801238-3.90174-0.

[35] J. Lenard. “Viral Membranes”. In: *Encyclopedia of Virology* (2008), pp. 308–314.

[36] X. Zhang et al. “Structure and function of capsid protein in flavivirus infection and its applications in the development of vaccines and therapeutics”. In: *Veterinary research* 52 (1 2021), p. 98. ISSN: 12979716. DOI: 10.1186/s13567-021-00966-2. URL: https://doi.org/10.1186/s13567-021-00966-2.

[37] P. E. Pellett, S. Mitra, and T. C. Holland. “Basics of virology”. In: *Handb Clin Neurol* 123 (2014), pp. 123–145. ISSN: 00729752. DOI: 10.1016/B978-0-444-53488-0.00002-X. URL: http://dx.doi.org/10.1016/B978-0-444-53488-0.00002-X.

[38] R. A. Simmons, C. B. Willberg, and K. Paul. “Immune Evasion by Viruses”. In: *eLS* (2013). DOI: 10.1002/9780470015902.a0024790.

[39] R. Marshall. “Glycoproteins”. In: *Annu. Rev. Biochem* 41 (34 1972), pp. 673–702.

[40] S. Moens and J. Vanderleyden. “Glycoproteins in prokaryotes”. In: *Archives of Microbiology* 168 (3 1997), pp. 169–175. ISSN: 03028933. DOI: 10.1007/s002030050484.

[41] L. Casalino et al. “Beyond shielding: The roles of glycans in the SARS-CoV-2 spike protein”. In: *ACS Central Science* 6 (10 2020), pp. 1722–1734. ISSN: 23747951. DOI: 10.1021/acscentsci.0c01056.

[42] J. Lan et al. “Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor”. In: *Nature* 581 (7807 2020), pp. 215–220. ISSN: 14764687. DOI: 10.1038/s41586-020-2180-5. URL: http://dx.doi.org/10.1038/s41586-020-2180-5.

[43] N. Banerjee and S. Mukhopadhyay. “Viral glycoproteins: biological role and application in diagnosis”. In: *VirusDisease* 27 (1 2016), pp. 1–11. ISSN: 23473517. DOI: 10.1007/s13337-015-0293-5.

[44] K. Candido et al. “Spike protein of SARS-CoV-2 variants: a brief review and practical implications”. In: *Brazilian Journal of Microbiology* 53 (3 2022), pp. 1133–1157. ISSN: 16784405. DOI: 10.1007/s42770-022-00743-z.

[45] S. Harrison. “Viral membrane fusion”. In: *Nature Structural and Molecular Biology* 15 (7 2008), pp. 690–698. ISSN: 15459993. DOI: 10.1038/nsmb.1456.

[46] A. Cho et al. “Anti-SARS-CoV-2 receptor-binding domain antibody evolution after mRNA vaccination”. In: *Nature* 600 (7889 2021), pp. 517–522. ISSN: 14764687. DOI: 10.1038/s41586-021-04060-7.

[47] S. Payne. “Family Coronaviridae”. In: *Viruses* (2017), pp. 149–158. ISSN: 2338-9567.

[48] J. Li et al. “The emergence, genomic diversity and global spread of SARS-CoV-2”. In: *Nature* 600 (7889 2021), pp. 408–418. ISSN: 14764687. DOI: 10.1038/s41586-021-04188-6.

[49] Z. Zhang et al. “Structure of SARS-CoV-2 membrane protein essential for virus assembly”. In: *Nature Communications* 13 (1 2022), pp. 1–7. ISSN: 20411723. DOI: 10.1038/s41467-022-32019-3.

[50] D. Schoeman and B. C. Fielding. “Coronavirus envelope protein: current knowledge”. In: *Cell Research* 9 (1 2019), pp. 278–280. ISSN: 0022-538X. DOI: 10.3390/ijerph17082932.

[51] X. Li et al. “Spike protein mediated membrane fusion during SARS-CoV-2 infection”. In: *Journal of Medical Virology* 95 (1 2023), pp. 1–12. ISSN: 10969071. DOI: 10.1002/jmv.28212.

[52] C. B. Jackson et al. “Mechanisms of SARS-CoV-2 entry into cells”. In: *Nature Reviews Molecular Cell Biology* 23 (1 2022), pp. 3–20. ISSN: 14710080. DOI: 10.1038/s41580-021-00418-x.

[53] M. Pizzato et al. “SARS-CoV-2 and the Host Cell: A Tale of Interactions”. In: *Frontiers in Virology* 1 (January 2022), pp. 1–29. DOI: 10.3389/fviro.2021.815388.

[54] P. V’kovski et al. “Coronavirus biology and replication: implications for SARS-CoV-2”. In: *Nature Reviews Microbiology* 19 (3 2021), pp. 155–170. ISSN: 17401534. DOI: 10.1038/s41579-020-00468-6. URL: http://dx.doi.org/10.1038/s41579-020-00468-6.

[55] J. Chen et al. “Competition of SARS-CoV-2 variants on the pandemic transmission dynamics”. In: *Chaos Solitons Fractals.* 169 (113193 2023).

[56] W. T. Harvey et al. “SARS-CoV-2 variants, spike mutations and immune escape”. In: *Nature Reviews Microbiology* 19 (7 2021), pp. 409–424. ISSN: 17401534. DOI: 10.1038/s41579-021-00573-0. URL: http://dx.doi.org/10.1038/s41579-021-00573-0.

[57] D. D. Farhud and N. Mojahed. “SARS-COV-2 Notable Mutations and Variants: A Review Article”. In: *Iranian Journal of Public Health* 51 (7 2022), pp. 1494–1501. ISSN: 22516093. DOI: 10.18502/ijph.v51i7.10083.

[58] B. Alberts, A. Johnson, and J. Lewis. “The Lipid Bilayer”. In: ed. by A. B, J. A, and L. J. 4th ed. Garland Science, 2002. URL: https://www.ncbi.nlm.nih.gov/books/NBK26871/.

[59] D. Casares, P. V. Escribá, and C. A. Rosselló. “Membrane lipid composition: Effect on membrane and organelle structure, function and compartmentalization and therapeutic avenues”. In: *International Journal of Molecular Sciences* 20 (9 2019). ISSN: 14220067. DOI: 10.3390/ijms20092167.

[60] N. A. Alhakamy et al. “Effect of Lipid Headgroup Charge and pH on the Stability and Membrane Insertion Potential of Calcium Condensed Gene Complexes”. In: *American Chemical Society* 14 (31 2017), pp. 4232–4245. DOI: 10.1021/la504970n.Effect.

[61] S. Pal et al. “Effect of Phospholipid Headgroup Charge on the Structure and Dynamics of Water at the Membrane Interface: A Terahertz Spectroscopic Study”. In: *Journal of Physical Chemistry B* 122 (19 2018), pp. 5066–5074. ISSN: 15205207. DOI: 10.1021/acs.jpcb.8b01633.

[62] E. Kilelee et al. “Lysyl-Phosphatidylglycerol Attenuates Membrane Perturbation Rather than Surface Association of the Cationic Antimicrobial Peptide 6W-RP-1 in a Model Membrane System: Implications for Daptomycin Resistance”. In: *Antimicrobial Agents and Chemotherapy* 54 (10 2010), pp. 4476–4479. ISSN: 00664804. DOI: 10.1128/AAC.00191-10.

[63] K. J. Seu et al. “Influence of lipid chemistry on membrane fluidity: Tail and headgroup interactions”. In: *Biophysical Journal* 91 (10 2006), pp. 3727–3735. ISSN: 00063495. DOI: 10.1529/biophysj.106.084590.

[64] J. Wang et al. “Functions of Sphingolipids in Pathogenesis During Host–Pathogen Interactions”. In: *Frontiers in Microbiology* 12 (701041 2021). ISSN: 1664302X. DOI: 10.3389/fmicb.2021.701041.

[65] E. Dufourc. “Sterols and membrane dynamics”. In: *J Chem Biol* 1 (1-4 2008), pp. 63–77. DOI: 10.1007/s12154-008-0010-6.

[66] S. Kim et al. “Bilayer Properties of Lipid A from Various Gram-Negative Bacteria”. In: *Biophysical Journal* 111 (8 2016), pp. 1750–1760. ISSN: 15420086. DOI: 10.1016/j.bpj.2016.09.001.

[67] G. Zhang, T. C. Meredith, and D. Kahne. “On the essentiality of lipopolysaccharide to Gram-negative bacteria”. In: *Current Opinion in Microbiology* 16 (6 2013), pp. 779–785. ISSN: 13695274. DOI: 10.1016/j.mib.2013.09.007. URL: http://dx.doi.org/10.1016/j.mib.2013.09.007.

[68] E. L. Wu et al. “Molecular dynamics and NMR spectroscopy studies of E. coli lipopolysaccharide structure and dynamics”. In: *Biophysical Journal* 105 (6 2013), pp. 1444–1455. ISSN: 00063495. DOI: 10.1016/j.bpj.2013.08.002.

[69] T. J. Beveridge. “Structures of gram-negative cell walls and their derived membrane vesicles”. In: *Journal of Bacteriology* 181 (16 1999), pp. 4725–4733. ISSN: 00219193. DOI: 10.1128/jb.181.16.4725-4733.1999.

[70] P. C. Hsu et al. “It Is Complicated: Curvature, Diffusion, and Lipid Sorting within the Two Membranes of Escherichia coli”. In: *Journal of Physical Chemistry Letters* 8 (22 Nov. 2017), pp. 5513–5518. ISSN: 19487185. DOI: 10.1021/acs.jpclett.7b02432.

[71] A. Steimle, I. B. Autenrieth, and J. S. Frick. “Structure and function: Lipid A modifications in commensals and pathogens”. In: *International Journal of Medical Microbiology* 306 (5 2016), pp. 290–301. ISSN: 16180607. DOI: 10.1016/j.ijmm.2016.03.001. URL: http://dx.doi.org/10.1016/j.ijmm.2016.03.001.

[72] C. Raetz et al. “Lipid A Modification Systems in Gram-Negative Bacteria”. In: *October* 76 (2007), pp. 295–329. ISSN: 08966273. DOI: 10.1146/annurev.biochem.76.010307.145803.LIPID.

[73] C. R. H. Raetz and C. Whitfield. “Lipopolysaccharide Endotoxins Endotoxins as Activators of Innate Immunity”. In: *Annual review of biochemistry* 71 (2008), pp. 635–700. DOI: 10.1146/annurev.biochem.71.110601.135414.Lipopolysaccharide.

[74] M. Bagheri, S. Keller, and M. Dathe. “Interaction of W-substituted analogs of cyclo-RRRWFW with bacterial lipopolysaccharides: The role of the aromatic cluster in antimicrobial activity”. In: *Antimicrobial Agents and Chemotherapy* 55 (2 2011), pp. 788–797. ISSN: 00664804. DOI: 10.1128/AAC.01098-10.

[75] P. R. Reeves. “Role of O-antigenvariation in the immune response”. In: *Trends in Microbiology* 3 (10 1995), pp. 381–386. URL: papers3://publication/uuid/59CD5B47-D329-4934-B613-F7BB085FB262.

[76] L. Sun, F. Pan, and S. Li. “Self-Assembly of Lipid Mixtures in Solutions: Structures, Dynamics Processes and Mechanical Properties”. In: *Membranes* 12 (8 2022). ISSN: 20770375. DOI: 10.3390/membranes12080730.

[77] H. Nikaido. “Multidrug resistance in bacteria”. In: *Annual Review of Biochemistry* 78 (2009), pp. 119–146. ISSN: 00664154. DOI: 10.1146/annurev.biochem.78.082907.145923.

[78] T. Dörr, P. J. Moynihan, and C. Mayer. “Editorial: Bacterial Cell Wall Structure and Dynamics”. In: *Frontiers in Microbiology* 10 (September 2019), pp. 4–7. ISSN: 1664302X. DOI: 10.3389/fmicb.2019.02051.

[79] L. Bessa, M. Ferreira, and P. Gameiro. “Evaluation of membrane fluidity of multidrug-resistant isolates of Escherichia coli and Staphylococcus aureus in presence and absence of antibiotics”. In: *Journal of Photochemistry and Photobiology B: Biology* 181 (2018), pp. 150–156. ISSN: 18732682. DOI: 10.1016/j.jphotobiol.2018.03.002. URL: https://doi.org/10.1016/j.jphotobiol.2018.03.002.

[80] J. Wang et al. “Core oligosaccharide portion of lipopolysaccharide plays important roles in multiple antibiotic resistance in Escherichia coli”. In: *Antimicrobial Agents and Chemotherapy* 65 (10 2021), pp. 1–14. ISSN: 10986596. DOI: 10.1128/AAC.00341-21.

[81] V. W. Rowlett et al. “Impact of membrane phospholipid alterations in Escherichia coli on cellular function and bacterial stress adaptation”. In: *Journal of Bacteriology* 199 (13 2017), pp. 1–22. ISSN: 0021-9193. URL: http://jb.asm.org/content/199/13/e00849-16.full.pdf%20precent0Ahttp://jb.asm.org/.

[82] S. Salinas-Almaguer et al. “Membrane rigidity regulates E. coli proliferation rates”. In: *Scientific Reports* 12 (933 2022). ISSN: 20452322. DOI: 10.1038/s41598-022-04970-0. URL: https://doi.org/10.1038/s41598-022-04970-0.

[83] C. Haest et al. “Changes in permeability of Staphylococcus aureus and derived liposomes with varying lipid composition”. In: *Biochimica et Biophysica Acta - Biomembranes* 255 (3 1972), pp. 720–733.

[84] R. M. Gould and W. J. Lennarz. "Metabolism of Phosphatidylglycerol and Lysyl Phosphatidylglycerol in Staphylococcus aureus". In: *Journal of Bacteriology* 104 (3 1970), pp. 1135–1144. ISSN: 0021-9193. DOI: 10.1128/jb.104.3.1135-1144.1970.

[85] O. Akanbi et al. "Antimicrobial susceptibility of Staphylococcus aureus isolated from recreational waters and beach sand in Eastern Cape Province of South Africa". In: *International Journal of Environmental Research and Public Health* 14 (9 2017), p. 1001. ISSN: 16604601. DOI: 10.3390/ijerph14091001.

[86] K.-S. Lim and P. C. A. Kam. "Chlorhexidine – pharmacology and clinical applications". In: *Anaesth Intensive Care* 36 (2008), pp. 502–512.

[87] B. P. Gomes et al. "Chlorhexidine in endodontics". In: *Brazilian Dental Journal* 24 (2 Mar. 2013), pp. 89–102. ISSN: 01036440. DOI: 10.1590/0103-6440201302188.

[88] T. Pattananandecha et al. "Pharmaceutical Incompatibility of Lubricating Gel Formulation Reduces Antibacterial Activity of Chlorhexidine Gluconate: In Vitro Study in Northern Thailand". In: *International Journal of Environmental Research and Public Health* 19 (19 2022). ISSN: 16604601. DOI: 10.3390/ijerph191912285.

[89] A. D. Russell and M. J. Day. "Antibacterial activity of chlorhexidine". In: *Journal of Hospital Infection* 25 (4 1993), pp. 229–238. ISSN: 01956701. DOI: 10.1016/0195-6701(93)90109-D.

[90] D. Adams et al. "Evaluation of a 2 precent chlorhexidine gluconate in 70 precent isopropyl alcohol skin disinfectant". In: *Journal of Hospital Infection* 61 (4 2005), pp. 287–290. ISSN: 01956701. DOI: 10.1016/j.jhin.2005.05.015.

[91] L. O. Ingram. "Mechanism of lysis of Escherichia coli by ethanol and other chaotropic agents". In: *Journal of Bacteriology* 146 (1 1981), pp. 331–336. ISSN: 00219193. DOI: 10.1128/jb.146.1.331-336.1981.

[92] M. E. Tóth, L. Vígh, and M. Sántha. "Alcohol stress, membranes, and chaperones". In: *Cell Stress and Chaperones* 19 (3 2014), pp. 299–309. ISSN: 14661268. DOI: 10.1007/s12192-013-0472-5.

[93] I. Komljenović et al. "Location of chlorhexidine in DMPC model membranes: A neutron diffraction study". In: *Chemistry and Physics of Lipids* 163 (6 2010), pp. 480–487. ISSN: 00093084. DOI: 10.1016/j.chemphyslip.2010.03.007.

[94] B. V. Oosten, D. Marquardt, and T. A. Harroun. "Testing High Concentrations of Membrane Active Antibiotic Chlorhexidine Via Computational Titration and Calorimetry". In: *Journal of Physical Chemistry B* 121 (18 2017), pp. 4657–4668. ISSN: 15205207. DOI: 10.1021/acs.jpcb.6b12510.

[95] G. Müller et al. "Residual Antimicrobial Effect of Chlorhexidine Digluconate and Octenidine Dihydrochloride on Reconstructed Human Epidermis". In: *Skin Pharmacology and Physiology* 27 (1 2014).

[96] T. J. Karpanen et al. "Penetration of chlorhexidine into human skin". In: *Antimicrobial Agents and Chemotherapy* 52 (10 Oct. 2008), pp. 3633–3636. ISSN: 00664804. DOI: 10.1128/AAC.00637-08.

[97] T. Kuyyakanond and L. B. Quesnel. “The mechanism of action of chlorhexidine”. In: *FEMS Microbiology Letters* 100 (1-3 1992), pp. 211–215. ISSN: 15746968. DOI: 10.1016/0378-1097(92)90211-6.

[98] M. Karplus. “Development of multiscale models for complex chemical systems: From H+H2 to biomolecules (nobel lecture)”. In: *Angewandte Chemie - International Edition* 53 (38 2014), pp. 9992–10005. ISSN: 15213773. DOI: 10.1002/anie.201403924.

[99] T. J. Piggot, D. A. Holdbrook, and S. Khalid. “Electroporation of the E. coli and S. aureus membranes: Molecular dynamics simulations of complex bacterial membranes”. In: *Journal of Physical Chemistry B* 115 (45 2011), pp. 13381–13388. ISSN: 15205207. DOI: 10.1021/jp207013v.

[100] A. Ebbensgaard et al. “The role of outer membrane proteins and lipopolysaccharides for the sensitivity of escherichia coli to antimicrobial peptides”. In: *Frontiers in Microbiology* 9 (2153 2018). ISSN: 1664302X. DOI: 10.3389/fmicb.2018.02153.

[101] F. Joodaki, L. M. Martin, and M. L. Greenfield. “Generation and Computational Characterization of a Complex Staphylococcus aureus Lipid Bilayer”. In: *Langmuir* 38 (31 2022), pp. 9481–9499. ISSN: 15205827. DOI: 10.1021/acs.langmuir.2c00483.

[102] S. A. Hollingsworth and R. O. Dror. “Molecular Dynamics Simulation for All”. In: *Neuron* 99 (6 2018), pp. 1129–1143. ISSN: 10974199. DOI: 10.1016/j.neuron.2018.08.011. URL: https://doi.org/10.1016/j.neuron.2018.08.011.

[103] J. Wildman et al. “General Force-Field Parametrization Scheme for Molecular Dynamics Simulations of Conjugated Materials in Solution”. In: *Journal of Chemical Theory and Computation* 12 (8 2016), pp. 3813–3824. ISSN: 15499626. DOI: 10.1021/acs.jctc.5b01195.

[104] A. Kukol. *Molecular modeling of proteins*. 2nd ed. 2014. ISBN: 9781493914654. DOI: 10.1007/978-1-4939-1465-4.

[105] A. T. Boags, F. Samsudin, and S. Khalid. “Binding from Both Sides: TolR and Full-Length OmpA Bind and Maintain the Local Structure of the E. coli Cell Wall”. In: *Structure* 27 (4 2019), 713–724.e2. ISSN: 18784186. DOI: 10.1016/j.str.2019.01.001. URL: https://doi.org/10.1016/j.str.2019.01.001.

[106] C. Pedebos et al. “The hitchhiker’s guide to the periplasm: Unexpected molecular interactions of polymyxin B1 in E. coli”. In: *Structure* 29 (5 2021), 444–456.e2. ISSN: 18784186. DOI: 10.1016/j.str.2021.01.009. URL: https://doi.org/10.1016/j.str.2021.01.009.

[107] J. K. Marzinek, R. G. Huber, and P. J. Bond. “Multiscale modelling and simulation of viruses”. In: *Current Opinion in Structural Biology* 61 (2020), pp. 146–152. ISSN: 1879033X. DOI: 10.1016/j.sbi.2019.12.019. URL: https://doi.org/10.1016/j.sbi.2019.12.019.

[108] L. Casalino et al. “AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics”. In: *International Journal of High Performance Computing Applications* 35 (5 2021), pp. 432–451. ISSN: 17412846. DOI: 10.1177/10943420211006452.

[109] S. Khalid, T. J. Piggot, and F. Samsudin. "Atomistic and Coarse Grain Simulations of the Cell Envelope of Gram-Negative Bacteria: What Have We Learned?" In: *Accounts of Chemical Research* 52 (1 2019), pp. 180–188. ISSN: 15204898. DOI: 10.1021/acs.accounts.8b00377.

[110] T. Polenova, R. Gupta, and A. Goldbourt. "Magic angle spinning NMR spectroscopy: A versatile technique for structural and dynamic analysis of solid-phase systems". In: *Analytical Chemistry* 87 (11 2015), pp. 5458–5469. ISSN: 15206882. DOI: 10.1021/ac504288u.

[111] E. Andrew, A. Bradbury, and R. Eades. "Nuclear Magnetic Resonance Spectra from a Crystal rotated at High Speed". In: *Nature* 182 (4650 1958), p. 1659.

[112] A. McDermott. "Structure and Dynamics of Membrane Proteins by Magic Angle Spinning Solid-State NMR". In: *Annual Review of Biophysics* 38 (2009), pp. 385–403.

[113] S. J. Opella. "Solid-state NMR and Membrane Proteins". In: *Journal of Magnetic Resonance* 253 (2015), pp. 129–137. DOI: 10.1053/j.gastro.2016.08.014.CagY.

[114] R. W. Hockney, S. P. Goel, and J. W. Eastwood. "Quiet high-resolution computer models of a plasma". In: *Journal of Computational Physics* 14 (2 1974), pp. 148–158. ISSN: 10902716. DOI: 10.1016/0021-9991(74)90010-2.

[115] H. Bekker et al. "Gromacs: A parallel computer for molecular dynamics simulations". In: *Physics computing* 92 (1993), pp. 252–256.

[116] M. Abraham et al. "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers". In: *SoftwareX* 1-2 (2015), pp. 19–25.

[117] H. Berendsen, D. Spoel, and R. Drunen. "GROMACS: A message-passing parallel molecular dynamics implementation". In: *Computer Physics Communications* 91 (1-3 1995), pp. 43–56.

[118] E. Lindahl, B. Hess, and D. van der Spoel. *GROMACS 3.0: A package for molecular simulation and trajectory analysis*. 2001. DOI: 10.1007/S008940100045.

[119] H. C. Andersen. "Molecular dynamics simulations at constant pressure and/or temperature". In: *J. Chem. Phys.* 72 (1980), pp. 2384–2393.

[120] M. Parrinello and A. Rahman. "Crystal structure and pair potentials: A molecular-dynamics study". In: *Physical Review Letters* 45 (1980).

[121] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals: A new molecular dynamics method". In: *Journal of Applied Physics* 52 (12 1981), pp. 7182–7190. ISSN: 00218979. DOI: 10.1063/1.328693.

[122] M. Parrinello and A. Rahman. "Strain fluctuations and elastic constants". In: *The Journal of Chemical Physics* 76 (5 1982), pp. 2662–2666. ISSN: 00219606. DOI: 10.1063/1.443248.

[123] S. E. Feller, D. Huster, and K. Gawrisch. "Interpretation of NOESY cross-relaxation rates from molecular dynamics simulation of a lipid bilayer". In: *Journal of the American Chemical Society* 121 (38 1999), pp. 8963–8964. ISSN: 00027863. DOI: 10.1021/ja991456n.

[124] L. L. Holte and K. Gawrisch. "Determining ethanol distribution in phospholipid multilayers with MAS- NOESY spectra". In: *Biochemistry* 36 (15 1997), pp. 4669–4674. ISSN: 00062960. DOI: 10.1021/bi9626416.

[125] J. Kremkow et al. "Membrane interaction of ibuprofen with cholesterol-containing lipid membranes". In: *Biomolecules* 10 (10 2020), p. 1384. ISSN: 2218273X. DOI: 10.3390/biom10101384.

[126] C. Waller et al. "Impact on S. aureus and E. coli membranes of treatment with chlorhexidine and alcohol solutions: insights from molecular simulations and nuclear magnetic resonance." In: *J Mol Biol* 435 (11 2023), p. 167953.

[127] J. Perry, N. Waglechner, and G. Wright. "The prehistory of antibiotic resistance". In: *Cold Spring Harbor Perspectives in Medicine* 6 (6 June 2016). ISSN: 21571422. DOI: 10.1101/cshperspect.a025197.

[128] M. Morehead and C. Scarbrough. "Emergence of Global Antibiotic Resistance". In: *Primary Care: Clinics in Office Practice* 45 (3 Sept. 2018), pp. 467–484.

[129] D. D. Blaney et al. "Use of alcohol-based hand sanitizers as a risk factor for norovirus outbreaks in long-term care facilities in northern New England". In: *American Journal of Infection Control* 39 (4 2011), pp. 296–301. ISSN: 01966553. DOI: 10.1016/j.ajic.2010.10.010. URL: http://dx.doi.org/10.1016/j.ajic.2010.10.010.

[130] E. J. Fendler et al. "The impact of alcohol hand sanitizer use on infection rates in an extended care facility". In: *American Journal of Infection Control* 30 (4 2002), pp. 226–233. ISSN: 01966553. DOI: 10.1067/mic.2002.120129.

[131] J. Hilburn et al. "Use of alcohol hand sanitizer as an infection control strategy in an acute care facility". In: *American Journal of Infection Control* 31 (2 2003), pp. 109–116. ISSN: 01966553. DOI: 10.1067/mic.2003.15.

[132] A. C. Foddai, I. R. Grant, and M. Dean. "Efficacy of instant hand sanitizers against foodborne pathogens compared with hand washing with soap and water in food preparation settings: A systematic review". In: *Journal of Food Protection* 79 (6 2016), pp. 1040–1054. ISSN: 19449097. DOI: 10.4315/0362-028X.JFP-15-492.

[133] S. E. Feller et al. "Nuclear Overhauser enhancement spectroscopy cross-relaxation rates and ethanol distribution across membranes". In: *Biophysical Journal* 82 (3 2002), pp. 1396–1404. ISSN: 00063495. DOI: 10.1016/S0006-3495(02)75494-5. URL: http://dx.doi.org/10.1016/S0006-3495(02)75494-5.

[134] M. Patra et al. "Under the influence of alcohol: The effect of ethanol and methanol on lipid bilayers". In: *Biophysical Journal* 90 (4 2006), pp. 1121–1135. ISSN: 00063495. DOI: 10.1529/biophysj.105.062364. URL: http://dx.doi.org/10.1529/biophysj.105.062364.

[135] M. Ghorbani et al. "Molecular dynamics simulations of ethanol permeation through single and double-lipid bilayers". In: *Journal of Chemical Physics* 153 (12 2020). ISSN: 10897690. DOI: 10.1063/5.0013430. URL: https://doi.org/10.1063/5.0013430.

[136] J. Ankarloo, S. Wikman, and I. A. Nicholls. "Escherichia coli mar and acrAB mutants display no tolerance to simple alcohols". In: *International Journal of Molecular Sciences* 11 (4 2010), pp. 1403–1412. ISSN: 14220067. DOI: 10.3390/ijms11041403.

[137] N. P. Franks and W. R. Lieb. "Partitioning of long-chain alcohols into lipid bilayers: Implications for mechanisms of general anesthesia". In: *Proceedings of the National Academy of Sciences of the United States of America* 83 (14 1986), pp. 5116–5120. ISSN: 00278424. DOI: 10.1073/pnas.83.14.5116.

[138] S. Šegota and D. urd ica Težak. "Spontaneous formation of vesicles". In: *Advances in Colloid and Interface Science* 121 (1-3 2006), pp. 51–75. ISSN: 00018686. DOI: 10.1016/j.cis.2006.01.002.

[139] G. Salvi, P. D. L. Rios, and M. Vendruscolo. "Effective interactions between chaotropic agents and proteins". In: *Proteins: Structure, Function and Genetics* 61 (3 2005), pp. 492–499. ISSN: 08873585. DOI: 10.1002/prot.20626.

[140] P. Bhaganna et al. "Hydrophobic substances induce water stress in microbial cells". In: *Microbial Biotechnology* 3 (6 2010), pp. 701–716. ISSN: 17517907. DOI: 10.1111/j.1751-7915.2010.00203.x.

[141] C. R. Myers and J. M. Myers. "Localization of cytochromes to the outer membrane of anaerobically grown Shewanella putrefaciens MR-1". In: *Journal of Bacteriology* 174 (11 1992), pp. 3429–3438. ISSN: 00219193. DOI: 10.1128/jb.174.11.3429-3438.1992.

[142] A. Bachmann et al. "A comparative study of the localization and membrane topology of members of the RIFIN, STEVOR and PfMC-2TM protein families in Plasmodium falciparum-infected erythrocytes". In: *Malaria Journal* 14 (2015), p. 274. ISSN: 14752875. DOI: 10.1186/s12936-015-0784-2.

[143] T. Kuil et al. "Ethanol tolerance of Clostridium thermocellum: the role of chaotropicity, temperature and pathway thermodynamics on growth and fermentative capacity". In: *Microbial Cell Factories* 21 (2022), p. 273. ISSN: 14752859. DOI: 10.1186/s12934-022-01999-8. URL: https://doi.org/10.1186/s12934-022-01999-8.

[144] Y. L. Tan et al. "Characterization of Denatured States and Reversible Unfolding of Sensory Rhodopsin II". In: *Journal of Molecular Biology* 430 (21 2018), pp. 4068–4086. ISSN: 10898638. DOI: 10.1016/j.jmb.2018.07.031. URL: https://doi.org/10.1016/j.jmb.2018.07.031.

[145] R. Zangi, R. Zhou, and B. J. Berne. "Urea's action on hydrophobic interactions". In: *Journal of the American Chemical Society* 131 (4 2009), pp. 1535–1541. ISSN: 00027863. DOI: 10.1021/ja807887g.

[146] D. B. Goldstein. "Effect of alcohol on cellular membranes". In: *Ann Emerg Med.* 15 (9 1986), pp. 1013–1018.

[147] H. I. Ingólfsson and O. S. Andersen. "Alcohol's effects on lipid bilayer properties". In: *Biophysical Journal* 101 (4 Aug. 2011), pp. 847–855. ISSN: 00063495. DOI: 10.1016/j.bpj.2011.07.013.

[148] N. Casanova-Morales et al. “Identifying Chaotropic and Kosmotropic Agents by Nanorheology”. In: *Journal of Physical Chemistry B* 122 (14 2018), pp. 3754–3759. ISSN: 15205207. DOI: 10.1021/acs.jpcb.7b12782.

[149] J. J. Minty et al. “Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli”. In: *Microbial Cell Factories* (2011). ISSN: 14752859. DOI: 10.1186/1475-2859-10-18.

[150] M. L. Rotter. “Arguments for alcoholic hand disinfection”. In: *Journal of Hospital Infection* 48 (2001), S4–S8. DOI: 10.4324/9780203936313.

[151] T. Kondela et al. “Alcohol interactions with lipid bilayers”. In: *Molecules* 22 (12 Nov. 2017), p. 2078. ISSN: 14203049. DOI: 10.3390/molecules22122078. URL: http://www.mdpi.com/1420-3049/22/12/2078.

[152] B. Griepernau et al. “1-Alkanols and membranes: A story of attraction”. In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1768 (11 2007), pp. 2899–2913.

[153] C. Jones. “Chlorhexidine: is it still the gold standard?” In: *Periodontology 2000* 15 (1997), pp. 55–62.

[154] P. Gilbert and L. E. Moore. “Cationic antiseptics: diversity of action under a common epithet”. In: *Journal of Applied Microbiology* 99 (4 2005), pp. 703–715. ISSN: 13645072. DOI: 10.1111/j.1365-2672.2005.02664.x.

[155] T. P. N. Hoang, M. U. Ghori, and B. R. Conway. “Topical antiseptic formulations for skin and soft tissue infections”. In: *Pharmaceutics* 13 (4 2021), p. 558. ISSN: 19994923. DOI: 10.3390/pharmaceutics13040558.

[156] X. Bai et al. “Toward a systematic exploration of nano-bio interactions”. In: *Toxicology and Applied Pharmacology* 323 (2017 2017), pp. 66–73. ISSN: 10960333. DOI: 10.1016/j.taap.2017.03.011. URL: https://doi.org/10.1016/j.taap.2017.03.011.

[157] J. Zhu et al. “Functional synergy of antimicrobial peptides and chlorhexidine acetate against gram-negative/ gram-positive bacteria and a fungus in vitro and in vivo”. In: *Infection and Drug Resistance* 12 (2019), pp. 3227–3239. ISSN: 11786973. DOI: 10.2147/IDR.S218778.

[158] E. Lei et al. “Potentiation of Vancomycin: Creating Cooperative Membrane Lysis through a "derivatization-for-Sensitization" Approach”. In: *Journal of the American Chemical Society* 144 (23 2022), pp. 10622–10639. ISSN: 15205126. DOI: 10.1021/jacs.2c03784.

[159] K. Murzyn, T. Róg, and M. Pasenkiewicz-Gierula. “Phosphatidylethanolamine-phosphatidylglycerol bilayer as a model of the inner bacterial membrane”. In: *Biophysical Journal* 88 (2 2005), pp. 1091–1103. ISSN: 00063495. DOI: 10.1529/biophysj.104.048835.

[160] T. Doherty and M. Hong. “2D 1H-31P solid-state NMR studies of the dependence of inter-bilayer water dynamics on lipid headgroup structure and membrane peptides”. In: *Journal of Magnetic Resonance* 196 (1 2009), pp. 39–47. ISSN: 10907807. DOI: 10.1016/j.jmr.2008.10.001. URL: http://dx.doi.org/10.1016/j.jmr.2008.10.001.

[161] A. Y. Peleg et al. “Whole genome characterization of the mechanisms of daptomycin resistance in clinical and laboratory derived isolates of staphylococcus aureus”. In: *PLoS ONE* 7 (1 2012). ISSN: 19326203. DOI: 10.1371/journal.pone.0028316.

[162] E. Wu et al. “CHARMM-GUI Membrane Builder Toward Realistic Biological Membrane Simulations”. In: *J. Comput. Chem.* 35 (2004), pp. 1997–2004.

[163] S. Jo et al. “CHARMM-GUI Membrane Builder for Mixed Bilayers and Its Application to Yeast Membranes”. In: *Biophys. J.* 97 (2009), pp. 50–58.

[164] J. Lee et al. “CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans”. In: *J. Chem. Theory Comput.* 15 (2019), pp. 775–786.

[165] B. V. Oosten et al. “Small molecule interaction with lipid bilayers: A molecular dynamics study of chlorhexidine”. In: *Journal of Molecular Graphics and Modelling* 48 (2014), pp. 96–104. ISSN: 10933263. DOI: 10.1016/j.jmgm.2013.12.007.

[166] A. S. Rose et al. “Web-based molecular graphics for large complexes”. In: Association for Computing Machinery, Inc, July 2016, pp. 185–186. ISBN: 9781450344289. DOI: 10.1145/2945292.2945324.

[167] E. Lindahl, B. Hess, and D. van der Spoel. “GROMACS 3.0: A package for molecular simulation and trajectory analysis”. In: *Journal of Molecular Modeling* 7 (8 2001), pp. 306–317. ISSN: 16102940. DOI: 10.1007/S008940100045.

[168] A. D. MacKerell et al. “All-atom empirical potential for molecular modeling and dynamics studies of proteins”. In: *J Phys Chem B* 102 (1998), pp. 3586–3616.

[169] B. Klauda J et al. “Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types”. In: *J Phys Chem B* 114 (23 2010), pp. 7830–7843. ISSN: 15378276.

[170] S. Nosé. “A molecular dynamics method for simulations in the canonical ensemble”. In: *Molecular Physics* 52 (2 1983), pp. 255–268.

[171] W. G. Hoover. “Canonical dynamics: Equilibrium phase-space distributions”. In: *Physical Review A* 31 (3 1985), pp. 1695–1697.

[172] S. Nosé and M. Klein. “Constant pressure molecular dynamics for molecular systems”. In: *Molecular Physics* 50 (5 1983), pp. 1055–1076.

[173] H. Yu et al. “Molecular dynamics simulation of the interactions between EHD1 EH domain and multiple peptides”. In: *Journal of Zhejiang University: Science B* 16 (10 2015), pp. 883–896. ISSN: 18621783. DOI: 10.1631/jzus.B1500106.

[174] W. Heumphrey, A. Dalke, and K. Schulten. “VMD - Visual Molecular Dynamics”. In: *Journal of molecular graphics* 14 (1996), pp. 33–38.

[175] T. Kluyver, B. Ragan-Kelley, and F. P. “Jupyter Notebooks – a publishing format for reproducible computational workflows”. In: ed. by F. Loizides and B. Schmidt. IOS Press, 2016, pp. 87–90.

[176] R. J. Gowers et al. “MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations”. In: ed. by S. Benthall and S. Rostrup. 2016, pp. 98–105.

[177] K. M. Dombek and L. O. Ingram. "Effects of ethanol on the Escherichia coli plasma membrane". In: *Journal of Bacteriology* 157 (1 1984), pp. 233–239. ISSN: 00219193. DOI: 10.1128/jb.157.1.233-239.1984.

[178] M. Rzycki et al. "Unraveling the mechanism of octenidine and chlorhexidine on membranes: Does electrostatics matter?" In: *Biophysical Journal* 120 (16 2021), pp. 3392–3408. ISSN: 15420086. DOI: 10.1016/j.bpj.2021.06.027.

[179] N. A. Berglund et al. "Interaction of the Antimicrobial Peptide Polymyxin B1 with Both Membranes of E. coli: A Molecular Dynamics Study". In: *PLoS Computational Biology* 11 (4 Apr. 2015). ISSN: 15537358. DOI: 10.1371/journal.pcbi.1004180.

[180] L. Lecoq et al. "Solid-state NMR for studying the structure and dynamics of viral assemblies". In: *Viruses* 12 (10 2020), p. 1069. ISSN: 19994915. DOI: 10.3390/v12101069.

[181] P. A. Bottomley. "The rise of human in vivo NMR spectroscopy". In: *Solid State Nuclear Magnetic Resonance* 9 (1 1997), pp. 29–40. ISSN: 09262040. DOI: 10.1016/S0926-2040(97)00040-4.

[182] D. E. Warschawski, A. A. Arnold, and I. Marcotte. "A New Method of Assessing Lipid Mixtures by 31P Magic-Angle Spinning NMR". In: *Biophysical Journal* 114 (6 2018), pp. 1368–1376. ISSN: 15420086. DOI: 10.1016/j.bpj.2018.01.025.

[183] J. B. Speyer et al. "Magnetic orientation of sphingomyelin-lecithin bilayers". In: *Biophysical Journal* 51 (4 1987), pp. 687–691. ISSN: 00063495. DOI: 10.1016/S0006-3495(87)83394-5.

[184] E. J. Dufourc et al. "Dynamics of phosphate head groups in biomembranes. Comprehensive analysis using phosphorus-31 nuclear magnetic resonance lineshape and relaxation time measurements". In: *Biophysical Journal* 61 (1 1992), pp. 42–57. ISSN: 00063495. DOI: 10.1016/S0006-3495(92)81814-3.

[185] A. W. Overhauser. "Polarization of nuclei in metals". In: *Physical Review* 92 (2 1953), pp. 411–415. ISSN: 0031899X. DOI: 10.1103/PhysRev.92.411.

[186] C. Doyen et al. "Nuclear Magnetic Resonance Spectroscopy: A Multifaceted Toolbox to Probe Structure, Dynamics, Interactions, and Real-TimeIn SituRelease Kinetics in Peptide-Liposome Formulations". In: *Molecular Pharmaceutics* 18 (7 2021), pp. 2521–2539. ISSN: 15438392. DOI: 10.1021/acs.molpharmaceut.1c00037.

[187] H. A. Scheidt and D. Huster. "The interaction of small molecules with phospholipid membranes studied by 1H NOESY NMR under magic-angle spinning". In: *Acta Pharmacologica Sinica* 29 (1 2008), pp. 35–49. ISSN: 16714083. DOI: 10.1111/j.1745-7254.2008.00726.x.

[188] K. Gawrisch, N. V. Eldho, and I. V. Polozov. "Novel NMR tools to study structure and dynamics of biomembranes". In: *Chemistry and Physics of Lipids* 116 (1-2 2002), pp. 135–151. ISSN: 00093084. DOI: 10.1016/S0009-3084(02)00024-5.

[189] E. L. Hahn. "Spin echoes". In: *Physical Review* 80 (4 1950), pp. 580–594. ISSN: 0031899X. DOI: 10.1103/PhysRev.80.580.

[190] J. Jeener et al. "Investigation of exchange processes by two- dimensional NMR spectroscopy". In: *J. Chem. Phys.* 71 (July 1979), pp. 4546–4553.

[191] D. Marion et al. “Rapid Recording of 2D NMR Spectra without Phase Cycling. Application to the Study of Hydrogen Exchange in Proteins”. In: *J. Magn. Reson.* 85 (1989), pp. 393–399.

[192] F. Delaglio et al. “NMRPipe: a multidimensional spectral processing system based on UNIX pipes”. In: *J Biol NMR* 6 (1995), pp. 277–293.

[193] T. M. Inc. *MATLAB and Statistics Toolbox Release R2021a*. The MathWorks Inc., Mar. 2021.

[194] J. D. van Beek. “matNMR: A flexible toolbox for processing, analyzing and visualizing magnetic resonance data in Matlab®”. In: *Journal of Magnetic Resonance* 187 (1 2007), pp. 19–26. ISSN: 10907807. DOI: 10.1016/j.jmr.2007.03.017.

[195] S. Akita et al. “Effects of Hand Hygiene Using 4 precent Chlorhexidine Gluconate or Natural Soap during Hand Rubbing Followed by Alcohol-Based 1 precent Chlorhexidine Gluconate Sanitizer Lotion in the Operating Room”. In: *Advances in Wound Care* 11 (1 2022), pp. 1–9. ISSN: 21621934. DOI: 10.1089/wound.2020.1352.

[196] J. K. Aronson. “Chlorhexidine.” In: 16th ed. Elsevier B.V., 2016, pp. 239–248. ISBN: 978-0-444-53716-4. DOI: 10.1016/b978-0-444-53717-1.00474-1. URL: https://www.sciencedirect.com/referencework/9780444537164/meylers-side-effects-of-drugs#book-info.

[197] M. Jansson et al. “Magnetic alignment and orientational order of dipalmitoylphosphatidylcholine bilayers containing palmitoyllsyophosphatidylcholine”. In: *Chem. Phys. Lipids* 54 (3-4 1990), pp. 157–170.

[198] J. Seelig, F. Borle, and T. A Cross. “Magnetic ordering of phospholipid membranes”. In: *Biochim Biophys Acta Biomembr* 814 (1 1985), pp. 195–198.

[199] E. L. A. Briggs et al. “Interaction between the NS4B amphipathic helix, AH2, and charged lipid headgroups alters membrane morphology and AH2 oligomeric state - Implications for the Hepatitis C virus life cycle”. In: *Biochimica et Biophysica Acta - Biomembranes* 1848 (8 2015), pp. 1671–1677. ISSN: 18792642. DOI: 10.1016/j.bbamem.2015.04.015. URL: http://dx.doi.org/10.1016/j.bbamem.2015.04.015.

[200] F. Separovic, D. W. Keizer, and M.-A. Sani. “In-cell Solid-State NMR Studies of Antimicrobial Peptides”. In: *Front Med Technol* 17 (2 2020), p. 610203.

[201] I. Solomon. “Relaxation Processes in a System of Two Spins”. In: *American Physical Society* 99 (2 1955), p. 559.

[202] D. Huster and K. Gawrisch. “NOESY NMR Crosspeaks between Lipid Headgroups and Hydrocarbon Chains: Spin Diffusion or Molecular Disorder?” In: *Solid State Nucl. Magn. Reson* 84 (3 1988), pp. 95–117. DOI: 10.1021/ja. URL: http://pubs.acs.org..

[203] L. Zhang et al. “SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity”. In: *Nature Communications* 11 (1 2020), pp. 1–9. ISSN: 20411723. DOI: 10.1038/s41467-020-19808-4. URL: http://dx.doi.org/10.1038/s41467-020-19808-4.

[204] P. Sashi, U. M. Yasin, and A. K. Bhuyan. "Unfolding action of alcohols on a highly negatively charged state of cytochrome c". In: *Biochemistry* 51 (15 2012), pp. 3273–3283. ISSN: 00062960. DOI: 10.1021/bi201716r.

[205] P. Sashi et al. "Preferential water exclusion in protein unfolding". In: *Journal of Physical Chemistry B* 118 (3 2014), pp. 717–723. ISSN: 15206106. DOI: 10.1021/jp4111103.

[206] M. Kinoshita, Y. Okamoto, and F. Hirata. "Peptide conformations in alcohol and water: Analyses by the reference interaction site model theory". In: *Journal of the American Chemical Society* 122 (12 2000), pp. 2773–2779. ISSN: 00027863. DOI: 10.1021/ja993939x.

[207] L. Zuzic et al. "Uncovering cryptic pockets in the SARS-CoV-2 spike glycoprotein". In: *Structure* 30 (8 2022), 1062–1074.e4. ISSN: 18784186. DOI: 10.1016/j.str.2022.05.006. URL: https://doi.org/10.1016/j.str.2022.05.006.

[208] Y. Watanabe et al. "Site-specific glycan analysis of the SARS-CoV-2 spike". In: *Science* 369 (6501 2020), pp. 330–333. ISSN: 10959203. DOI: 10.1126/science.abb9983.

[209] B. Chapman and J. Chang. "Biopython". In: *ACM SIGBIO Newsletter* 20 (2 2000), pp. 15–19. ISSN: 0163-5697. DOI: 10.1145/360262.360268.

[210] R. T. McGibbon et al. "MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories". In: *Biophysical Journal* 109 (8 2015), pp. 1528–1532. ISSN: 15420086. DOI: 10.1016/j.bpj.2015.08.015.

[211] H. G. Niesters et al. "Epitopes on the peplomer protein of infectious bronchitis virus strain M41 as defined by monoclonal antibodies". In: *Virology* 161 (2 1987), pp. 511–519. ISSN: 10960341. DOI: 10.1016/0042-6822(87)90145-0.

[212] M. D. Parker et al. "Primary structure of the S peplomer gene of bovine coronavirus and surface expression in insect cells". In: *Journal of General Virology* 71 (2 1990), pp. 263–270. ISSN: 00221317. DOI: 10.1099/0022-1317-71-2-263.

[213] R. Asor et al. "The spike-ACE2 interaction underlying SARS-CoV-2 infection and inhibition is enhanced by intermolecular cross linking". In: *Biophysical Journal* 121 (3 2022), 333a. ISSN: 00063495. DOI: 10.1016/j.bpj.2021.11.1120.

[214] D. J. Benton et al. "Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion". In: *Nature* 588 (7837 Dec. 2020), pp. 327–330. ISSN: 14764687. DOI: 10.1038/s41586-020-2772-0.

[215] B. Turoňová et al. "In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges". In: *Science* 370 (6513 2020), pp. 203–208. ISSN: 10959203. DOI: 10.1126/science.abd5223.

[216] Y. K. Choi et al. "Structure, Dynamics, Receptor Binding, and Antibody Binding of Fully-glycosylated Full-length SARS-CoV-2 Spike Protein in a Viral Membrane". In: *JTC* 17 (2021), pp. 2479–2487.

[217] P. Raghuvamsi et al. "Sars-cov-2 s protein:Ace2 interaction reveals novel allosteric targets". In: *eLife* 10 (e63646 2021). ISSN: 2050084X. DOI: 10.7554/eLife.63646.

[218] F. Hikmet et al. “The protein expression profile of ACE2 in human tissues”. In: *Molecular Systems Biology* 16 (e9610 2020). ISSN: 1744-4292. DOI: 10.15252/msb.20209610.

[219] A. J. Turner. “ACE2 Cell Biology, Regulation, and Physiological Functions”. In: *The Protective Arm of the Renin Angiotensin System (RAS): Functional Aspects and Therapeutic Implications* (January 2015), pp. 185–189. DOI: 10.1016/B978-0-12-801364-9.00025-0.

[220] W. Ni et al. “Role of angiotensin-converting enzyme 2 (ACE2) in COVID-19”. In: *Crit Care* 24 (1 2020), p. 422.

[221] Y. Huang et al. *Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19*. Sept. 2020. DOI: 10.1038/s41401-020-0485-4.

[222] B. J. Bosch et al. “The Coronavirus Spike Protein Is a Class I Virus Fusion Protein: Structural and Functional Characterization of the Fusion Core Complex”. In: *Journal of Virology* 77 (16 2003), pp. 8801–8811. ISSN: 0022-538X. DOI: 10.1128/jvi.77.16.8801-8811.2003.

[223] T. Heald-Sargent and T. Gallagher. “Ready, set, fuse! the coronavirus spike protein and acquisition of fusion competence”. In: *Viruses* 4 (4 2012), pp. 557–580. ISSN: 19994915. DOI: 10.3390/v4040557.

[224] J. L. Sampaio et al. “Membrane lipidome of an epithelial cell line”. In: *Proceedings of the National Academy of Sciences of the United States of America* 108 (5 Feb. 2011), pp. 1903–1907. ISSN: 00278424. DOI: 10.1073/pnas.1019267108.

[225] A. Santamaria et al. “Investigation on the relationship between lipid composition and structure in model membranes composed of extracted natural phospholipids”. In: *Journal of Colloid and Interface Science* 637 (2023), pp. 55–66. ISSN: 10957103. DOI: 10.1016/j.jcis.2023.01.043. URL: https://doi.org/10.1016/j.jcis.2023.01.043.

[226] L. Monticelli et al. “The MARTINI coarse-grained force field: Extension to proteins”. In: *Journal of Chemical Theory and Computation* 4 (5 2008), pp. 819–834. ISSN: 15499618. DOI: 10.1021/ct700324x.

[227] D. H. D. Jong et al. “Improved parameters for the martini coarse-grained protein force field”. In: *Journal of Chemical Theory and Computation* 9 (1 2013), pp. 687–697. ISSN: 15499618. DOI: 10.1021/ct300646g.

[228] G. Bussi, D. Donadio, and M. Parrinello. “Canonical sampling through velocity rescaling”. In: *Journal of Chemical Physics* 126 (1 2007). ISSN: 00219606. DOI: 10.1063/1.2408420.

[229] L. Onsager. “Electric moments of molecules in liquids”. In: *Journal of the American Chemical Society* 58 (8 1936), pp. 1486–1493.

[230] J. Barker and R. Watts. “Monte Carlo studies of the dielectric properties of water-like models”. In: *Molecular Physics* 26 (3 1973), pp. 789–792.

[231] J. D. Watts and R. J. Bartlett. “The coupled-cluster single, double, and triple excitation model for open-shell single reference functions”. In: *Journal of Chemical Physics* 93 (1998), p. 6104. ISSN: 00219606. DOI: 10.1063/1.459002.

Chapter 9

Supplementary tables

Peak	EcIM	EcIM+CHX	EcIM+PROH	EcIM+CHX+PROH
G2	0.7559	0.8989	0.8309	0.5944
H_2O	3.7600	4.3460	3.249	3.858
α	1.2340	0.6998	0.3459	0.7089
G1	0.9685	0.6627	0.4703	0.6026
G1/G3	0.8009	0.7041	0.5782	0.6094
β	1.0250	0.6671	0.6031	0.5622
C2	0.7007	0.6431	0.5836	0.5772
CH_2=C	0.7386	0.6591	0.6116	0.5947
C3	0.6616	0.6457	0.6014	0.6500
$(CH_2)_n$	0.7820	0.6903	0.7	0.6467
CH_3	1.0690	0.8704	1.085	0.8796
CHX(1)	-	0.8575	-	0.4243
CHX(2)	-	0.8353	-	0.4167
CHX(3)	-	0.8108	-	-
PROH(6)	-	-	0.8304	0.8108
*	-	1.7820	-	1.137

Table 9.1: T1 relaxation times for peaks in 1D proton NMR of the EcIM without and with sanitisers.

Protein	Residue (ps)	Glycan (ns)
S	17	fa3g2
S	61	m5
S	74	fa3g2
S	122	m5
S	149	fa2g1
S	165	fa2g1
S	234	m8
S	282	fa3g2
S	330	fa2g1
S	343	fa2g1
S	603	m5
S	616	fa2g1
S	657	m5a1g1
S	709	m5
S	717	m5
S	801	m5
S	1074	m5
S	1098	a2g1
S	1134	fa2g1
S	1158	a2g1
S	1173	fa4
S	1194	fa4g2s1
ACE2	53	fa2g1
ACE2	90	fa2g2s2
ACE2	103	fa2g1
ACE2	322	fa2g1
ACE2	432	fa2g1
ACE2	546	fa2g1

Table 9.2: Position and glycan added the proteins in the system by the Martini glycosylator script.