The effects of oxidative damage on peptide-lipid interactions in human lung surfactant and model lipid membranes

Birkbeck
UNIVERSITY OF LONDON

The Effects of Oxidative Damage on Peptide-Lipid Interactions in Human Lung Surfactant and Model Lipid Membranes

Vivian Wang

1-31-2024

I, Vivian Wang, declare the work presented in this thesis, titled 'The Effects of Oxidative Damage on Peptide-Lipid Interactions in Human Lung Surfactant and Model Lipid Membranes' as my own. I confirm that information derived from the literature has been cited accordingly, with a list of references provided.

……

Table of Contents

Abstract

Lung surfactant comprises surfactant lipids and proteins, that work synergistically in maintaining the biophysical function of the lungs by reducing the surface tension at the air-water interface in the alveoli. To date, an experimentally determined structure of full-length Sp-B has not been found. This research established a protocol to express and purify recombinant Sp-B, using BL21 (DE3) *E. coli*. Removal of the His6 tag caused Sp-B to be less stable especially at higher concentrations, however the protein was seen to be more stable when the pH of the buffer is lowered. The expressed His6 Sp-B and Sp-B were analysed using circular dichroism spectroscopy, where both samples were shown to have a predominantly alpha helical structure. The addition of increasing concentrations of trifluoroethanol caused both samples to have increased alpha helicity, which agrees with structural predictions from AlphaFold2, and previous experimental data on peptide mimics of Sp-B; Sp-B_{1-25} and Mini-B, which also formed a more helical structure when exposed to a lipid membrane environment. Subsequently, molecular dynamics simulations on Sp-B_{1-25} were performed to investigate the effects of oxidative damage in surfactant lipid monolayers, on lipid-lipid and peptide-lipid interactions. Oxidative damage to PG lipids increased monolayer compressibility, changed how the surface pressure varied during expansion and compression of the monolayer and influenced the positioning of Sp-B_{1-25} within the lipid monolayers. Additionally, damage to the unsaturated lipid tails destabilised the alpha helical structure of Sp-B_{1-25}. Finally, neutron reflectivity was used to examine the insertion of the antimicrobial peptide protegrin (PG-1) into model lipid membranes. Oxidative damage to PC lipids within the membrane was shown to have only had a modest effect on the insertion of PG-1 into the lipid membrane.

Acknowledgements

I would like to thank my supervisor Dr Katherine Thompson, who has given me her unwavering support and guidance with this project. Her extensive work on surfactant lipids and proteins has highlighted her expertise and allowed me to gain valuable insight whilst completing both my MSc and PhD projects at Birkbeck. I could not have completed this project without her, and for that, I am extremely grateful to have been under her supervision. I would also like to thank my other supervisors, Dr Sanjib Bhakta and Dr Amandine Marechal, who gave me important feedback on my progression during our thesis committee meetings.

I would also like to thank Dr Claire Bagneris, for her guidance and extensive knowledge of protein purification techniques, while I was working in the Rosalind Franklin lab. I greatly appreciate Dr Riccardo Zenezini Chiozzi, the lab manager of the UCL MS facilities, and thank him for setting up the mass spectrometry experiments, and his help with compiling and analysing the relevant data.

I also extend my gratitude towards Dr Dave Houldershaw and Dr Mark Williams for their contributions in helping me set up the initial molecular dynamics simulations on Sparrow, where I collected most of my experimental data.

The neutron experiments conducted at the ISIS Neutron and Muon Source would not have been possible without the help of Dr Becky Welbourn and Dr Adrian Rennie. I would like to thank them for providing their insights into the fitting the neutron data and interpreting the results.

Finally, I would like to thank my friends and family for showing their continual support, as they have been a huge motivation for me to work towards my goal. Although this has been a long and arduous journey, I am truly grateful that I was given the opportunity to pursue a PhD.

List of Figures

List of Tables

List of Abbreviations

ABCA3	Adenosine triphosphate-Binding Cassette subfamily A member 3
AMPs	Anti-Microbial Peptides
APL	Area per Lipid
BLAST	Basic Local Alignment Search Tool
BRICHOS	BRI2, CHOndromodulin-1 and proSurfactant protein C domain
BSA	Bovine Serum Albumin
CATH	Class, Architecture, Topology, Homology
CD	Circular Dichroism
cDNA	Complementary Deoxyribonucleic Acid
CDSSTR	Modified VARSLC Program
CG	Coarse-Grained Simulations
CHARMM-GUI	Chemistry at Harvard Macromolecular Mechanics Graphical User Interface
CHL	Cholesterol
CONTIN	Ridge Regression Analysis Program
COPD	Chronic Obstructive Pulmonary Disease
CRD	Carbohydrate Recognition Domain
CTL	Cholesta-5,7,9(11)-trien-3-β-ol
d31POPG	1-palmitoyl-d31-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt)
d62DPPG	Deuterated 1,2-dipalmitoyl(D62)-sn-glycero-3-phosphoglycerol
DAMPs	Disease-Associated Molecular Patterns
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DNA	Deoxyribonucleic acid
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine
DOSY	Diffusion-Ordered Spectroscopy
DPPC	1,2-dipalmitoyl-sn-glycero-3-phosphocholine
DPPE	Di-palmitoyl-phosphatidylethanolamine
DPPG	1,2-dipalmitoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)
DTT	Dithiothreitol

E. Coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic Acid
ESI	Electrospray Ionisation
FA	Formic Acid
FTIR	Fourier-Transform InfraRed spectroscopy
GRAVY	GRAnd aVerage hydropathicitY index values
GROMACS	GROningen MAchine for Chemical Simulations
His6 tag	Polyhistidine Tag
HSQC	Heteronuclear Single Quantum Coherence
HT	High Tension
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-d-1-thiogalactopyranoside
K2D	Neural Network Program
LB	Luria Broth
LC	Liquid Chromatography
LC-MS	Liquid Chromatography-Mass Spectrometry
L_d	Liquid-Disordered
LJ	Lennard-Jones
LMPC	Lysomyristoyl-sn-glycero-3-phosphocholine
LMPG	Lysomyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)
L_o	Liquid-Ordered
MD	Molecular Dynamics
MP180	Membrane Proteins 180 cut-off
mRNA	Messenger Ribonucleic Acid
MWCO	Molecular Weight Cut-Off
NL	Neutral Lipids
NMR	Nuclear Magnetic Resonance
NPT	Isothermal-isobaric ensemble (fixed number of atoms, pressure, and temperature)
NR	Neutron Reflectivity
NRDS	Neonatal Respiratory Distress Syndrome
NRMSD	Normal Root Mean Squared Deviation
NRW	Null-Reflecting Water

NusA	N-utilisation substance A
NVT	Canonical ensemble (fixed number of atoms, volume and temperature)
OPLS-AA	Optimised potentials for Liquid Simulations All Atom forcefield
PA clan	Proteases of mixed nucleophile, superfamily A
PAMPs	Pathogen-Associated Molecular Patterns
PBS	Phosphate-Buffered Saline
PC	Phosphatidylcholine
PDB	Protein Data Bank
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PL	Phospholipids
pLDDT	Predicted Local Distance Difference Test
PLPC	1-palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine
POnPC	1-palmitoyl-2-(9′-oxo-nonanoyl)-sn-glycero-3-phosphocholine
POPC	1-palmitoyl-2-oleoyl-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)
PRRs	Pattern Recognition Receptors
PXPG	Oxidised 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)
RMSD	Root Mean Squared Deviation
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
RP-HPLC	Reversed-Phase High Performance Liquid Chromatography
SAPLIP	Saposin-like Proteins
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate–PolyAcrylamide Gel Electrophoresis
SELCON3	SELf-CONsistent 3 program
SFTPA1	Surfactant protein A1 gene
SFTPA2	Surfactant protein A2 gene
SFTPB	Surfactant protein B gene

SFTPC	Surfactant protein C gene
SFTPD	Surfactant protein D gene
SLD	Scattering Length Density
SMP180	Soluble and Membrane Proteins 180 cut-off
SP175	Soluble Protein 175 cut-off
Sp-A	Surfactant Protein A
Sp-B	Surfactant Protein B
$Sp\text{-}B_{1\text{-}25}$	Residues 1 to 25 of surfactant protein B
Sp-C	Surfactant Protein C
Sp-D	Surfactant Protein D
SSH	Secure SHell Server
SUMO	Small Ubiquitin-like MOdifier
TEV	Tobacco Etch Virus
TFE	Trifluoroethanol
TOF	Time-of-Flight
TTF1	Transcription termination factor 1
VARSLC	VARiable SeLeCtion Program
VFS	Volume Fraction Solvent
VMD	Visual Molecular Dynamics software
VOC	Volatile Organic Compound

Chapter 1 – Introduction

1.1 The respiratory system

The three major components of the respiratory system include the airway, lungs, and respiratory muscles. The lungs are the primary organs of respiration in humans, with each lung lying on its side of the thoracic cavity, surrounded by the pleura, with the left lung consisting of two lobes and the right consisting of three. Functionally, the respiratory system is separated into two zones, namely the conducting zone which forms a conduction path for inhaled gases travelling from the nose to the bronchioles, and the respiratory zone where gaseous exchange occurs in the alveolar ducts and alveoli. Anatomically, the upper respiratory tract consists of the nose, pharynx and larynx, and the lower respiratory tract is made of the trachea, bronchi branching into bronchioles, alveolar ducts, and alveoli (Pallav L. Shah, Stradling and Craig, 2020). Bronchioles lead into alveolar ducts that are lined with many alveoli. Alveoli, being the smallest functional units of the respiratory tract are found in the alveolar sacs of the pulmonary lobules and are made of a layer of squamous epithelial cells, referred to as type I pneumocytes, or type I alveolar cell, forming the alveolar membrane, with an extracellular matrix surrounded by capillaries to facilitate gaseous exchange. Two other cell types are also present, type II pneumocytes (or type II alveolar cells), which secrete pulmonary surfactant and alveolar macrophages. The alveolar membrane has three distinct layers – the surfactant layer, the epithelial layer, and the basement membrane (Figure 1.1).

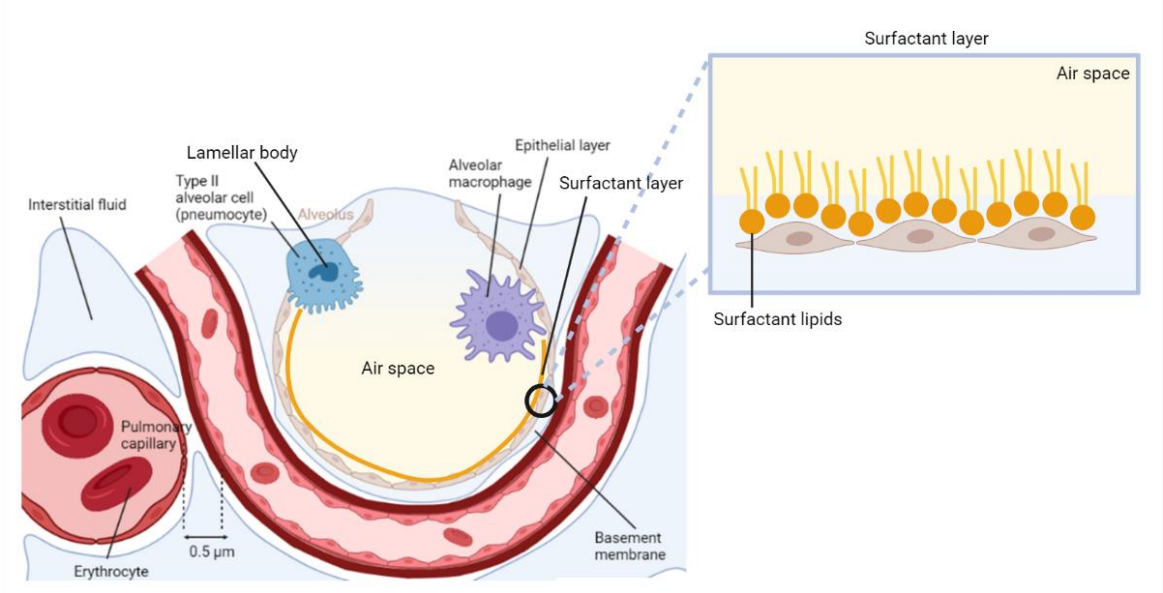


Figure 1.1: Anatomy of the alveolar membrane, consisting of the surfactant layer, epithelial layer, and basement membrane. Type I alveolar cells cover approximately 95% of the alveolar surface area, and type II alveolar cells, which secrete pulmonary surfactant, cover around 5%. Created in BioRender.

1.1.1 Gaseous exchange in the lungs

Type II alveolar cells repair damage to cells in the epithelium and secrete pulmonary surfactant into the alveolar space, forming a surfactant-lined aqueous hypophase. This pulmonary surfactant is comprised of proteins and phospholipids and is synthesised by the type II alveolar cells in the alveolar sac stage during lung development. Pulmonary surfactant minimises surface tension experienced at the air-liquid interface, a process which contributes strongly to its biophysical function of regulating respiratory mechanics, as a lowered alveolar surface tension causes a reduction in energy needed for lung inflation, leading to increased pulmonary compliance. Pulmonary surfactant also reduces elastic recoil of the lungs, hence reducing the chances of alveolar collapse during respiratory expiration (Zuo *et al.*, 2008). Pulmonary surfactant also exists in bronchioles and small airways, to prevent cohesion of the bronchiolar walls by decreasing the surface tension of the mucosal airway linings. In humans, it can protect epithelial cells from potential damage during the reopening of collapsed airways (Bilek, Dee and Gaver, 2003). Lamellar bodies appear in the cytoplasm of Type II alveolar cells (Figure 1.1) around 20 weeks gestation and are subsequently secreted via exocytosis into the aqueous layer lining the alveolar airspace. The surfactant produced forms a highly ordered, lung-

specific meshwork called tubular myelin. A study showed that tubular myelin constructed from fetal murine lungs were approximately 2-3 microns in diameter, spherical and surrounded by lamellar bodies. It was also discovered that saturated phospholipids, Ca^{2+} and surfactant apoproteins A and B are key to the structure and function of tubular myelin (Young, Fram and Larson, 1992).

1.2 Composition of pulmonary surfactant

Approximately 90% of pulmonary surfactant consists of lipids, mainly zwitterionic phosphatidylcholines (PC), phosphatidylinositol (PI) and anionic phosphatidylglycerol (PG) lipids (Walther, Gordon, and Waring, 2019), and the remaining 10% being key surfactant proteins (Sp-A, Sp-B, Sp-C and Sp-D). At the air-liquid interface, a lipid-protein surfactant monolayer forms where the hydrophobic lipid tails point outwards facing the air and the hydrophilic heads face inwards, immersed in the aqueous layer. In humans, the lipid fraction of pulmonary surfactant consists of approximately 41% 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC) phospholipids, 25% unsaturated phosphatidylcholine (PC) lipids such as 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC), 9% phosphatidylglycerol (PG) lipids such as 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (POPG), 8% cholesterol, 4% of other phospholipids and 5% neutral lipids (Parra and Pérez-Gil, 2015). As illustrated in Figure 1.2, key surfactant proteins make up for approximately 10% of human surfactant, with Sp-A being the most abundant surfactant protein, followed by Sp-B, Sp-C and finally Sp-D.

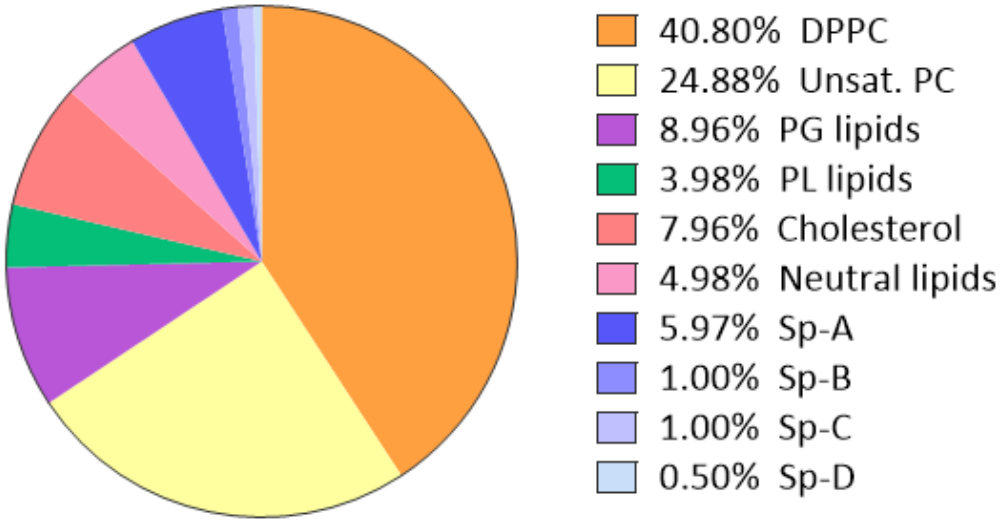


Figure 1.2: The lipid and protein composition of human pulmonary surfactant, where percentages are calculated with respect to total mass. **Abbreviations:** DPPC (dipalmitoyl phosphatidylcholine); Unsaturated PC (phosphatidylcholine); PG (phosphatidylglycerol); Other PL (phospholipids); Chol (cholesterol); Other NL (neutral lipids); Sp-A/B/C/D (Surfactant protein A/B/C/D). Created in GraphPad Prism based on data from Parra and Pérez-Gil, 2015.

1.2.1 The key surfactant lipids

Phosphatidylcholines (PC) are lipids of the lecithin group, which are characterised by a choline headgroup, glycerophosphoric acid and an array of saturated and unsaturated fatty acid chains (Agassandian and Mallampalli, 2013). PC lipids make up approximately 80% of all surfactant lipids, where a decrease in surface tension at this air-liquid interface is largely aided by the formation of a surface film, consisting of the major surface-active phospholipid DPPC which has two saturated tails (Wüstneck *et al.*, 2003). Phosphatidylinositol (PI) lipids contain a glycerol backbone, two non-polar fatty acid tails and an inositol polar headgroup. It comprises a non-polar and polar region, making it amphiphilic. Phosphorylated PI lipids (phosphoinositides) usually contains stearic acid in the SN_1 position and arachidonic acid in the SN_2 position. The inositol headgroup may be phosphorylated by kinases, which confer the lipid roles in both lipid and cell signalling. Phosphatidylglycerol (PG) lipids consists a glycerol substituent headgroup bonded through a phosphomonoester, an L-glycerol 3-phosphate backbone, and an ester bond to either unsaturated or saturated fatty acids on carbons 1 and 2. PG lipids are surfactant precursors, which give an indication of neonatal lung maturity, and is hypothesised to be

essential in the spread of surfactant over the type I cell surface area (King and MacBeth, 1981).

Cholesterol is a sterol lipid produced in all animal cells as a precursor to synthesise steroid hormones, vitamin D and bile acid. In pulmonary surfactant, cholesterol is essential in promoting rapid interfacial adsorption at the air-liquid interface and the formation of surface-associated structures (De La Serna *et al.*, 2004). The concentration of cholesterol in pulmonary surfactant also defines the lateral structure of the surfactant bilayer – when cholesterol makes up around 11% total dry weight of pulmonary surfactant, a single liquid-ordered (L_o) lamellar phase is adopted at physiological temperatures. An increase in physiological temperature would cause this phase to gradually adopt a liquid-disordered (L_d) lamellar phase. If cholesterol is absent, surfactant segregates into a lamellar gel phase and a L_d lamellar phase (Andersson *et al.*, 2017).

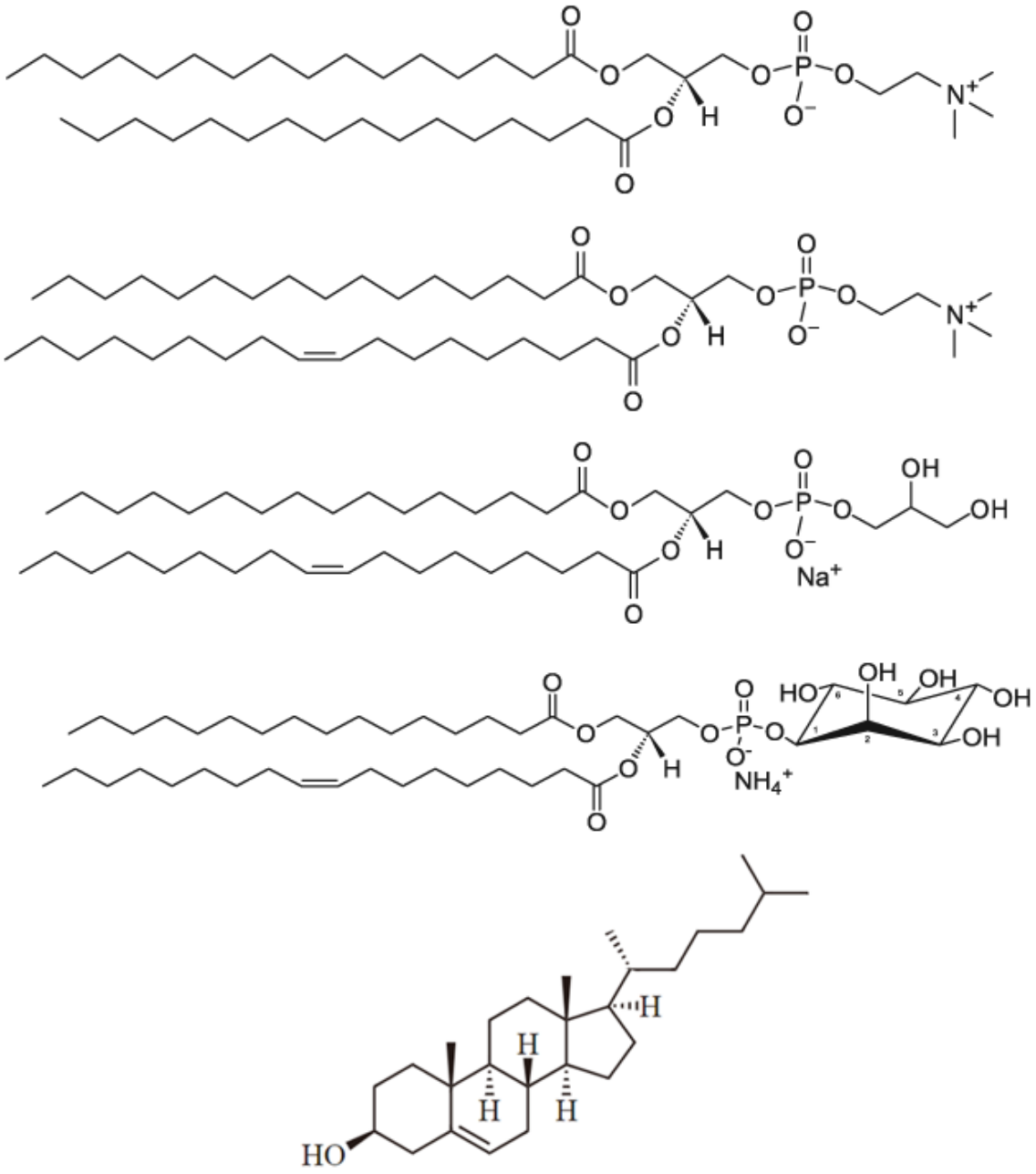


Figure 1.3: Chemical structure of lipids commonly found in pulmonary surfactant. **Top to bottom:** 1,2-dipalmitoyl-sn-glycero-3-phosphocholine (DPPC), 1-palmitoyl-2-oleoyl-glycero-3-phosphocholine (POPC), 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) (POPG), 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol (ammonium salt) (POPI) and cholesterol. Taken from Avanti Lipids (https://avantilipids.com/).

1.2.2 The surfactant proteins

The lungs face complex physical and environmental challenges, by coping with dynamic changes in volumes and pressures, as well as constant exposure to a variety of foreign particles and pathogens. This is regulated by pulmonary surfactant situated at the air-liquid interface.

The four surfactant proteins associated with pulmonary surfactant have differing compositions and structures catering to regulating the biophysical and immune function of the lungs. Surfactant protein A (Sp-A) (M_w 35 kDa in monomeric state) (Vieira, Kung and Bhatti, 2017) and surfactant protein D (Sp-D) (M_w 43 kDa) are glycoproteins that are members of the Ca^{2+}-dependent carbohydrate-binding collectin family involved in host defence mechanisms, such as inhibiting the activity of foreign pathogens. These proteins are also largely hydrophilic in nature, and often exist as multimers. Surfactant protein B (Sp-B) (M_w 8.7 kDa) and surfactant protein C (Sp-C) (M_w 4.2 kDa) however, are much smaller and predominantly hydrophobic proteins (Parra and Pérez-Gil, 2015) involved in biophysical functions of surfactant. These include the regulation of interfacial lipid adsorption, reversible lipid sequestration into the surfactant hypophase and lipid recruitment towards the surface to maintain the integrity and surface tension of the lipid film (Orgeig *et al.*, 2010).

Sp-B causes lipid arrangement in the surfactant lining, which reduces surface tension at the air-liquid interface in alveoli hence minimising the mechanical forces that collapses the lungs which prevents alveolar collapse (Schurch, Goerke and Clements, 1976). Incorporating Sp-B into the lipid component of surfactant promotes interfacial adsorption, film stability and the capacity to re-spread while surfactant lipids ensure the maximal reduction of surface tension upon lung compression (Serrano and Pérez-Gil, 2006).

1.2.3 Structure and function of surfactant proteins A and D

Sp-A and Sp-D are high-molecular weight, hydrophilic proteins belonging to the collectin protein family, contributing to the immune function of the lungs. Sp-A. the most abundant surfactant protein, comprises 248 residues, and is involved in immune defence against respiratory pathogens. Mature Sp-A1 (M_w 35 kDa in monomeric state) differs from Sp-A2 in the coding region by four amino acids, giving them a 98% positive sequence identity (244-248 amino acids). The human Sp-A locus contains both the *SFTPA1* and *SFTPA2* genes in the long arm q of Chromosome 10, each encoding Sp-A1 and Sp-A2 respectively. There are more than 30 variants characterised for both genes, the most abundant being the Sp-A1 variants 6A, $6A^2$, $6A^3$, $6A^4$ and the Sp-A2 variants 1A, $1A^0$, $1A^1$, $1A^2$, $1A^3$, $1A^5$ (Wang *et al.*, 2019). *SFTPA2* is mainly produced in the lung, but the presence of both *SFTPA1* and *SFTPA2* have been detected in other tissues such as the colon, eye, pancreas, trachea, and salivary glands. To produce mature Sp-A in the lungs, both Sp-A1 and Sp-A2 must be expressed, where the Sp-A complex contains two Sp-A1 gene products and a single Sp-A2 gene product per trimer (Vieira, Kung and Bhatti, 2017). In the presence of Ca^{2+}, Sp-A1 and Sp-A2 binds to surfactant phospholipids to lower the surface tension at the alveolar air-liquid interface, and Sp-A1 enhances the expression of Sp-A receptor Sp-R210 and the myosin-XVIIIa protein in alveolar macrophages.

Like Sp-A, Sp-D has an immunoregulatory role in the lungs and is encoded by the *SFTPD* gene on Chromosome 10. As shown in Figure 1.4, Sp-A and Sp-D's primary structure comprises 4 defined structural domains, namely a proline-rich collagenous region, an N-terminal region with inter-chain disulfide bonds, a hydrophobic neck region and a carbohydrate recognition domain that is highly homologous to proteins of the collectin family (McCormack, 1998). These proteins possess both a collagen-like region and a C-type lectin domain, making them pattern recognition receptors (PRRs) that contribute to their function for innate immunity in the lungs where they recognise pathogen and disease-associated molecular patterns (PAMPs and DAMPs) (Watson *et al.*, 2018).

MWLCPLALNLILMAASGAACEVKDVCVGSPGIPGTPGSHGLPGRDGRDGVKGDPGPPGPMGPPGET
PCPPGNNGLPGAPGVPGERGEKGEAGERGPPGLPAHLDEELQATLHDFRHQILQTRGALSLQGSIMTV
GEKVFSSNGQSITFDAIQEACARAGGRIAVPRNPEENEAIASFVKKYNTYAYVGLTEGPSPGDFRYSDGT
PVNYTNWYRGEPAGRGKEQCVEMYTDGQWNDRNCLYSRLTICEF

Signal peptide (1-20) Sp-A mature peptide chain (21-248)

MLLFLLSALVLLTQPLGYLEAEMKTYSHRTMPSACTLVMCSSVESGLPGRDGRDGREGPRGEKGDPGLP
GAAGQAGMPGQAGPVGPKGDNGSVGEPGPKGDTGPSGPPGPPGVPGPAGREGPLGKQGNIGPQG
KPGPKGEAGPKGEVGAPGMQGSAGARGLAGPKGERGVPGERGVPGNTGAAGSAGAMGPQGSPGA
RGPPGLKGDKGIPGDKGAKGESGLPDVASLRQQVEALQGQVQHLQAAFSQYKKVELFPNGQSVGEKI
FKTAGFVKPFTEAQLLCTQAGGQLASPRSAAENAALQQLVVAKNEAAFLSMTDSKTEGKFTYPTGESLV
YSNWAPGEPNDDGGSEDCVEIFTNGKWNDRACGEKRLVVCEF

Signal peptide (1-20) Mature Sp-D peptide chain (21-375)

Figure 1.4: Amino acid sequences of full-length surfactant proteins obtained from UniProt. **Top to bottom:** Pulmonary surfactant-associated protein A1 (252 residues, UniProt ID Q8IWL2), pulmonary surfactant-associated protein A2 (252 residues, UniProt ID Q8IWL1) and pulmonary surfactant-associated protein D (381 residues, UniProt ID P35247).

Native Sp-A obtained from the bronchoalveolar lavages consist primarily of 6 Sp-A trimers bound by van der Waals forces, forming octodecamers of approximately 630 kDa. Similarly, each trimeric Sp-D (M_w 129 kDa) is made up of 3 monomeric subunits of Sp-D monomers (M_w 43 kDa). As seen in Figure 1.5, Sp-D can form a hexameric, dodecameric and 'fuzzy ball' structure which are anti-inflammatory in nature. The conserved cysteine residues Cys15 and Cys20 located in the N-terminal region stabilises the oligomerisation of trimeric Sp-D via inter-chain disulfide crosslinks, causing the hydrophobic tails to be buried with the hydrophilic heads exposed (Guo *et al.*, 2008).

Generally, mature Sp-A functions in the octadecameric form except when bound to detergent and lipid micelles mimicking pulmonary surfactant phospholipids, in which they function in a smaller oligomeric form. There is evidence of Sp-A interacting with Sp-B in enhancing phospholipid adsorption in the air-liquid interface (Warr *et al.*, 1987),

inducing Ca^{2+}-dependent aggregation of surfactant lipid vesicles regardless of the presence of Sp-B or Sp-C. Both Sp-A and Sp-B induce the lateral redistribution of hydrophobic surfactant proteins in lipid films and are essential for tubular myelin assembly in alveolar spaces (Martínez-Calle *et al.*, 2019).

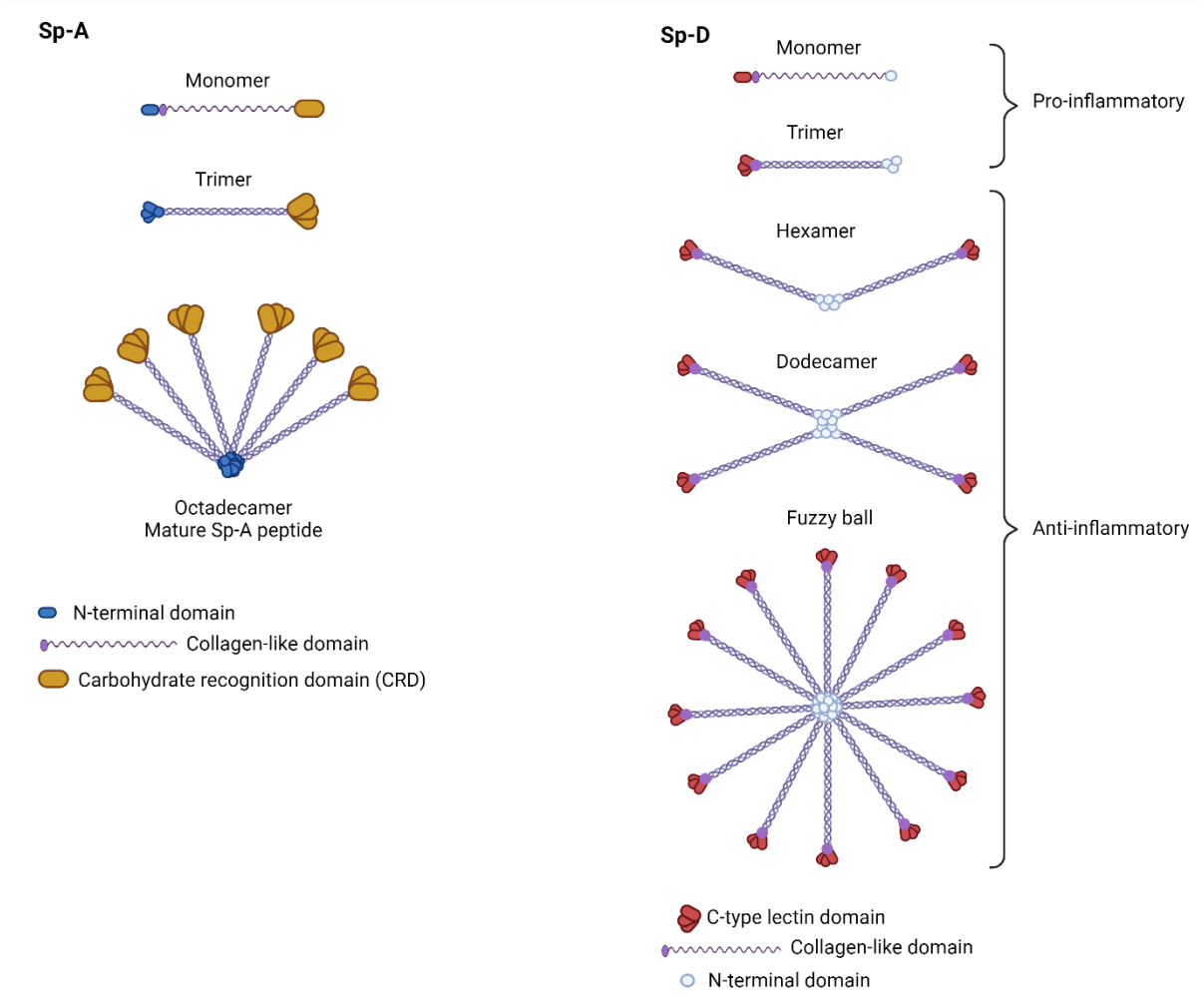


Figure 1.5: Structure of the hydrophilic surfactant proteins Sp-A and D. Sp-A and Sp-D share four common structural domains (proline-rich collagenous domain, N-terminal domain, hydrophobic neck region, carbohydrate recognition domain (CRD). Three Sp-A monomer subunits form a Sp-A trimer, where six trimers form the mature Sp-A octadecamer (King and Chen, 2020) Similarly, the multimerisation of Sp-D allows it to gain anti-inflammatory properties, with the docadecameric structure being the most common. Acidic environments cause Sp-D monomerisation or trimerisation, which have pro-inflammatory properties. The presence of pathogens causes ligand occupation of the C-type lectin domain, which blocks anti-inflammatory signaling (Colmorten, Nexoe and Sorensen, 2019). Created in BioRender.

1.2.4 Structure and function of surfactant proteins B and C

Sp-B and Sp-C are low-molecular weight, lipophilic proteins necessary for the biophysical functions of surfactant phospholipids. As seen in Figure 1.7, both proteins exist in close association with the surfactant phospholipid monolayer. Human Sp-C is expressed as a 197 residue proprotein encoded by the *SFTPC* gene on Chromosome 8. It consists of 2 unique domains, namely the N-terminal propeptide and C-terminal BRICHOS domain. Residues 1-23 and 59-197 are cleaved off, leaving residues 24-58 producing the 35-residue Sp-C protein (M_w 3.7 kDa), as shown in Figure 1.6. Mature Sp-C contains 2 thioester-linked palmitoyl groups (M_w 4.2 kDa), covalently linked fatty acyl chains and is made of approximately 70% non-polar residue hence giving it its highly hydrophobic nature (Johansson, 1998). It exists in many isoforms, with the most abundant form being the biologically active molecule, and minor forms providing insight into its surfactant activity. Sp-C isoforms containing one or two thioester-palmitoyl chains missing were found, showing the unstable nature of the thioester bond. Additionally, an isoform with a palmitoyl group linked to the amino acid Lys-11 making up approximately 4% of the main isoform was identified, likely formed by the transfer of an acyl group via neighbouring phospholipids present in pulmonary surfactant (Johansson, 1998).

Sp-B is a saposin-like protein with a molecular weight of approximately 8.7 kDa. It is generated from its 381-residue precursor, pro-Sp-B, which is proteolytically cleaved into the mature 79 residue Sp-B highlighted in yellow below. Sp-B comprises 52% hydrophobic amino acids with a conserved negatively charged residue and 8 conserved positively charged residues (Olmeda, García-Álvarez and Pérez-Gil, 2013), contributing to its amphipathic nature. Sp-B exists as a covalently linked homodimer with a molecular weight of 17.4 kDa, with an inter-subunit disulfide bond at the Cys-48 residue (Cys48-Cys48'), allowing Sp-B to crosslink various lipid membranes and distinguishing it from its structural saposin protein homologues (Zaltash *et al.*, 2000). Members of the saposin family share a common structural feature of three conserved intramolecular cysteine bridges. In comparison, Sp-B is more hydrophobic and contain 7 cysteines in its primary structure, with six highly conserved cysteines (highlighted in bold in the amino acid sequence of the mature Sp-B chain) involved in the formation of 3 intramolecular disulfide bridges forming at Cys8-Cys77, Cys11-Cys71 and Cys35-Cys46 which stabilises its tertiary structure (Serrano and Pérez-Gil, 2006), and the 7th cysteine allowing the

covalent dimerisation of the protein. Circular dichroism (CD) spectroscopy and Fourier transform infrared spectroscopy of Sp-B have shown a high alpha helical content of around 40-60% in its secondary structure (Vandenbussche *et al.*, 1992), however, a crystal structure of full-length human Sp-B has not been determined due to its highly hydrophobic nature and challenges in producing a recombinant form of the protein.

MDVGSKEVLMESPPDYSAAPRGRFGIPCCPVHLKRLLIVVVVVVLIVVVIVGALLMGLHMSQKHTEMVL
EMSIGAPEAQQRLALSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCYIMKIAPESIPSLEALNRKV
HNFQMECSLQAKPAVPTSKLGQAEGRDAGSAPSGGDPAFLGMAVNTLCGEVPLYYI

N-terminal propeptide (1-23) C-terminal propeptide (59-197)

Mature Sp-C peptide (24-58)

MAESHLLQWLLLLLPTLCGPGTAAWTTSSLACAQGPEFWCQSLEQALQCRALGHCLQEVWGHVGAD
DLCQECEDIVHILNKMAKEAIFQDTMRKFLEQECNVLPLKLLMPQCNQVLDDYFPLVIDYFQNQTDSN
GICMHLGLCKSRQPEPEQEPGMSDPLPKPLRDPLPDPLLDKLVLPVLPGALQARPGPHTQDLSEQQFPI
PLPYCWLCRALIKRIQAMIPKGALAVAVAQVCRVVPLVAGGICQCLAERYSVILLDTLLGRMLPQLVCRL
VLRCSMDDSAGPRSPTGEWLPRDSECHLCMSVTTQAGNSSEQAIPQAMLQACVGSWLDREKCKQFV
EQHTPQLLTLVPRGWDAHTTCQALGVCGTMSSPLQCIHSPDL

Signal peptide (1-24) Mature Sp-B peptide (201-279)

N-terminal propeptide (25-200) C-terminal propeptide (280-381)

Figure 1.6: Amino acid sequences of full-length surfactant proteins obtained from UniProt. **Top to bottom:** Pulmonary surfactant-associated protein C (197 residues, UniProt ID P11686); the two Cys residues highlighted in yellow form a thioester bond to the palmitic chains. Pulmonary surfactant-associated protein B (381 residues, UniProt ID P15285).

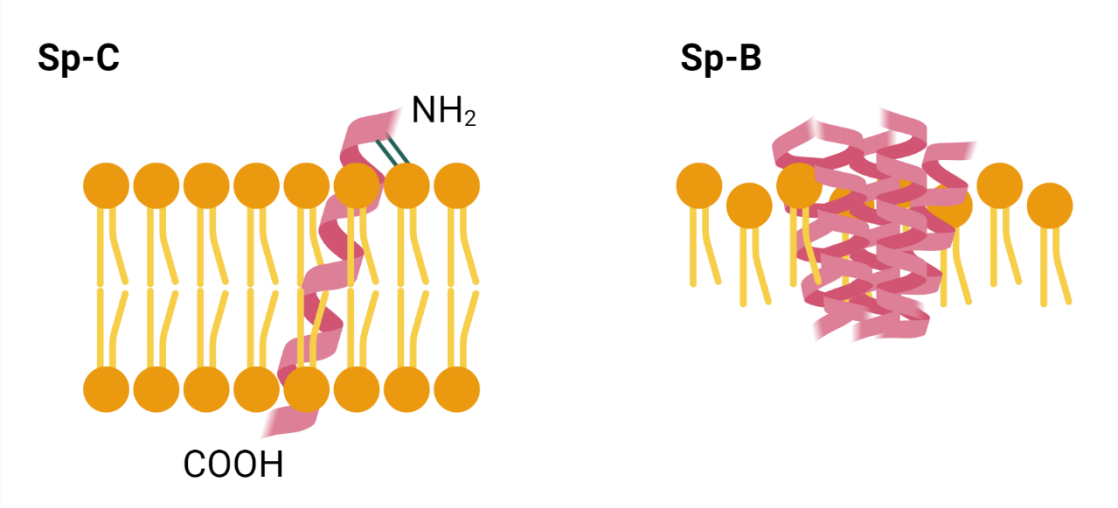


Figure 1.7: Structure of the hydrophobic surfactant proteins Sp-B and C. The hydrophobic surfactant proteins Sp-B and Sp-C exist embedded in the surfactant phospholipid monolayer (Christmann et al., 2009). Sp-B is hypothesised to associate closely with surfactant lipids, where a compact globular multimer containing five monomeric alpha helices of mature Sp-B is formed (Olmeda, García-Álvarez and Pérez-Gil, 2013). Created in BioRender.

A study by Yang *et al.* in mouse models showed that the N-terminal Sp-B propeptide contained 3 saposin-like protein (SAPLIP) domains, which relates to the surfactant function of the mature Sp-B peptide. Interestingly, the N-terminal propeptide of Sp-B reinforced the anti-microbial function of macrophage cells, by promoting bacterial uptake in mice alveoli. The study also suggested that the propeptide inhibited bacterial growth at lower pH, although its mechanisms remain largely unknown (Yang et al., 2010). Another study found that mature Sp-B peptides have cytolytic properties such as increasing membrane permeability *in vitro*, however this is inhibited when the peptide associates with surfactant lipids. This reinforces that endogenous Sp-B does not play a major role in the immune function in the alveoli. However, synthetic Sp-B derivatives with the helix containing residues 7 to 22 were seen to retain significant antimicrobial activity in native surfactant (Ryan *et al.*, 2006).

1.2.5 Saposin-like proteins, SAPLIP

As stated earlier, a full length, three-dimensional structure of Sp-B has yet to be produced recombinantly due to the protein's highly hydrophobic nature. Proteins of the saposin-like (SAPLIP) family with a high sequence alignment with Sp-B have however allowed the prediction of theoretical models of SP-B, based on data extrapolation (Zaltash *et al.*, 2000). These proteins have key structural features allowing them to interact with lipids, such as having a conserved hydrophobic amino acid distribution, 3 intramolecular disulfide bridges formed from 6 conserved cysteines at specific positions and the presence of 4 to 5 amphipathic alpha helices (Olmeda, García-Álvarez and Pérez-Gil, 2013).

Saposins A, B, C and D are highly homologous, membrane active proteins that share the common saposin structural fold, shown in Figure 1.8, characterised by disulfide bonding of amphipathic alpha helices, one of which includes the C-terminal helices (highlighted in red), N-terminal helices (highlighted in blue), and the other containing two or three central helices (highlighted in green and yellow), allowing them to exist in the lipid-bound and soluble states. They share a common precursor prosaposin, with the mature protein binding to lipids to activate lysosomal enzymes needed for sphingolipid degradation. The conformational flexibility of saposins allow them to bind lipids in their hydrophobic cavities whilst in the closed conformation, which also allows the extraction of lipids from the membrane and subsequently, enzyme hydrolysis.

Saposin A is an 85-residue protein with a molecular weight of 9.3 kDa. In the absence of lipid interaction, it adopts a closed monomeric apo conformation (Figure 1.8, saposin A (2DOB)) and in the presence of detergent, the lipid-bound protein assembles into a dimeric, open conformational form (Figure 1.8, saposin A (4DDJ)) with a hydrophobic core solubilising lipids into monodisperse particles (Popovic *et al.*, 2012). Saposin B, an 81-residue protein with a molecular weight of 9.13 kDa, has a shell-like dimer, see Figure 1.8, enclosing a hydrophobic cavity. It adopts an open V shaped conformation, with the hydrophobic side of the helices facing the inner surface of the V. The Y54 kink refers to the conserved tyrosine 54 residue determining the conformational flexibility of the saposin proteins, which in turn affects the protein's interfacial activity. Saposin B has 2 conformations which differ in the Y54 kink, causing a higher bending of the V in the second chain (Ahn *et al.*, 2006). On the other hand, saposin C, an 83-residue protein with

a molecular weight of 9 kDa formed trimers at pH 4.8 in the presence of detergent. Both saposins B and C possess conformational flexibility, where a variation in chain conformation produces 2 different types of dimers, both having an inner hydrophobic cavity either having a closed or open formation depending on their chain conformations. Saposin D, an 80-residue protein with a molecular weight of 9.13 kDa is proposed to exist as a dimer stabilised with polar interactions between conserved residues, which prevents non-specific binding of substrates to the protein prior to its association with lysosomal membranes and lipid solubilisation in low pH conditions (Rossmann *et al.*, 2008). In the closed conformation, the globular packing of saposins A, C and D causes the hydrophobic residues of the alpha helices to be buried and therefore stable in aqueous environments.

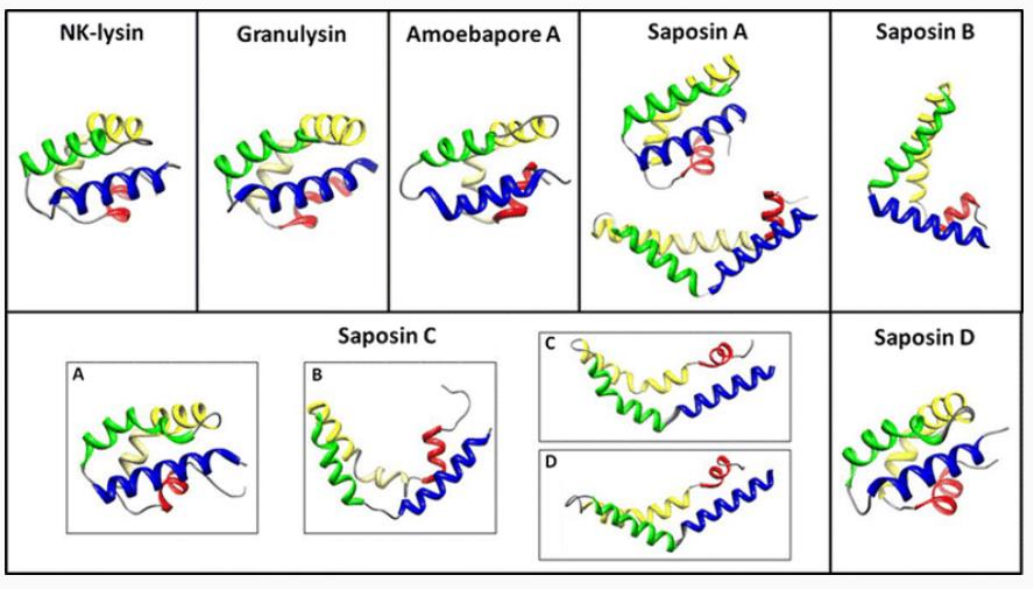


Figure 1.8: 3D structures of SAPLIP proteins, with C-terminal helices in red, and N-terminal helices in blue. NMR-solved pig NK-lysin (1NKL), crystallised human granulysin (1L9L), NMR amoebapore A from *Entamoeba histolytica* (1OF9), crystallised soluble human saposin A (top) (2DOB), crystallised human saposin A in complex with LDAO (bottom) (4DDJ), crystallised human saposin B (chain A) (1N69), NMR human saposin C (A) (1M12), NMR saposin C in SDS micelles (B) (1SN6), tetragonal crystals of saposin C (C) (2Z9A), orthorhombic crystals of saposin C (D) (2QYP) and crystallised human saposin D (2RB3) (Ahn et al., 2006; Rossmann et al., 2008; Popovic et al., 2012; Olmeda, García-Álvarez and Pérez-Gil, 2013).

The saposins can alternate its structure between the open and closed conformations essential to their function. For instance, saposin A in the open state allows effective

substrate solubilisation for lipid hydrolysis (Popovic *et al.*, 2012). Saposins A, B, C and D are acidic in nature, where the net charges for each protein at pH 7 are -8.2, -8.0, -9.1 and -8.0 respectively (Bruhn, 2005; Ahn *et al.*, 2006). NK lysin and granulysin however, have many positively charged residues leading to a net charge of +7.5 and +11 respectively (Bruhn, 2005). Arginine residues in granulysin mediate contact with anionic lipids and its hydrophobic residues orientate the protein to allow the insertion of tryptophan (Trp41) into the lipid membrane (Anderson *et al.*, 2003). The Trp41 residue is in Helix 3 of the N terminus of granulysin, where granulysin's orientation is mediated by charges on the membrane surface to facilitate the insertion of Trp41 into the membrane. Tryptophan aids the anchoring of phospholipase A_2, and NK lysin is seen to partially bury its tryptophan (Trp58) residue into the membrane without losing its secondary structure. The interaction of amoebapores with the membrane requires an acidic pH, mediated by the conserved H75 residue, without exposing cationic clusters (groups of positively charged residues) due to their hydrophobic surface (Bruhn *et al.*, 2003; Hecht *et al.*, 2004). Human Sp-B has 79 residues, as shown in Figure 1.6, with leucine (14 residues), valine (9 residues), alanine (8 residues), cysteine (7 residues) and arginine (7 residues) being the most abundant amino acids present. It has 1 negatively charged and 8 positively charged residues, leaving it with a net charge of +7 at pH 7. Since the degree of membrane association of SAPLIP proteins rely on their charge distributions, some insight can be drawn into Sp-B's interactions with anionic PG and zwitterionic PC lipids in the lipid bilayer.

1.2.6 Using saposin-like proteins (SAPLIPs) for the structure prediction of Sp-B

Homology modelling allows the theoretical prediction of the three-dimensional structure of Sp-B, based on its similarity to closely related solved structures of SAPLIP proteins. Figure 1.9 illustrates the protein sequence alignments of Sp-B to SAPLIP proteins of known structure, providing insight into the structure of Sp-B. Both SAPLIP proteins and Sp-B show the presence of highly conserved cysteine residues (highlighted in yellow) at 6 positions (Cys 8, 11, 31, 46, 71, 77). Human granulysin however, has just 5 cysteine residues, with Cys8 missing. The formation of disulfide bridges between Cys8-Cys77,

Cys11-Cys71 and Cys31-Cys46 generates the saposin fold which are characteristic to SAPLIP proteins.

Figure 1.9 also show the grand average hydropathicity (GRAVY) index values of the peptides, which calculates the total hydrophobicity of a peptide from the total sum of hydropathy values of all residues, divided by the length of the sequence. Hydrophobic proteins have positive GRAVY values and hydrophilic proteins have negative values. Sp-B's GRAVY value of 1.027 implies that it is a largely hydrophobic protein, which is far larger than the values of the SAPLIP proteins shown in Figure 1.9.

Since homology modelling is heavily reliant on the sequence similarity between a known structure of closely related proteins and the protein model, the percentage similarities between the SAPLIP proteins and human Sp-B were calculated using the BLAST sequence alignment tool and shown in Table 1.1. As can be seen, the closest similarity to Sp-B is 28.57%. Ideally, a sequence alignment similarity of about 30-35% can predict a protein structure up to an accuracy of a low-resolution X-ray structure (Xiang, 2006) but a reasonable model of Sp-B can be generated from a model with less than 30% sequence identity since Sp-B is related to a large saposin-like protein family where a good sequence alignment can be defined.

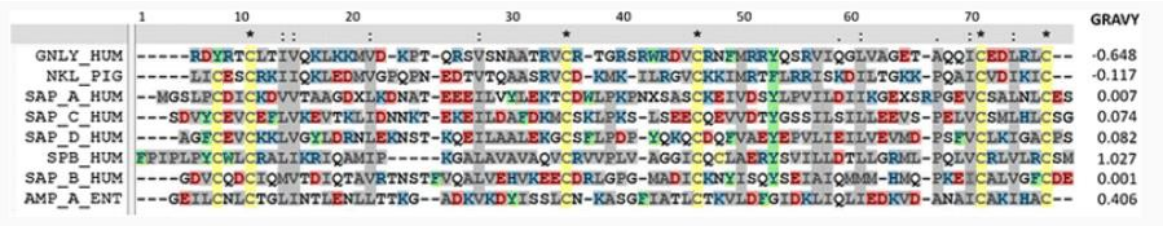


Figure 1.9: Sequence alignment of SAPLIP proteins and human Sp-B generated by CLUSTAL-W. Top to bottom: Human granulysin (GNLY_HUM), pig NK-lysin (NKL_PIG), human saposin proteins A/B/C/D (SAP_A/B/C/D_HUM), human Sp-B (SPB_HUM) and amoebapore A from *Entamoeba histolytica* (AMP_A_ENT). Cysteine residues are highlighted in yellow, aromatic residues in green, hydrophobic residues in grey, positively charged residues in blue and negatively charged residues in green. Asterisks represent identical residues, while colons and dots show conserved (.) and semi-conserved (:) substitutions respectively. Numberings 1-79 corresponds to the Sp-B sequence, and the grand average hydropathicity (GRAVY) values were calculated to define how hydrophilic or hydrophobic the protein is (Olmeda, García-Álvarez and Pérez-Gil, 2013a).

SAPLIP protein	Percentage identity (%) when aligned with SP-B using BLAST
GNLY_HUM (Human granulysin)	20.31%
NKL_PIG (Porcine NK-lysin)	24.64%
SAP_A_HUM (Human saposin A)	22.08%
SAP_C_HUM (Human saposin C)	27.63%
SAP_D_HUM (Human saposin D)	16.22%
SAP_B_HUM (Human saposin B)	28.57%
AMP_A_ENT	No significant similarity

Table 1.1: Sequence identity (%) of Sp-B aligned against SAPLIP proteins, calculated using the BLAST sequence alignment tool. Obtained from (Wang, 2019).

Despite the sequence alignment similarity of Sp-B with the SAPLIP proteins, Sp-B covalently dimerises and is permanently associated to the phospholipid membranes of pulmonary surfactant. A sequence alignment of Sp-B with NK-lysin showed Sp-B arranging into four amphipathic alpha helices, with a model of Sp-B generated based on monomeric NK-lysin's three-dimensional structure. This model predicts Sp-B having a closed, compact globular saposin fold consisting of five packed monomeric alpha helices, with a disulfide bridge between the C48 residue of each monomer, with salt bridges forming from charged residues E51 and R52 (Zaltash *et al.*, 2000). In the dimeric state, two molecules of Sp-B form a disulfide bridge at residues C48-C48'. An absence of the C48 residue resulted in the formation of monomeric Sp-B which relies on concentration to maintain proper function, suggesting the formation of non-covalent Sp-B dimers involving the reciprocal ion pairs E51 and R52 (Beck *et al.*, 2000; Ikegami, Takabatake and Weaver, 2002). Fourier-transform infrared spectroscopy (FTIR) spectroscopy and molecular modelling of folds in the synthetic Sp-B peptide Mini-B supported the model of Sp-B in the closed conformation, but a lack of a second saposin fold in Mini-B resulted in insufficient evidence of Sp-B existing in the open conformation (Waring et al., 2005). However, an open conformation of Sp-B may occur in response to variations in surface tension due to the protein's conformational flexibility.

As supported by Olmeda *et al.*'s molecular dynamics (MD) simulations, SAPLIP proteins can adopt several water-soluble conformations, suggesting the possibility of protein-protein and protein-lipid interactions. The soluble monomer has a closed structure with

a hydrophobic inner cavity, initiating electrostatic interactions with the membrane surface. Membrane-associated monomers adopt a more open conformation, exposing the hydrophobic cavity to detergents/lipids. The presence of an open conformation also promotes protein dimerisation, forming a hydrophobic cavity. A conformational change of a closed dimeric protein into the open oligomeric state facilitates the exposure the saposin dimer's hydrophobic cavities to the lipid acyl chains. Saposin oligomers can permeabilise lipid membranes and expose the hydrophobic surfaces towards the lipid acyl chains, causing the charged polar surfaces of the oligomer core to allow the permeabilisation of various polar molecules through the membrane (Olmeda, García-Álvarez and Pérez-Gil, 2013).

It was proposed that Sp-B shared similar structures and conformations to SAPLIP proteins, as illustrated in Figure 1.10. The interaction of Sp-B with the phospholipid membrane was caused by the oligomerisation of Sp-B dimers and a conformational change from the closed to open state, allowing the protein to be loaded with lipid molecules. The loaded protein can then adopt a shallower orientation to facilitate both the interfacial adsorption and transfer of lipids at the air-liquid interface, hence reducing the alveolar surface tension. However, the large difference in GRAVY values between Sp-B and SAPLIP proteins, as previously stated, still implies that Sp-B is far more hydrophobic compared to SAPLIP proteins.

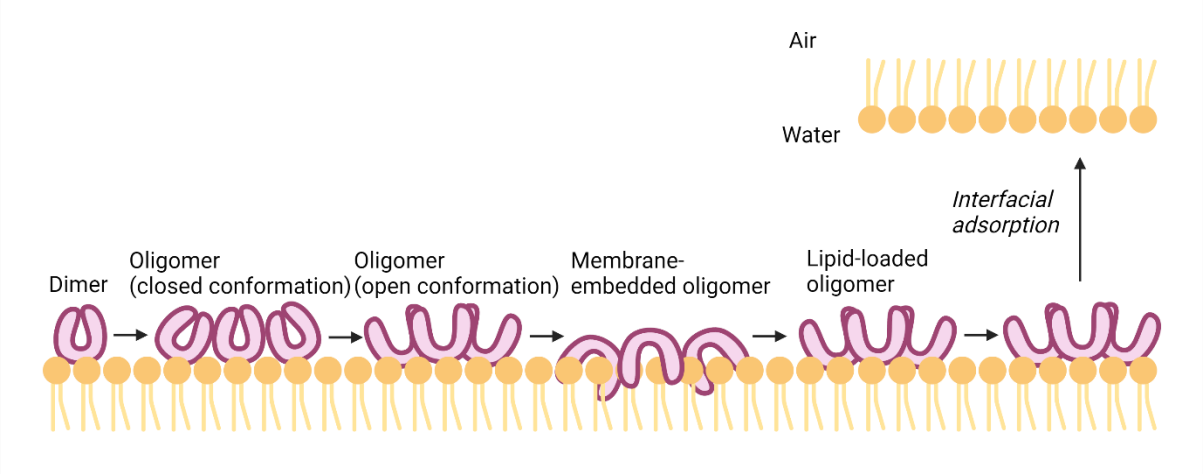


Figure 1.10: Possible conformations of permanently membrane-associated Sp-B derived from the conformations of SAPLIP proteins. The lipid loaded oligomeric form promotes Sp-B's biophysical function by facilitating lipid exposure at the air-liquid interface. Created in BioRender based on (Olmeda, García-Álvarez and Pérez-Gil, 2013).

1.3 Using AlphaFold2 to predict the structure of surfactant proteins

In recent years, the development of artificial intelligence-based computational software like AlphaFold2 has aided the prediction of protein structures based on their amino acid sequence. The Critical Assessment of Structure Prediction (CASP14), which assesses how accurate structure prediction methods are at predicting the structure of unsolved proteins, determined that AlphaFold2 had a high accuracy in predicting structures of protein backbones, domain structures and side chains (Jumper et al., 2021).

1.3.1 Mechanisms of AlphaFold2

AlphaFold2 starts by generating multiple sequence alignments (MSAs) of the protein. This compares its amino acid sequence with an extensive database of previously determined protein sequences, to detect homologous protein structures. The software then extracts specific pairwise features that provide insight into the target protein sequence and its structural homologues.

The input structure is processed through layers of a neural network block known as Evoformer. In each block, each column corresponds to the MSA of the target protein while each row corresponds to the sequences in which those residues appear. This produces a MSA and an array of several sequences (N_{seq}) multiplied by the number of residues (N_{res}). An MSA transformer processes the MSA columns to obtain information regarding the evolutionary relationships between the sequences. Evoformer also processes the extracted features for the prediction of pairwise angles and distances between the amino acid residues, which help to predict the 3D structure of the target protein.

The attention mechanism allows Evoformer to focus on relevant sections of the MSA and pairwise features to model long-range interactions between residues in the protein sequence. Additionally, Evoformer iteratively refines the prediction process by predicting the pairwise angles and distances of the protein structure, which repeats until the model converges on the most stable structure (Jumper et al., 2021).

1.3.2 Structure prediction of the surfactant proteins using AlphaFold2

After predicting the angles and distances of individual residues, a full 3D model of the protein is generated from the model that has the least violations of predicted constraints. A confidence measure known as the predicted Local Distance Difference Test (pLDDT) score is used to assess the reliability of the predicted structure of each residue in the model. This indicates how confident AlphaFold2 is in predicting the position of a particular residue in a 3D space. The pLDDT score ranges from 0 to 100, with 0 being the lowest confidence and 100 being the highest. Generally, a score above 90 would indicate that the predicted model has a high accuracy.

Figure 1.11 shows the pLDDT scores and the 3D structure of full length Sp-B, the mature Sp-B peptide and Sp-B_{1-25}. The pLDDT value was plotted against the residue number. The structures were generated from the predicted model with the best fit (rank_1). Generally, the alpha-helical structure of Sp-B_{1-25} predicted by AlphaFold2 with a pLDDT of 50 to 90, agrees with the structure obtained from the Protein Data Bank (PDB) (Gordon et al., 2000). AlphaFold2 predicts the mature Sp-B peptide with a pLDDT between 45 to 85 as being mostly alpha-helical. This agrees with the "loop" structure as proposed by Banfi *et al.*, where the alpha helix loops back at residues 40 and 41 to accommodate disulfide bonding between the cysteine residues 35 and 46, 11 and 71, and finally 8 and 77. This agrees with the structure (Banfi and Agostoni, 2016). In full-length Sp-B, AlphaFold2 predicts the signal peptide with low confidence (pLDDT ~40). Part of the N-terminal propeptide is predicted with high confidence (pLDDT ~80) however the last 50 residues of the propeptide have a pLDDT score of approximately 30. The software predicts most parts of the mature Sp-B peptide and C-terminal propeptide with high confidence (pLDDT ~80). This highlights the difficulty in predicting the 3D structure of Sp-B especially due to challenges faced in producing an experimentally determined structure of the protein (Martinez-Calle et al., 2020).

The AlphaFold2 predicted structure of mature Sp-B shared similar characteristics to the structure predicted by Olmelda *et al.*, where it was agreed that Sp-B had a predominantly alpha helical structure. However, AlphaFold2 predicts mature Sp-B to contain three alpha helices, instead of five in the latter study.

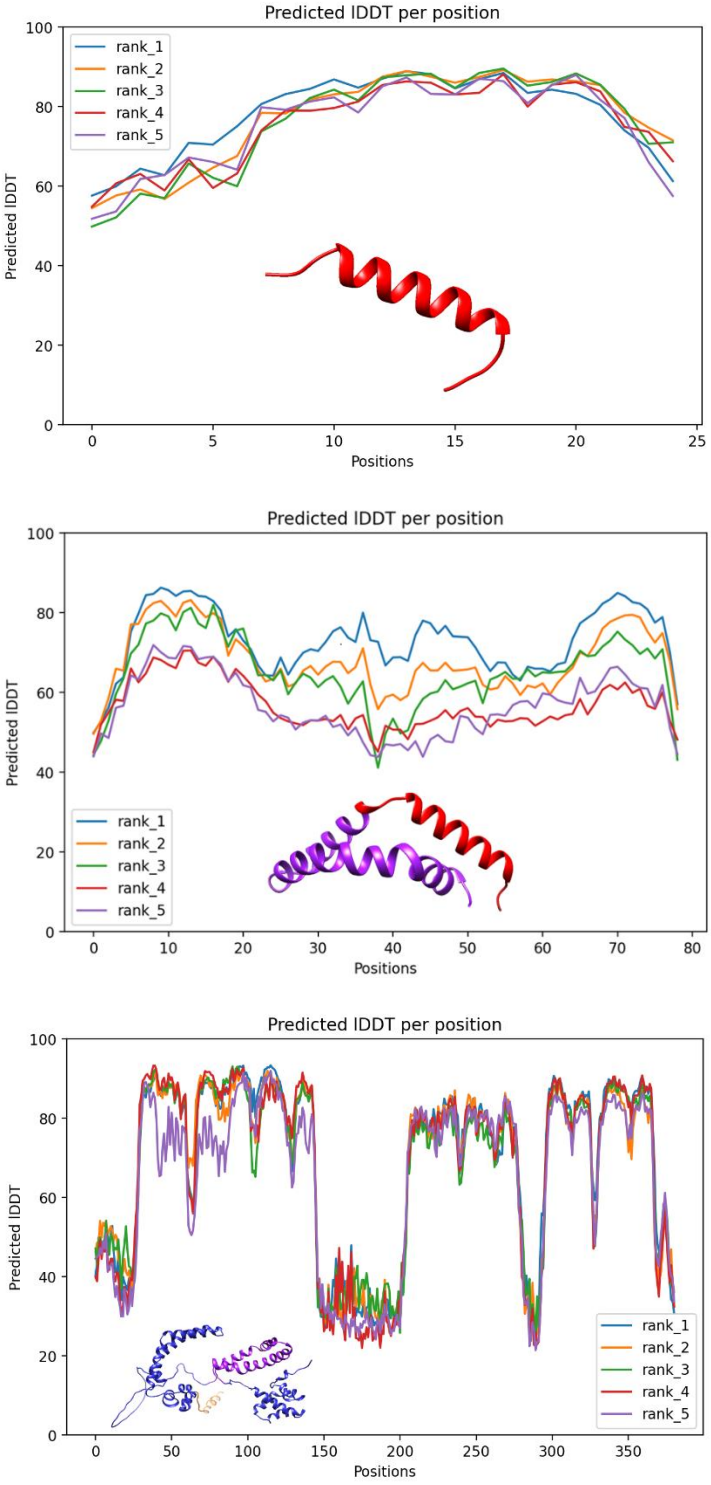


Figure 1.11: The pLDDT scores and 3D structures of Sp-B generated using AlphaFold2. **Top to bottom:** Sp-$B_{1\text{-}25}$ (red), mature Sp-B peptide (residues 26-79 in purple, residues 1-25 in red), full-length Sp-B (signal peptide in yellow, N and C terminal propeptides in blue, mature Sp-B peptide in purple).

1.4 Interactions between key surfactant proteins and membrane lipids

To fulfil the functions of pulmonary surfactant, it has been proposed that there are many key interactions between surfactant proteins and surfactant phospholipids. For instance, *in vitro* studies revealed functional interactions between peptide mimics of Sp-A and mini-B. Mini-B (1SSZ) is a 34-residue, biologically active synthetic peptide mimic of Sp-B (CWLCRALIKRIQAMIPKGGRMLPQLVCRLVLRCS). In the presence of lysomyristoylphosphatidylcholine (LMPC) and lysomyristoylphosphatidylglycerol (LMPG) micelles, the interaction of Sp-A with mini-B was analysed using nuclear magnetic resonance (NMR) techniques involving the ^{15}N-^{1}H HSQC spectra of Mini-B after adding Sp-A. It was discovered that the extent of their interaction depended on the lipid headgroups, ultimately resulting in a structural rearrangement of Sp-A. The absence of an Sp-A-induced chemical shift in Mini-B's HSQC peaks suggested that there was no direct interaction between Sp-A and Mini-B. Mini-B preferentially binds to Sp-A within micelles even in an excess of Sp-A-free micelles. Furthermore, diffusion-ordered spectroscopy (DOSY) data showed the interactions between Sp-A, Mini-B, and micelles, where the hydrodynamic diameter of Sp-A-Mini-B-micelle complexes were smaller than Sp-A-micelle complexes but larger than Mini-B-micelle complexes. This also suggested Mini-B affecting the oligomerisation of Sp-A when associating with LMPC/LMPG micelles (Sarker, Jackman, and Booth, 2011).

Sp-A interacts beneath the interfacial lipid film and induces lateral redistribution of hydrophobic proteins like Sp-B, while Sp-B enhances the adsorption of Sp-A onto the lipid films. Sp-A and Sp-B also work synergistically to reduce alveolar surface tension, where both surfactant proteins are essential for the formation of tubular myelin (Weaver and Whitsett, 1991). In a study by Taneva and Keough, both Sp-B and Sp-C were found to enhance the adsorption of Sp-A to the lipid-protein film when pure protein monolayers were formed, or when the proteins were supplemented into DPPC:PG films. Hence, it could be assumed that the interactions between Sp-A/Sp-B and Sp-A/Sp-C were affected by lipid-induced alterations in the protein conformations. CD and external reflection absorption infrared spectroscopy showed Sp-A adsorbing to DPPC monolayers, which promoted interactions between Sp-A molecules which subsequently formed extensive protein fibrous networks. Furthermore, studies using transmission electron microscopy showed the binding of Sp-A to the lipid monolayer resulted in structural changes to Sp-A

induced by cations (Ca^{2+}) and lipids. On the other hand, Sp-B and Sp-C were shown to exist stably in lipid monolayers. This suggested the conformational changes in SP-A's structure accompanied the adsorption of Sp-A to the lipid-protein monolayer. Interactions of Sp-A with Sp-B only occurred in the presence of lipids, and measurements of the monolayer revealed no specificity in the interactions of Sp-A with Sp-B or Sp-C (Taneva and Keough, 2000).

1.4.1 Interaction of cholesterol with the membrane monolayer

The dynamic lateral structure of surfactant membranes and monolayers depend on its composition of lipids such as cholesterol. Cholesterol aids in lateral phase segregation of the liquid-ordered (L_o) and liquid-disordered (L_d) phases in surfactant, and effectively improves the spreading capability of lipids in surfactant at the air-liquid interface (Kim *et al.*, 2013). The L_o phase found in PC monolayers contain high concentrations of cholesterol, which causes a more ordered and compact formation of the acyl chains (Figure 1.12), promoting mobility of the lipids laterally and restricting small molecules from moving across the monolayer. L_d monolayers, however, have irregularly packed lipid molecules and unsaturated acyl chains, where the kinks of the acyl chains weaken the Van der Waals forces between the chains. The highly fluid nature of the L_d phase allows lipids to move laterally across the membrane surface.

Cholesterol is crucial in defining the lateral structure of pulmonary surfactant layers at mammalian physiological temperatures of about 37 °C. The spreading properties of native surfactant is largely influenced by the presence of cholesterol, suggesting that cholesterol plays a complementary role in aiding surfactant lipids or proteins like Sp-B and Sp-C in regulating the biophysical function of surfactant. In a study be De La Serna *et al.*, fluorescently labelled Sp-B and Sp-C were seen to reside in the L_d phase of the lipid monolayer. It was observed that suspensions lacking surfactant proteins resulted in decreased adsorption rates of lipids, proving that the hydrophobic proteins Sp-B and Sp-C were crucial in catalysing the interfacial transfer of phospholipids. Furthermore, a depletion of cholesterol in the lipid fraction of surfactant resulted in further impairment of the interfacial transfer, showing that the presence of cholesterol improved interfacial dynamics in surfactant structures (De La Serna *et al.*, 2004).

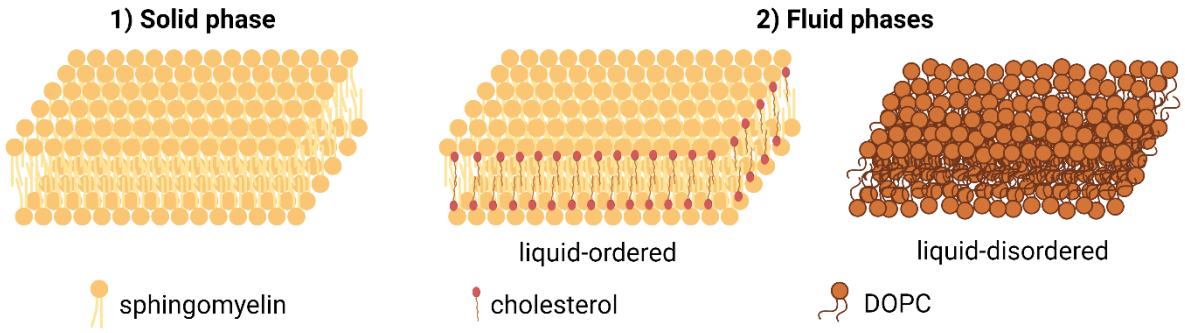


Figure 1.12: L_o phases made up of high-melting temperature lipids like sphingomyelin with high concentrations of cholesterol, and L_d phases made of low-melting temperature lipids like 1, 2-dioleoyl-sn-glycero-3-phosphocholine. The kinks present in DOPC allow for lateral lipid movement across the membrane surface. Created in BioRender based on (Sych, Mély and Römer, 2018).

The Sp-C lipopeptide also aids in the incorporation of cholesterol by increasing its miscibility in surfactant-mimicking membranes in lung surfactant (Roldan *et al.*, 2016). The hydrophobic surfactant proteins and cholesterol therefore aid in the viscosity and fluidity of pulmonary surfactant to contribute to its biophysical functions. Increased levels of cholesterol beyond 20-22 mol % of pulmonary surfactant, however, can impair surfactant function (Leonenko *et al.*, 2007), reinforcing the importance of cholesterol in the lateral organisation of native pulmonary surfactant (De La Serna *et al.*, 2004). This is seen in respiratory disorders such as pulmonary alveolar proteinosis (Seymour and Presneill, 2002) and acute respiratory distress (Markart *et al.*, 2007). In a study done by Roldan *et al.*, a fluorescent analogue of cholesterol (CTL) was used to analyse how the transmembrane hydrophobic surfactant proteins Sp-B and Sp-C influenced the distribution of cholesterol in surfactant membranes. Both Sp-B and Sp-C helped increase the motion of CTL in surfactant membranes and change in its distribution towards a more fluid phase, which could be effective to achieve the minimum tension required for the interfacial film to sustain breathing mechanics, suggesting that both surfactant proteins are essential in regulating the dynamic nature of surfactant lipids.

1.5 Diseases associated with surfactant protein deficiencies

Genetic studies and sequencing of Sp-B and Sp-C have suggested that hereditary Sp-B deficiency causes underdeveloped lungs in neonates, resulting in neonatal respiratory distress syndrome (NRDS) (Zimmermann *et al.*, 2005) which is lethal if untreated with exogenous surfactant. Deficiencies in tubular myelin were also found to be associated with NRDS, suggesting a possible deficiency in Sp-A. In the corners of the tubular myelin lattice, Sp-A has its collectin-like domains located in membrane intersections with its N-terminal domains extending towards the centre of the lattice square. Whilst Sp-A does not directly affect the biophysical functioning of surfactant, mutations in the *SFTPA* gene affect the formation of tubular myelin and interactions of Sp-A with other surfactant proteins crucial in maintaining surface activity (Ikegami *et al.*, 2001). Incubating Sp-A with DPPC:DPPG lipid bilayers containing Sp-B in the presence of Ca^{2+} resulted in the production of a lattice-like structure resembling tubular myelin. The *in vitro* formation of tubular myelin required the presence of Ca^{2+} ions to modify both charge and hydration states of the lipid monolayer from adjacent lipid bilayers. In response to the chelation of Ca^{2+} ions, Sp-A was still found to associate to the collapse tubular myelin, suggesting that the presence of Ca^{2+} was unnecessary in the interaction of Sp-A and lipids for assembling tubular myelin. Neonates with NRDS often have reduced lung compliance, impaired alveolar gaseous exchange, and alveolar collapse with decreased functional residual capacity (Walther et al., 2016), which is a result of the absence of surfactant to prevent alveolar wall cohesion. Approximately 10% of premature neonates in developed nations suffer from NRDS, however the incidence rate drastically declines with an increase in gestational age (Zuo *et al.*, 2008).

The mature Sp-B peptide is encoded by exons 6 and 7 of the 11 exon *SFTPB* gene, that corresponds to codons 201 to 279 on the 2 kb Sp-B mRNA (Warr *et al.*, 1987). Mutations in the *SFTPB* gene usually result from an insertion of 2 base pairs into codon 121 of the SP-B cDNA, causing a frameshift GAA to C substitution mutation, accounting for the absence of pro Sp-B and mature Sp-B causing respiratory failure in neonates associated with alveolar proteinosis (Nogee *et al.*, 1994). Low levels of Sp-C may indicate the presence of disease caused by mutations in various genes involved in surfactant metabolism (Griese *et al.*, 2016). For instance, mutations in the *SFTPC* gene may cause interstitial lung disease predominantly manifesting in postnatal or early infancy neonates

(Kröner *et al.*, 2015). Mutations in the *TTF1* gene, which encodes a protein mediating the termination of RNA polymerase I involved in gene transcription, can also affect the production of surfactant proteins. The membrane associated *ABCA3* protein facilitates phospholipid transportation into lamellar bodies containing surfactant proteins, to produce mature and functional pulmonary surfactant proteins.

1.6 The effects of environmental air pollution on lung surfactant

The human respiratory system is in constant contact with gases and particles in the atmosphere. There are several layers of defence mechanisms to protect the host from a myriad of diseases and air pollutants. The nasal hairs and mucus in the nasal passages help to block and trap large dust particles. Smaller particles may be trapped by mucus lining the respiratory tract, with the cilia propelling the mucus out into the stomach. Any pollutants that were not trapped by the mechanisms would therefore enter the alveoli and interact with lung surfactant.

Chronic exposure to high concentrations of air pollutants can lead to respiratory irritation and conditions such as asthma, or the exacerbation of diseases such as lung cancer and chronic obstructive pulmonary disease (COPD). CT imaging done by Paulin *et al.* revealed that long term exposure to ozone was linked to emphysema and reduced lung function; individuals with a history of smoking were at higher risk for respiratory exacerbations (Paulin *et al.*, 2020). Additionally, a study by Wiegman et al. revealed that chronic inflammation by ozone and reactive oxygen species induces oxidative stress, which may result in the activation of pro-inflammatory cytokines, when previously it was thought that cigarette smoking contributed to COPD symptoms. The direct contact of ozone with epithelial cells and macrophages in the alveoli activates inflammatory pathways, for example, through the induction of chemokine ligand 1 and C-C motif chemokine ligand 2, mediated by interleukin-1α (Manzer et al., 2012). There was evidence of inflammatory intracellular pathways linked to ozone exposure, ultimately leading to prolonged inflammation in the lungs (Wiegman et al., 2020). This section will therefore explore the effects of ozone, one of the most harmful pollutants to human health (Kurt, Zhang, and Pinkerton, 2016), on surfactant lipids.

1.6.1 The formation of ozone

Ozone is a highly reactive species and an environmental oxidant, formed when ultraviolet (UV) light and electrical discharges within the Earth's atmosphere react with oxygen at high altitudes. Ozone formation can also occur in the troposphere. The equations below (1.1 and 1.2) show nitrogen dioxide reacting with light to produce nitrous oxide and atomic oxygen. The highly reactive free oxygen atom subsequently encounters atmospheric oxygen and reacts to form ozone.

$$NO_2 + UV\ light = NO + O \quad (1.1)$$

$$O + O_2 = O_3 \ (1.2)$$

Ozone in the stratosphere can also move down to the troposphere, as shown in Figure 1.13, further affecting the background levels of tropospheric ozone. Ozone can react with high-energy visible light at wavelengths below 342 nm, to produce an excited state oxygen atom and oxygen gas. The combustion of fossil fuels (anthropogenic emissions) contributes to the formation of nitrogen oxides, carbon monoxides and VOCs. As illustrated in Figure 1.13, the excited state oxygen atom reacts with water molecules in the air to produce hydroxyl ions, which then react with volatile organic compounds (VOCs), such as ethane, to give an alkylradical ($R^{\cdot}$) and water. The alkylradical reacts with oxygen to form an alkylperoxyradical (RO_2), which reacts with NO, producing an alkoxyradical ($RO^{\cdot}$) and NO_2. The alkoxyradical reacts with oxygen to form a peroxyradical (HO_2), which reacts with NO in the air to form hydroxyl ions and NO_2. As seen in Equations 1.1 and 1.2, NO_2 is photolyzed to produce NO and O, where O reacts with atmospheric O_2 to give O_3 (Wang et al., 2022).

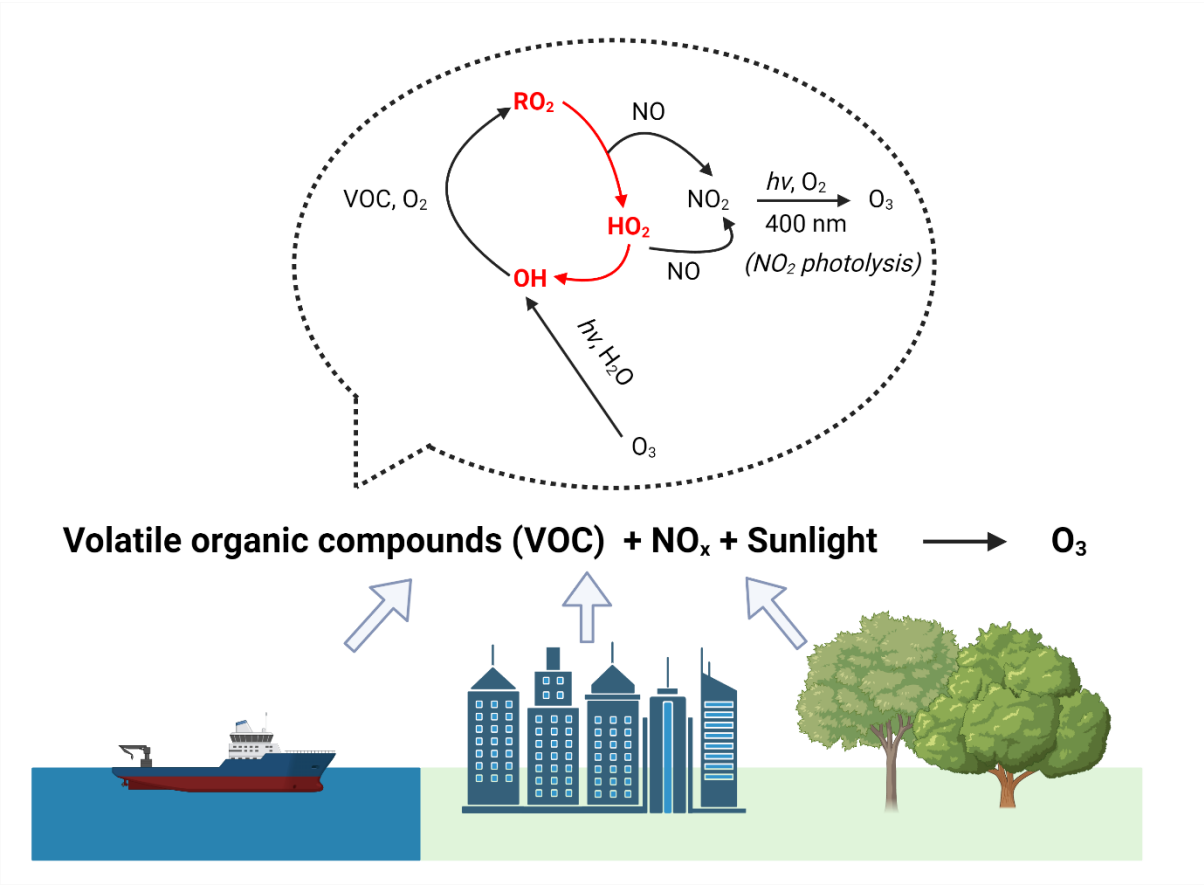


Figure 1.13: Diagram showing the various sources of ozone and oxidation pathways of VOC and CO radicals in pollution, resulting in the production of tropospheric ozone. Conventional radicals are highlighted in red. Examples where VOCs are produced include diesel fuels from ships, seaweed from the ocean releasing organic compounds into the air, trees and green plants emit large amounts of VOCs, and cars in cities burning petrol which releases NO, NO_2 and VOCs. The main source of ozone is derived from the photolysis of NO_2 with high-energy visible light (*hv*) at 400 nm wavelengths to produce NO and O, where the O radical reacts with O_2 to produce O_3 (Wang et al., 2022).

1.6.2 Oxidation of surfactant lipids

Ozone readily bypasses most defence mechanisms of the lungs, due to its limited solubility in water. The upper respiratory tract is unable to remove most of the ozone from inhaled air, resulting in ozone directly interacting with lung surfactant. The highly oxidative nature of ozone causes it to readily react with the carbon-carbon double bonds present in the unsaturated tail of surfactant lipids. Figure 1.14 shows the reaction of the carbon double bond of an alkene group with ozone, to form unstable molozonide. An O-O bond and a C-C bond breaks in the decomposition of molozonide, to form trioxolanes such as ozonide, which subsequently forms carboxylic acid, ketones, or aldehydes (Horibe and Goto, 2012).

Ozone may impact the effectiveness of lung surfactant due to the presence of unsaturated PG and PC lipids. The initiation of lipid peroxidation by ozone can lead to a chain reaction which forms reactive oxygen species (ROS), exacerbating the damage to surfactant function. A study by Thompson *et al.* showed that the reaction of POPC with ozone causes rapid loss of the terminal C9 portion of the oleoyl strand, leading to an increase in surface pressure at the air-water interface, and following that, a loss of material from the palmitoyl strand was observed in a secondary reaction step. It was revealed that there was no loss of DPPC in the PG:PC lipid mixtures upon exposure to ozone, however, the changes observed in surface pressure of the lipid films at the air-water interface imply the significance of unsaturated lipid tails in surfactant lipid films (Thompson et al., 2013). The effects of ozone on unsaturated lipids will therefore be explored further in Chapters 4 and 5, using MD simulations and neutron reflectivity (NR) experiments, to investigate if oxidative damage would lead to significant impairments in the biophysical function of lung surfactant.

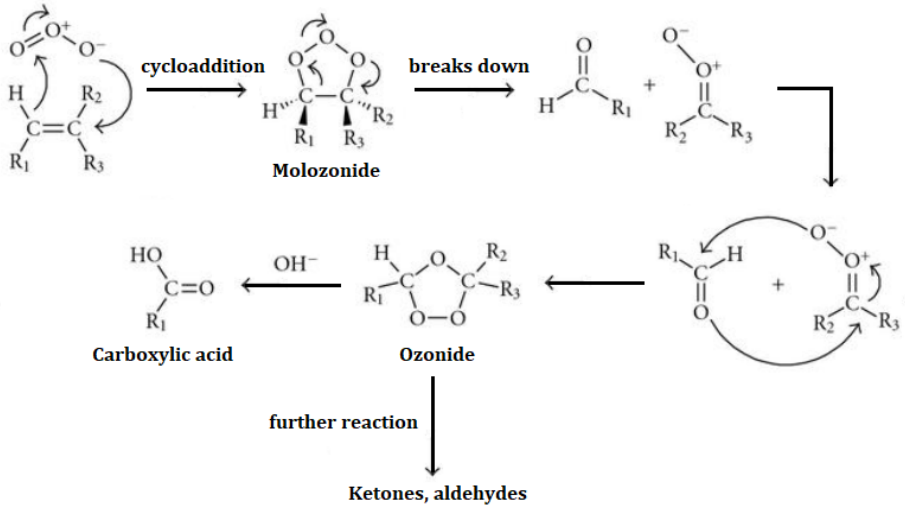


Figure 1.14: Interaction of ozone with a carbon double bond. Based on (Horibe and Goto, 2012).

1.7 Conclusions

The role of surfactant is key in maintaining the structure of the alveoli in the lungs to allow for gaseous exchange and immune function. Generally, it is known that hydrophilic surfactant proteins A and D facilitates immune function, while hydrophobic surfactant proteins B and C regulate biophysical function (Han and Mallampalli, 2015). However, there is a lot to be discovered about the collective mechanisms and interactions of surfactant lipids and Sp-B.

Presently, no literature has suggested that the structure of full-length mature Sp-B has been determined *in vitro*, due to the lack of techniques available in expressing a highly hydrophobic and small protein. It is widely proposed that the alpha-helical nature of Sp-B helps in maintaining surfactant function; for instance, synthetic peptides containing residues 1-25 of Sp-B were seen to retain the biophysical function of surfactant protein (Gordon et al., 2000). Not much is known about how oxidative damage to surfactant lipid tails may affect the peptide-lipid interactions and thus, the secondary structure of Sp-B.

Exposure to ozone is known to have detrimental effects on human health, as changing the structure of unsaturated surfactant lipid tails via ozonolysis reduces lung compliance and surfactant function. The formation of ROS and aldehydes may cause tissue inflammation and damage in the lungs by damaging cells lining the alveolar epithelium and capillary endothelium, activating an immune response in the lungs, and further exacerbating lung injury.

1.8 Research objectives

This research focuses on studying the effects of physical changes in the form of oxidative damage, on the interactions between peptides and lipids in model lipid membranes, particularly in lung surfactant where a protocol will be established to express recombinant Sp-B. Due to limitations in determining an experimental structure of Sp-B, the mechanisms by which Sp-B regulates biophysical functions in surfactant are largely unknown. This project therefore aims to explore ways to optimise the production of Sp-B using recombinant protein expression in *E. coli* as an extension of my previous MSc work done on the expression and characterisation of Sp-B (Chapter 2), which can be scaled up in future work to create alternative forms of therapy for diseases such as NRDS and ARDS. The expressed protein will subsequently be purified by chromatography methods, and its structure analysed using biophysical methods such as CD spectroscopy and mass spectroscopy, to identify key secondary structural elements (Chapter 3).

Currently, the lack of an experimentally determined full-length structure of Sp-B confirms the challenges faced in isolating or synthesising the protein. MD simulations of an N-terminal segment of Sp-B (Sp-B_{1-25}) would further explore the key structural features of Sp-B in various monolayers containing lipids commonly found in surfactant. Simulating the monolayers at different sizes (area per lipid, APL) would aid in understanding the behaviour of surfactant lipids and peptide during expansion and compression in the alveoli. Sp-B_{1-25} was also simulated in monolayers containing oxidised lipid, to investigate how oxidatively damaged lipid tails affect interactions between the peptide and surfactant lipids (Chapter 4).

Neutron reflectivity experiments on lipids commonly found in lipid membranes aim to study the interaction of protegrin (PG-1), an anti-microbial drug, with membrane lipid mixtures (Chapter 5). This chapter extends the research to further investigate how oxidative damage may affect the effectiveness of PG-1 in permeabilising the model bacterial lipid membranes containing unoxidised and/or oxidised lipids. Additionally, the degree of the membrane permeability may shed insight on whether PC lipids significantly affect the degree of PG-1 insertion, which had been established in the literature to primarily affect PG lipids.

Chapter 2 –Expression of Recombinant Sp-B using *E. coli*

2.1 Introduction

This chapter aims to explore the methods for expressing recombinant Sp-B. *Escherichia coli (E. coli)* is a well-known bacterial expression system used in the production of recombinant prokaryotic and eukaryotic proteins. Approximately 80% of protein structures registered in the PDB with known three-dimensional structures were produced with *E. coli* expression hosts (Terpe, 2006). *E. coli* expression systems have been developed to facilitate the controlled expression of heterologous proteins.

2.1.1 The T7 expression system

The T7 expression system comprises four main components: the host organism, T7 promoter, T7 RNA polymerase and the cloning vector. The cloning vector, typically a plasmid or virus, is designed to carry regulatory elements and the gene containing the protein to be expressed. The protein's size, disulfide bonds and any post-translational modifications should be considered before recombinantly expressing it with a suitable vector. *E. coli* systems can effectively express small molecular weight proteins containing a few disulfide bonds, potentially making them a suitable expression vector for Sp-B. Furthermore, proteins with a molecular weight lower than 30 kDa, 3 or less contiguous hydrophobic amino acids and 1 or less low complexity regions are likely to yield higher levels of soluble protein (Dyson *et al.*, 2004). Although Sp-B is considered a low molecular weight protein, it is highly hydrophobic, which may influence how much soluble protein can be produced.

The bacterial strain used can largely influence the production of the target protein. BL21 (DE3) Competent *E. coli* strains are exceptional at taking up foreign DNA for high-level recombinant protein expression, making them suitable for the purposes of this research, in particular, protein purification and structural studies. These cells, derived from the BL21 *E. coli* strain, contain the prophage DE3 derived from the λDE3 lysogen. The λDE3 prophage is a modified bacteriophage specifically designed to facilitate overexpression of the gene of interest, whilst under control by the T7 promoter. The λDE3 prophage attaches to an *E. coli* cell, introducing phage DNA into the cell, which circularises and integrates into the *E. coli* chromosome in the *lacZ* gene or in a non-essential region. The

T7 promoter DNA sequence is recognised by T7 RNA polymerase to drive efficient transcription of the target gene. The λDE3 prophage carries the gene for T7 RNA polymerase as shown in Figure 2.1, where it contains the the *lac*O repressor, and T7 RNA polymerase controlled by the lacUV5 promoter (P_{lacUV5}) (Jeong, Kim and Lee, 2016). Upon the addition of isopropyl β-d-1-thiogalactopyranoside (IPTG), the T7 promoter activates, and target gene expression occurs.

E. coli naturally contain a set of genes called the lac operon, which aids in the metabolism of lactose. The lac operon contains the *lac*Z gene, which expresses β-galactosidase that digests lactose into glucose and galactose, and the *lac*I gene, which expresses the lac repressor protein. When lactose is absent, the lac repressor protein binds to the operating region of the lac operon promoter, inhibiting binding of RNA polymerase to the promoter and thus, transcription of the *lac*Z gene. Similarly, the BL21 *E. coli* strain contains the pLysS plasmid which encodes the gene for T7 lysozyme, which naturally binds to and inhibits T7 RNA polymerase activity, reducing expression of the gene of interest. IPTG is a structurally analogous compound to the natural inducer lactose, which activates the lac operon in *E. coli*. However, IPTG cannot be metabolised by *E. coli*. The addition of IPTG to a bacterial culture causes the compound to freely enter the cells, where it binds to the lac repressor, causing a conformational change in the T7 lysozyme, preventing it from binding to the operating region of the lac operon promoter. Since IPTG inhibits the activity of T7 lysozyme, this in turn this stops the inhibition of T7 RNA polymerase by T7 lysozyme, allowing RNA polymerase to bind to the promoter to initiate transcription and target gene expression (see Figure 2.1). In addition, the pLysS plasmid contains the gene for chloramphenicol resistance.

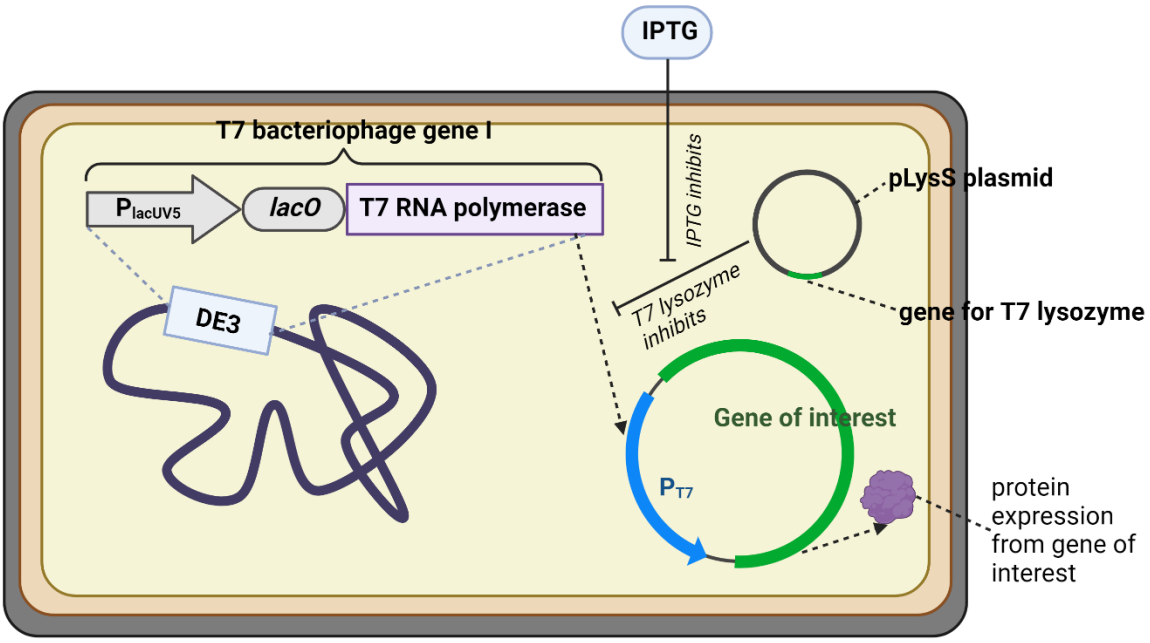


Figure 2.1: Diagram showing the components of a BL21 (DE3) pLysS *E. coli* cell. The BL21 strain contains the λDE3 lysogen (dark blue, left), which expresses the T7 RNA polymerase (purple) that is controlled by the mutant P_{lacUV5} promoter (brown arrow), with *lac*O being the repressor binding site. The T7 lysozyme naturally inhibits the transcriptional activity of T7 RNA polymerase in the absence of IPTG. In the presence of IPTG/during IPTG induction, the activity of the T7 lysozyme is inhibited. As a result, this allows the λDE3 lysogen to express the T7 RNA polymerase where the gene of interest (green) can be cloned downstream to express the protein of interest. Created in BioRender.

In a study by Zhang *et al.*, BL21 (DE3) cells were used to express hydrophobic membrane proteins, resulting in an increased and stable production yield of the protein grown in luria broth (LB) media, even in the absence of IPTG induction. With IPTG, non-producing clones get selected for and aggregation of the membrane proteins occur, implying that the Sec-translocon capacity is minimally saturated if IPTG is absent, causing increased protein yields (Zhang *et al.*, 2015). The BL21 (DE3) strains are also deficient in both the Lon and OmpT proteases. The Lon protease is an ATP-dependent protease which selectively degrades misfolded proteins and maintains mitochondrial homeostasis (Pinti et al., 2016). OmpT, an aspartyl protease, cleaves peptides between two positively charged residues to break down foreign material encountered by the bacterial cells (Haiko et al., 2009). The lack of such proteases gives the BL21 (DE3) strains the advantages of recombinant protein expression in large quantities through high

transformational efficiency and rapid cell growth, however, the lack of regulation may result in the accumulation of unstable or misfolded proteins.

2.1.2 Formation of the recombinant plasmid

The gene containing the protein of interest is typically isolated from an organism source or synthesised artificially. Before the protein can be expressed, the gene must be cloned into the chosen bacterial plasmid. This involves the usage of restriction enzymes to cut the plasmid and gene of interest at specific recognition sites, creating complementary overhangs. In previous experiments, the SspI and BamHI restriction enzymes were chosen to cleave the region before the plasmid's terminator sequence and after the promoter sequence, allowing the insertion of the Sp-B gene (Figure 2.3). The treated plasmids and gene fragments are mixed with DNA ligase, to covalently join them and create a recombinant plasmid (Figure 2.2).

In this research, the *E. coli* cells used had ampicillin and chloramphenicol resistance. This maintains the selective pressure of the cells in carrying the recombinant plasmids that contain the Sp-B gene, where only cells that have uptaken the Sp-B cloning plasmid will grow, which is crucial for the large-scale expression of Sp-B. Dual antibiotic selection gives more control over the expression of T7 polymerase, so protein expression does not occur prior to induction by IPTG. It also greatly reduces the changes of resistance gene loss, leading to plasmid loss or mutations, which are problems commonly encountered in strains with single antibiotic resistance genes.

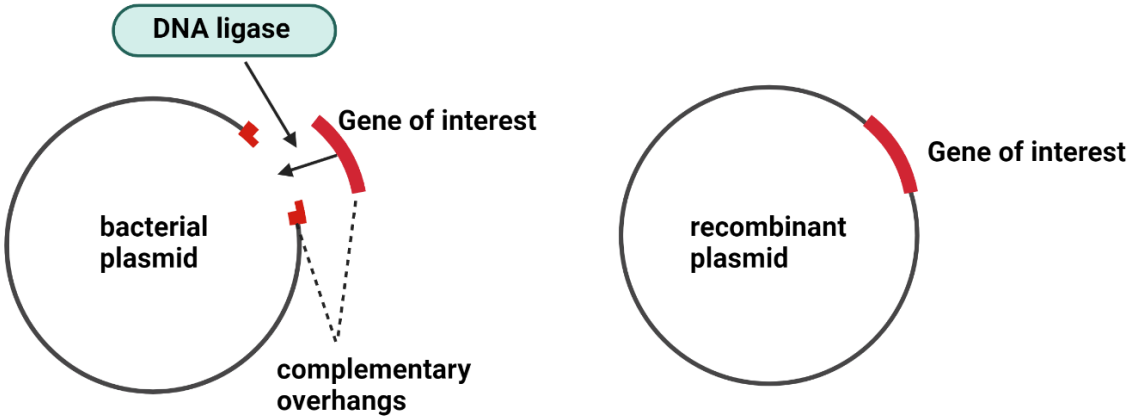


Figure 2.2: Cloning the gene of interest into a bacterial plasmid to form a recombinant plasmid. Created in BioRender.

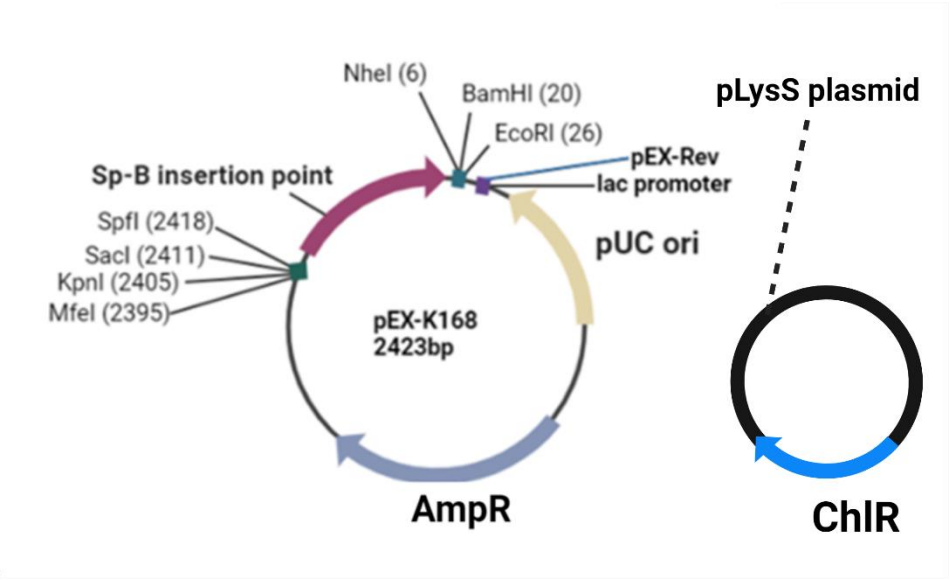


Figure 2.3: Full-sequence plasmid vector map of pEX-K168-Sp-B and the pLysS plasmid of BL21 (DE3) *E. coli*. The Sp-B vector shows the origin of replication (pUC ori, beige), lac promoter, ampicillin resistance open reading frame (bases 1316-2110, blue), multiple cloning sites (NheI, BamHI, EcoRI, SpfI, SacI, KpnI, MfeI). The pLysS plasmid carries the resistance gene for chloramphenicol. Created in BioRender by (Wang, 2019).

2.1.3 Solubility tags in protein expression

An advantage of using *E. coli* expression systems would be the ability to express recombinant Sp-B using affinity tags. These are short peptide sequences that help improve purification of the recombinant protein, since the tags specifically bind with high affinity to ligands or resins on chromatography columns. These tags may also improve the solubility and stability of the recombinant protein, which can be a concern when expressing hydrophobic proteins that are prone to forming precipitates.

Prior to commencing the work described in this thesis, samples of Sp-B linked to genes for three different solubility tags were prepared and described in this section. This was done in previous MSc work by Wang (Wang, 2019). The 6-polyhistidine (His6) is a 0.8 kDa affinity tag containing 6 consecutive histidine residues. It has a strong affinity to nickel ions, allowing for tagged recombinant proteins to be purified and isolated via immobilised metal affinity chromatography (IMAC), as they bind to the nickel-rich chromatography column. The small ubiquitin-like modifier (SUMO) tag, approximately 11 kDa in size, functions similarly to ubiquitin where it covalently attaches to the recombinant protein at a lysine residue (SUMOylation). This improves the stability and solubility of the recombinant protein, which may increase protein yield. Similarly, the N-

utilisation substance A (NusA) tag, approximately 55 kDa, improves the solubility and yield of the recombinant protein. However, the relatively large size of the SUMO and NusA tag may affect the structure and function of the recombinant protein. Hence, to find the most suitable affinity tag, a combination of 3 cloning vectors containing His6 were used in the expression of recombinant Sp-B, as seen in Figure 2.4.

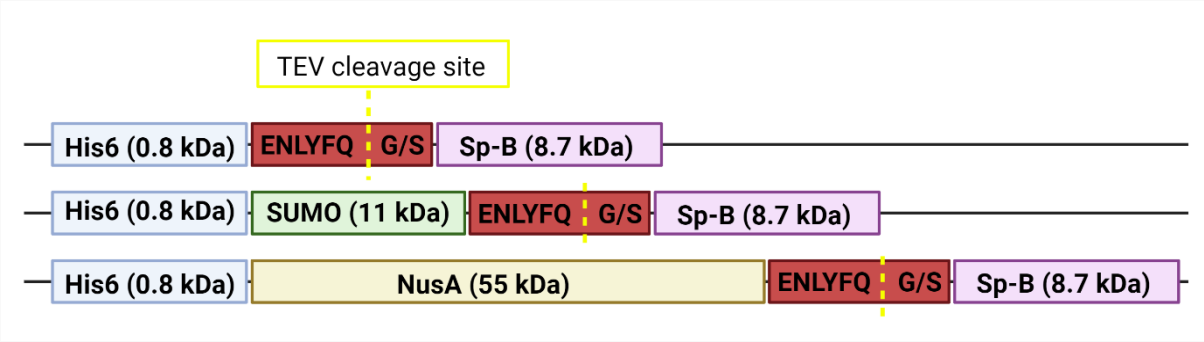


Figure 2.4: Cloning vectors used in the cloning of recombinant Sp-B in BL21 (DE2) pLysS cells. **Top to bottom:** His6 Sp-B, His6-SUMO Sp-B, His6-NusA Sp-B. The TEV protease cleaves between the 6th and 7th residues of the cleavage sites (sequence in red) as shown by the dotted lines in yellow. Created in BioRender.

Affinity tags were added to the N-terminal end of the Sp-B sequence and a TEV cleavage site inserted between the tag and the Sp-B. This allows tobacco etch virus (TEV) protease to cleave the bond between the glutamine and glycine (or serine) residue on the His6 tag, removing the tags from the recombinant protein after purification. The sequence was subsequently cloned into the bacterial cloning plasmid vector pEX-K168. The cloning vectors used in the plasmid constructs were pET-His6-TEV, pET-His6-NusA-TEV, and pET-His6-SUMO-TEV.

2.1.4 Recombinant plasmid transformation

After cloning, the recombinant plasmids can be introduced into competent bacterial cells through a process known as transformation. Competent cells are typically treated with divalent cations, making them susceptible to changes in temperature. The cells are subjected to heat shock, increasing their membrane permeability to encourage plasmid uptake.

In previous studies, the recombinant plasmids pET-His6-TEV, pET-His6-NusA-TEV, and pET-His6-SUMO-TEV, were transformed into *E. coli* Stellar™ competent cells. All plasmids contained the gene for ampicillin and chloramphenicol resistance, hence successfully transformed bacteria were identified by growing the cells on agar containing ampicillin and chloramphenicol.

The plasmids were subsequently purified and digested using SspI and BamHI and sent for DNA sequencing. The cleaved plasmid fragments were run on an agarose gel (Figure 2.5) to check for successful cloning of Sp-B into the recombinant vectors containing the solubility tag, indicated by the presence of two distinct bands per lane, corresponding to the base pair (bp) of the recombinant protein (Sp-B) at 244 bp, and the solubility tag used; 6289 bp for the His6-NusA cloning vector (Addgene plasmid #29709), 5104 bp for His6-SUMO cloning vector (Addgene plasmid #29711) and 4786 bp for the His6 cloning vector (Addgene plasmid #29666).

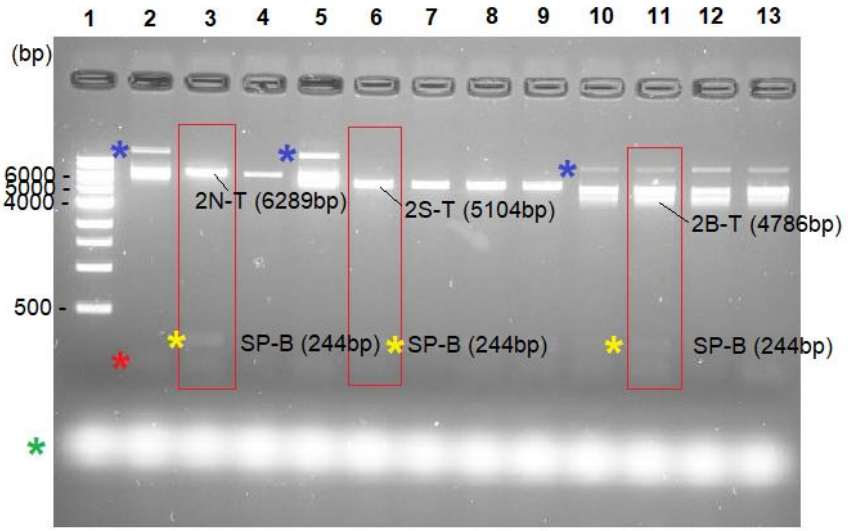


Figure 2.5: Agarose gel electrophoresis of restriction digested plasmid clones. **Lane 1:** 1kb (10,000bp) DNA ladder, **Lane 2-5:** cleaved His6-NusA Sp-B recombinant clones, **Lane 6-9:** cleaved His6-SUMO Sp-B recombinant clones, **Lane 10-13:** cleaved His6 Sp-B recombinant clones (yellow asterisk). Blue asterisk: partially digested plasmids, red asterisk: small RNAs, green asterisk: Orange G dye. Obtained from (Wang, 2019).

2.1.5 Optimising the expression conditions for Sp-B

Before expressing a recombinant protein on a larger scale, conducting a test expression would determine the suitability of the parameters used to express the protein of interest. Such parameters include the temperature, induction concentration and time, and competent cells used.

Previous studies by Wang confirmed that affinity tagged Sp-B could not be expressed when transformed into Origami B DE3 competent cells, compared to BL21 (DE3) pLysS competent cells. It was also seen that the His6 affinity tag was preferable over the His6-NusA and His6-SUMO tags used. Moreover, His6 Sp-B was seen to be expressed at higher yields when induced at 30˚C for 3 hours, rather than 18˚C overnight. The conditions in which Sp-B is expressed and purified in this research, are therefore based on the conclusions reached in these experiments (Wang, 2019).

2.1.6 Purification of His6 (2B-T) Sp-B

IMAC is a common purification technique used to rapidly isolate recombinant proteins. Previous research has shown affinity-tagged protein purification via IMAC to achieve purity yields of up to 95% (Bornhorst and Falke, 2000), including *E. coli* expression systems (Van Dyke, Sirito and Sawadogo, 1992).

Chromatography columns such as HisTrap, are coated with agarose or Sepharose beads, where nickel (Ni^{2+}) ions are added using chelating agents like nitrilotriacetic (NTA) acid, forming the Ni^{2+}-NTA resin. The nickel-charged resin has a strong affinity for histidine residues, which are abundantly found in the His6 affinity tag. The His6 tag shifts the association/disassociation equilibrium towards the association with the immobilised ligand (Ni^{2+}) leading to strong binding to the Ni^{2+}-NTA resin.

After protein expression the cells are lysed and the soluble fraction, the cell extract passes through the column, most of the His6-tagged protein will bind to the Ni^{2+}-NTA resin. Proteins and other compounds that do not possess the His6 tag will not bind to the resin and are eluted in the flowthrough. Subsequently, the addition of a buffer with a high imidazole concentration will force the His6-tagged protein to be eluted in the flowthrough (Figure 2.6). Imidazole competitively binds to the binding sites on the Ni^{2+}-NTA resin, due to its cyclic ring structure resembling the side chain of histidine.

The high specificity and concentration-dependent elution mechanism of the HisTrap column makes it ideal in the purification and isolation of His6-tagged protein. Furthermore, the robust interaction between the histidine tag and the Ni^{2+}-NTA resin works under both native and denaturing (8 M urea or 6 M guanidinium hydrochloride) conditions and oxidising and reducing conditions while withstanding a wide variety of chemicals.

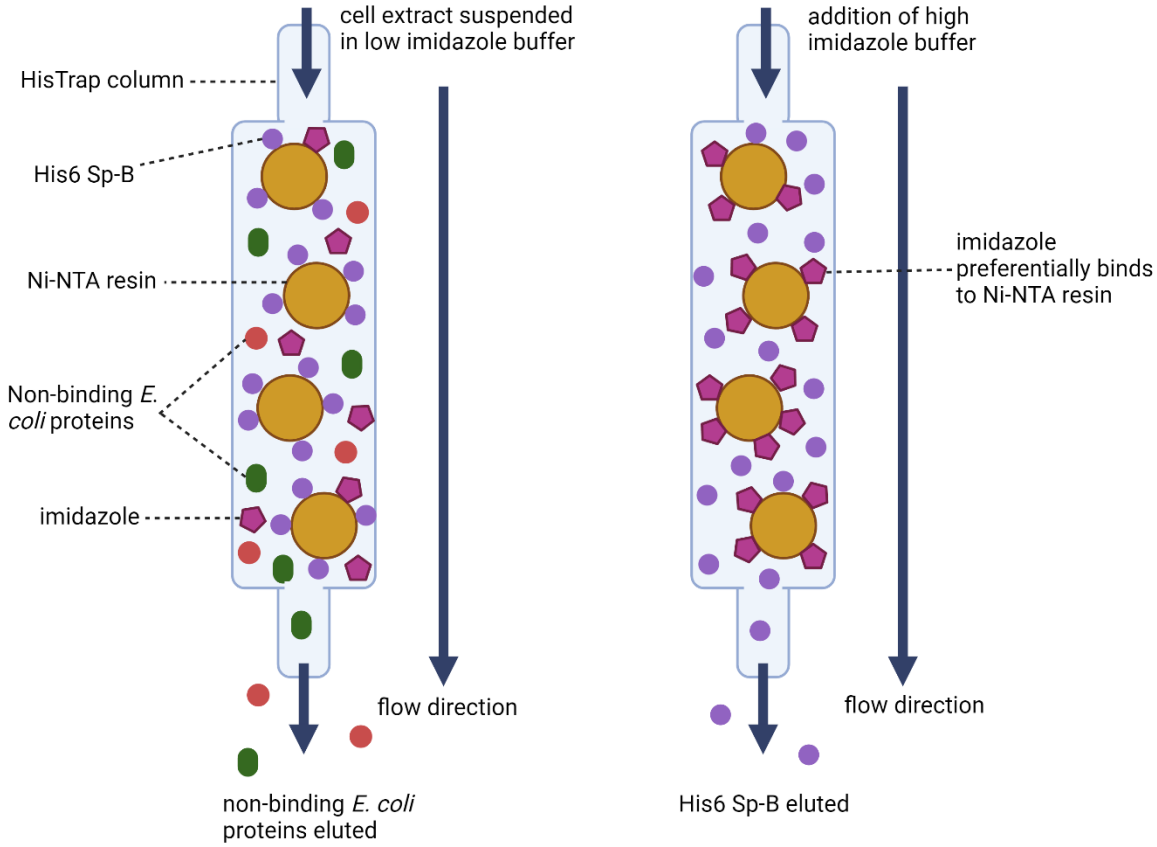


Figure 2.6: Diagram of an IMAC column containing Ni^{2+}-NTA resin (yellow). **Left:** Non-binding *E. coli* proteins (green, red) are eluted when cell extract in low imidazole buffer is added through the column. His6 Sp-B (purple) is bound to the resin. **Right:** The addition of high imidazole buffer causes imidazole (purple) to bind to the resin, eluting His6 Sp-B. Made in BioRender.

2.2 Protein assay techniques

The NanoDrop™ spectrophotometer is a commonly used method used in the quantification of protein concentration. This is based on the principle of the Beer-Lambert law, where the protein concentration is directly proportional to the protein absorbance at a constant pathlength and specified wavelength. The NanoDrop™ operates via ultraviolet-visible (UV-Vis) spectrophotometry, where light absorbed by the sample, at 280 nm in the case of proteins, is measured. Typically, this is the absorbance at which aromatic residues like tryptophan and tyrosine absorb light.

The NanoDrop™ allows the measurement of microvolume (0.5 - 2 µL) samples, where a drop of sample is pipetted onto the pedestal (detection surface), retained by its surface tension properties. A second optical surface touches the sample, forming a liquid column between the optical fibers, in turn forming a vertical optical path. This provides a great alternative to measuring protein concentrations using cuvettes, since a much smaller volume of sample is required.

The automatic detection of the pathlength (1 - 0.05 mm) is crucial for accurately measuring samples with a broad range of concentrations. For instance, larger concentrations of protein can be measured at shorter pathlengths. A UV light source is directed through the sample, and the NanoDrop™ measures the amount of light absorbed by the sample. The absorbance data is processed by its software to determine the concentration in mg mL^{-1}, using the extinction coefficient and molecular weight of the protein entered by the user, which has been calculated based on pre-determined values for individual amino acids, by inputting the amino acid sequence into the ProtParam website. The concentration is therefore calculated by multiplying the absorbance recorded from the NanoDrop at 280 nm, with the extinction coefficient calculated from ProtParam. In the case of His6 Sp-B and Sp-B, the extinction coefficients used were 0.9 and 0.974 $(mg\ mL^{-1})^{-1}\ cm^{-1}$ respectively, under the assumption that all cysteine residues were reduced.

This method allows the rapid measurement of fluorescence across a broad spectrum, spanning both UV and visible light. However, the protein sample should be largely free from contaminants for accurate measurements. If the protein sample does not strongly absorb at 280 nm, colorimetric assays may be a suitable alternative (Desjardins, Hansen and Allen, 2010).

2.3 Mass spectrometry

Mass spectrometry (MS) is a widely used method to quantitatively and qualitatively characterise proteins (Domon and Aebersold, 2006), by measuring the mass-to-charge (m/z) ratio of ions. This is done by bombarding the compound ($\boldsymbol{M}$) with a beam of electrons ($\boldsymbol{e^-}$) to produce gas-phase ions (ionisation). The molecular ion ($\boldsymbol{M^{\bullet+}}$) has an odd number of electrons, causing it to undergo fragmentations to produce a free radical ($\boldsymbol{R^{\bullet}}$) and an ion containing an even number of electrons ($\boldsymbol{X^+}$), or a molecule ($\boldsymbol{N}$) and a new radical cation ($\boldsymbol{Y^{\bullet+}}$). The product ion generated from the molecular ion can then undergo subsequent fragmentations.

$$\boldsymbol{M + e^- \rightarrow M^{\bullet+} + 2e^-}$$

$$\boldsymbol{M^{\bullet+} \rightarrow X^+ + R^{\bullet}}$$

$$\boldsymbol{M^{\bullet+} \rightarrow Y^{\bullet+} + N}$$

Uncharged free radicals and molecules do not produce a peak on the spectra; however, the charged ions are accelerated in the mass spectrometer, and separated according to their m/z ratio. A mass spectra plot of relative ion abundance, determined by the base peak, is plotted against the m/z ratio.

2.3.1 Bottom-up and top-down approach

The analysis of protein fragments in mass spectrometry are done using the bottom-up or top-down approach. In the bottom-up approach, protein samples are digested into smaller peptide fragments with enzymes such as trypsin (Figure 2.7). The fragments are separated according to size, by liquid chromatography (LC) methods and combined with mass spectrometry (LC-MS). To identify the protein, the mass spectrometer measures the m/z ratios of the fragments, and the resulting spectra are compared to known theoretical spectra produced from protein databases.

In contrast, the top-down approach allows protein samples to be directly infused into the mass spectrometer. The intact protein undergoes electrospray ionisation (ESI) and subsequently fragmentation by electron-capture or electron-transfer dissociation methods to produce fragment ions.

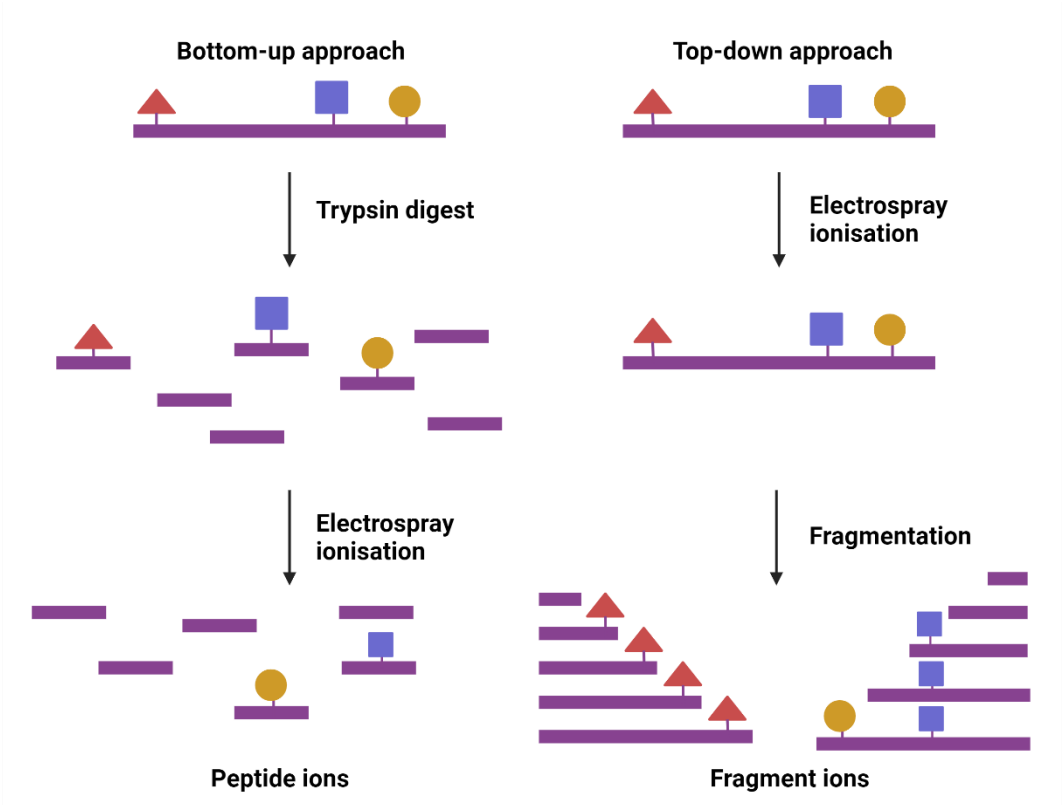


Figure 2.7: Diagram showing bottom-up (left) and top-down (right) proteomics. In the bottom-up approach, a peptide is digested using trypsin into smaller fragments, where a fraction of peptide ions generated via ESI reveal sequence data to identify the original proteoform based on unique sequences. In the top-down approach, intact peptide ions are generated using ESI and fragmented to give fragment ions that provide sequence data to identify the original proteoform.

A study by Klont *et al.* used the bottom-up approach, digesting Sp-D with trypsin and analysing the peptide fragments using LC-MS to quantitatively determine the levels of Sp-D in serum (Klont et al., 2019). The sensitivity of the bottom-up approach results in improved ionisation and analysis that are compatible with complex protein mixtures, attributed to the ease of separating peptide fragments compared to whole protein samples.

Using the top-down approach analyses native protein samples provide comprehensive structural and modification sites on the protein. The top-down approach may also identify the presence of low molecular weight proteins which can go undetected by the bottom-up approach. Moreover, it provides ease of sample preparation since the samples do not require enzymatic digestion, which can be advantageous for membrane proteins resistant to trypsin digestion.

2.3.2 Applications of mass spectrometry in protein digestion

In proteomics, MS and protein digestion are often combined to analyse protein structure, where various strategies have been optimised to study membrane proteins. Compared to soluble proteins, the optimisation of the extraction, digestion and detection conditions of membrane proteins are much more challenging. Due to its high cleavage specificity and proteolytic activity, trypsin is widely used to digest proteins to generate peptides to be analysed via mass spectrometry. Trypsin is a serine protease belonging to the PA clan superfamily, characterised by a chymotrypsin-like fold and similar proteolysis mechanisms. Trypsin contains a catalytic triad of residues namely histidine-57, aspartate-102 and serine-195, where protons are abstracted from serine to histidine, and subsequently histidine to aspartate (Polgár, 2005). The alkoxide serine has a stronger affinity on the protons compared to the histidine imidazole ring, where both serine and histidine form short low-barrier hydrogen bonds with the proton. The nucleophilicity of serine facilitates the attack on the amide carbon during proteolysis, and the formation of an oxyanion hole from the amide hydrogen atoms of glycine-193 and serine-195 stabilises the negative charge accumulating on the amide oxygen after the nucleophilic attack. This causes the planar amide carbon to assume a tetrahedral geometry to decrease the energy barrier needed to form the intermediate to facilitate the preferential binding of the transition state. Trypsin proteolytically cleaves proteins into peptide chains of approximately 700-1500 Da, at the carboxyl (C-terminal) end of arginine or lysine residues (Laskay *et al.*, 2013) as shown in Figure 2.8. The size of these peptides and the charged residues at the ends facilitates detection via mass spectrometry. The negatively charged residue, aspartate-189, located in the S1 catalytic pocket of trypsin is responsible for the specificity of trypsin, attracting and stabilising the positively charged arginine and lysine residues.

A study by Min, Choe and Lee analysed various protease digestion conditions to detect the sequence coverages of several membrane proteins. A combination of different additives at low concentrations were used to facilitate membrane protein solubilisation whilst minimising potential interferences with the protease digestion and MS. Bacteriorhodopsin was digested using either trypsin or chymotrypsin, along the addition of methanol, RapiGest or sodium dodecyl sulfate (SDS) additives. Peptides detected from digestion by either enzyme were subsequently combined to increase sequence coverage,

and the improved conditions applied to selenoprotein S and *E. coli* membrane proteins. High concentrations of additives were found to negatively impact the activity of trypsin and chymotrypsin. For instance, the addition of 10% methanol did not negatively affect enzymatic activity, whereas the addition of 30% methanol caused a 30% decrease in enzymatic activity. The sequence coverage of Selenoprotein S reached 46% and 25% for trypsin and chymotrypsin respectively, with a combined sequence coverage of 54% and 64% for SDS and methanol/RapiGest respectively. The addition of MeOH/RapiGest allowed the detection of more sequences and in addition, a transmembrane domain peptide, which could not be detected under standard conditions. This shows that trypsin and chymotrypsin can aid in the digestion of highly hydrophobic membrane proteins (eg. bacteriorhodopsin), under improved digestion conditions. These conditions resulted in a 40% sequence coverage of bacteriorhodopsin, compared to 0% under standard conditions (Min, Choe and Lee, 2015).

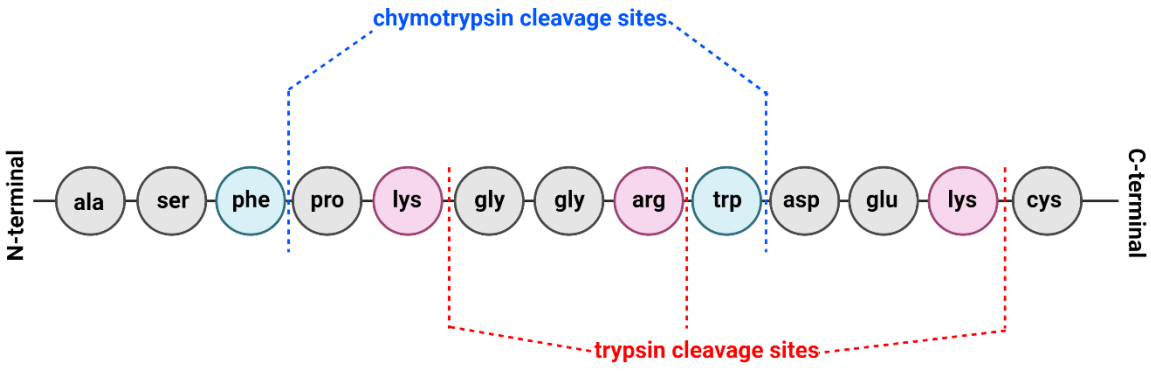


Figure 2.8: Cleavage sites of trypsin and chymotrypsin in a peptide sequence. Trypsin cleaves peptide bonds at the C-terminal side of lysine and arginine residues (red dotted lines), while chymotrypsin cleaves peptide bonds at the C-terminal side of aromatic amino acids (blue dotted lines).

2.4 Materials and methods

The reagents and their respective concentrations used in the expression and purification of recombinant His6 Sp-B are listed in Table 2.1. For most purification experiments, Tris-based buffer was used. Phosphate-based buffers were used for CD spectroscopy. Additionally, sodium fluoride (NaF) was used in place of sodium chloride (NaCl) for CD experiments. TEV protease was prepared using the TEV Protease Purification Protocol (Tropea, Cherry and Waugh, 2009). and stored in 1 mL aliquots at -80˚C.

Reagents and buffers prepared	Concentrations
Ampicillin, soluble in 1M HCl (Melford)	50 μg mL^{-1}
Chloramphenicol, soluble in ethanol (Melford)	34 μg mL^{-1}
Ultrapure isopropyl β-d-1-thiogalactopyranoside, non-animal origin (IPTG) (NeoBiotech)	1 M
Low imidazole purification buffer (pH 8.5)	20 mM imidazole (99% purity, Sigma Aldrich), 25 mM Trizma base (≥ 99.9% purity, Sigma Aldrich), 500 mM NaCl (≥ 99.5% purity, ThermoFisher Scientific), 10% glycerol (≥ 99% purity, ThermoFisher Scientific)
High imidazole purification buffer (pH 8.5)	500 mM imidazole (99% purity, Sigma Aldrich), 25 mM Trizma base (≥ 99.9% purity, Sigma Aldrich), 500 mM NaCl (≥ 99.5% purity, ThermoFisher Scientific), 10% glycerol (≥ 99% purity, ThermoFisher Scientific)
Dialysis buffer (pH 8.5)	25 mM Trizma base (≥ 99.9% purity, Sigma Aldrich) or sodium phosphate (monobasic ≥ 98%, dibasic ≥ 99% purity, Sigma Life Science), 500 mM NaCl (≥ 99.5% purity, ThermoFisher Scientific), 10% glycerol (≥ 99% purity, ThermoFisher Scientific)

Size exclusion buffer (pH 8.5)	25 mM Trizma base base (≥ 99.9% purity, Sigma Aldrich) or sodium phosphate (monobasic ≥ 98% purity, dibasic ≥ 99% purity, Sigma Life Science) , 500 mM NaCl (≥ 99.5% purity, ThermoFisher Scientific)/300 mM NaF (≥ 99% purity, Sigma Aldrich), 1 mM DTT (Melford), 10% glycerol (≥ 99% purity, ThermoFisher Scientific)
TEV protease	0.573 mg mL^{-1}

Table 2.1: Solutions prepared for expression and purification of recombinant His6 Sp-B.

2.4.1 Cloning and transformation of recombinant plasmid

Glycerol stock solutions of the plasmid were prepared and stored at -80˚C. The vector contains the Sp-B gene with a His6 tag at the N-terminus. Agar plates were prepared containing 50 μg mL^{-1} ampicillin (Melford) and 34 μg mL^{-1} chloramphenicol (Melford). The transformed cells were plated out onto the solid agar and incubated overnight at 37˚C. A colony was picked from the agar plate and used in the inoculation into sterile luria broth (LB) miller (Formedium) containing 50 μg mL-1 ampicillin and 34 μg mL^{-1} chloramphenicol. The recombinant cells were incubated at 37˚C and 200 rpm overnight.

2.4.2 Bacterial culture growth and induction

5 mL of the overnight culture was subsequently added in a 1:100 ratio to 12 separate flasks containing 500 mL LB (50 μg mL^{-1} ampicillin, 34 μg mL^{-1} chloramphenicol). The cells were grown in a shaking incubator at 220 rpm, at 37˚C for 3 hours to an optical density (OD_{600}) of 0.6-0.8.

Expression of the Sp-B protein was subsequently induced with 0.5 mM IPTG (Melford) at 30˚C for 3 hours. The soluble supernatant and insoluble cell pellet fractions were separated by centrifugation using the Beckman Avanti™ J20-I, at 3024 *g* (5000 rpm), at 4˚C for 20 minutes. Approximately 8.4 g of cell pellet was obtained from the 12 x 500 mL cultures. The cell pellets were stored at -20˚C for further use.

2.4.3 Recombinant protein purification

The cell pellets were resuspended with low imidazole Tris-based buffer (see Table 2.1), where 5 mL of buffer was added per gram of cells. The cells were lysed with 10 µg of bovine pancreas grade II DNAse I (LOT 55707500, Roche) per 47 mL of resuspended cells, and ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail tablets (Roche Diagnostics). The cells were sonicated at 60% amplitude (78 W) and 10 s pulse for 6 minutes, using the Sonics VCX130 Vibra Cell Ultrasonic Liquid Processor (130 W). The lysed cells were centrifuged at 75,600 *g*, at 4˚C for 1 hour in the JA Beckman 2550 rotor. The supernatant containing recombinant His6 Sp-B was filtered into a Falcon tube using a sterile cellulose acetate (CA) 0.45 Minisart® filter unit (Sartorius Stedim Biotech).

Purification of recombinant His6 Sp-B carried out using IMAC, using a 1 mL HisTrap HP column (High Performance) (GEHealthcare Life Sciences). The column was washed and equilibrated using the AKTA chromatography system. The column was set at a flow rate of 1 mL/min and washed with 10 mL of filtered ethanol and ddH_2O. The column was subsequently equilibrated with 10 mL of high imidazole buffer (Tris or phosphate-based, see Table 2.1), followed by 10 mL of low imidazole buffer.

The equilibrated column was attached to a Pharmacia LKB peristaltic Pump P-1, trapping recombinant His6 Sp-B in the column, removing unbound proteins in the flowthrough. The column with the bound proteins was reattached to the AKTA chromatography system, where more unbound proteins were washed out with 30 mL of low imidazole buffer (25 mM). A 20 mL gradient elution was done to elute the bound protein, with a final step where 5 mL of buffer with a high concentration of imidazole (500 mM) was added to the column.

2.4.4 SDS-PAGE of protein fractions eluted from affinity chromatography

His6 Sp-B was eluted from the column in 1 mL fractions by the AKTA chromatography system. The fractions containing recombinant His6 Sp-B were pooled into a 3500 molecular weight cut-off (MWCO) SnakeSkin™ Dialysis tubing (ThermoFischer Scientific) and dialysed overnight at 4˚C in dialysis buffer.

To remove the His6 tag from the protein, 1 mg mL^{-1} of TEV protease was added to His6 Sp-B and the protein was dialysed under the same conditions. For CD spectroscopy experiments, recombinant protein was dialysed overnight in phosphate-based dialysis buffer. TEV protease and other possible contaminants were subsequently separated from untagged Sp-B via affinity chromatography.

To determine the fractions in which Sp-B was eluted after incubation with TEV protease, samples from the eluted fractions were collected and run on a Bolt™ 4-12% Bis-Tris Plus SDS-PAGE gel (ThermoFischer Scientific) using the Bolt™ MES SDS Running Buffer (ThermoFischer Scientific), at 200 V for 24 minutes in the PowerPac 300 (BioRad). The gel was stained using the Pierce™ PageBlue™ Protein Staining solution (ThermoFischer Scientific), using the Fast-Staining Protocol (Figure 2.8).

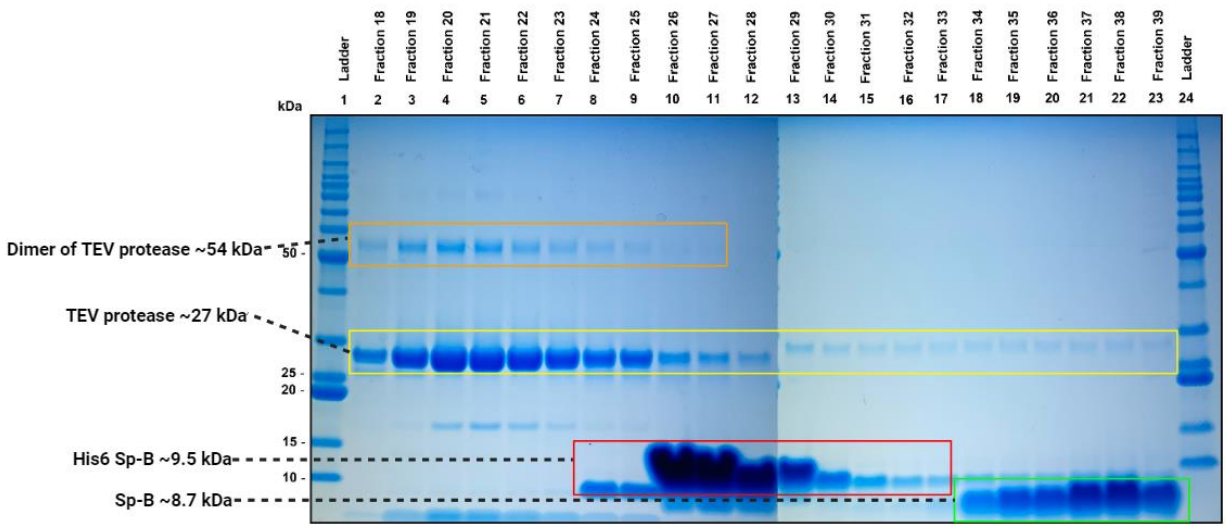


Figure 2.8: SDS-PAGE gel showing the fractions (18-39) eluted from the 1 mL HisTrap column by affinity purification. The gel was stained with PageBlue™ Protein Staining solution. **Lane 1 and 24:** BenchMark™ Protein Ladder (Novex). **Lane 2-23:** Fractions 18 to 39 eluted from the column. The key bands are highlighted in orange (dimerised TEV protease), yellow (TEV protease), red (His6 Sp-B) and green (Sp-B).

2.4.5 Size exclusion chromatography of Sp-B

To further purify Sp-B, the sample was concentrated in a Vivaspin® Turbo 4 Ultrafiltration unit (Sartorius) and centrifuged at 3200 rpm at 4˚C down to a volume of approximately 500 µL. The concentrated sample was then centrifuged at maximum speed of 18.624 *g* (14,000 rpm) at 4˚C in a microfuge for 5 minutes and injected into the AKTA Purifier 10 System (GEHealthcare), pre-equilibrated to a flow rate of 1 mL/min with ddH_2O and size exclusion buffer. Size exclusion chromatography was performed using the HiLoad® 16/600 Superdex® 75 pg size exclusion column (GEHealthcare). The sample was eluted through the column with 120 mL of size exclusion buffer. The eluted fractions were collected and run on an SDS-PAGE gel at 200 V for 24 minutes to determine the fractions where Sp-B was eluted (Figure 2.9).

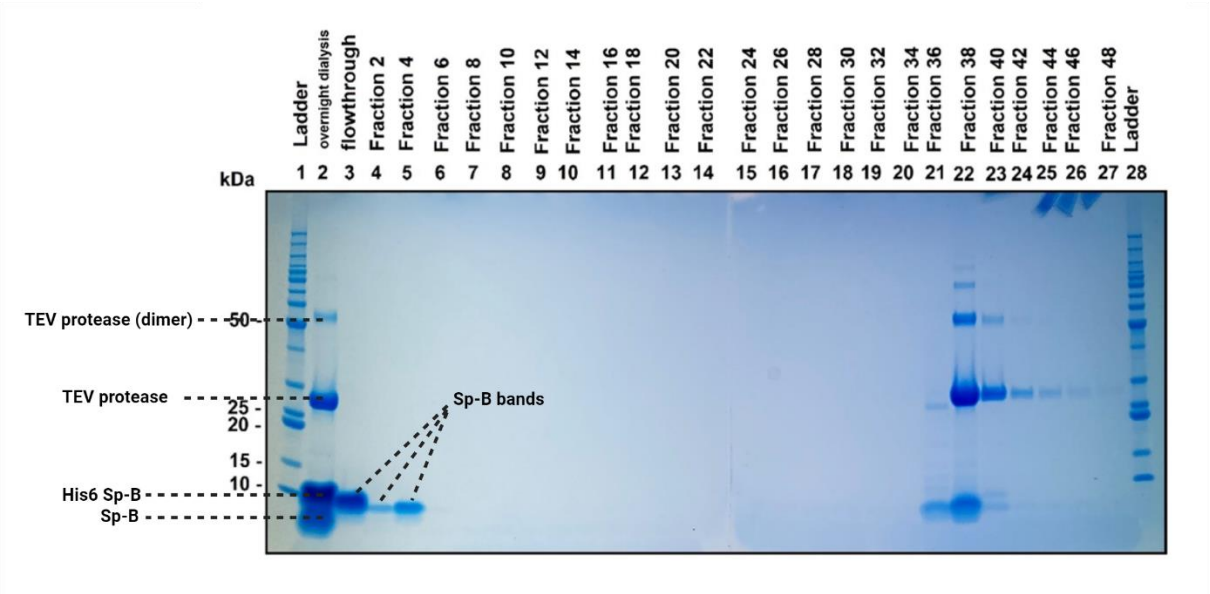


Figure 2.9: SDS-PAGE gel showing the fractions obtained from the size exclusion chromatography of Sp-B using a 24 mL Superdex 75 pg column. The gel was stained with PageBlue Protein Staining solution. **Lane 1 and 28:** BenchMark™ Protein Ladder. **Lane 2:** Sp-B after overnight dialysis with TEV. **Lane 3:** Flowthrough obtained after removing TEV with a HisTrap column. **Lane 4-27:** Fractions eluted from the size exclusion chromatography column. The key bands in the gel are highlighted accordingly.

2.4.6 Western blot analysis of recombinant His6 Sp-B and Sp-B

Samples of tagged and untagged Sp-B were run on a Bolt™ 4-12% Bis-Tris Plus SDS-PAGE gel for 200 V and 24 minutes. The gel was blotted using the Invitrogen iBlot 2 Dry Blotting System using the dry transfer method, for 7 minutes: 1 minute at 20 V, 4 minutes at 23 V

and 2 minutes at 25 V. The proteins transferred onto the nitrocellulose membrane were blocked for 1 hour with blocking solution (5% w/w of non-fat dry milk powder dissolved in phosphate buffered saline (PBS)). The membrane was subsequently incubated with antibody solution, anti-His6, for 1.5 hours, washed twice with PBS, and staining solution (U-1380 Fast urea H_2O_2 (SigmaFast) and D9292 DAB (SigmaFast) tablets dissolved in dH_2O) was added until bands were visible on the membrane. The staining reaction was inhibited by washing and incubating the membrane with dH_2O.

2.4.7 Addition of formic acid to recombinant protein samples

To test the stability of recombinant proteins His6 Sp-B and Sp-B at lower pH for mass spectrometry analysis, formic acid (96%, ACS reagent CAS 64-18-6, Sigma Aldrich) was added in a 1:10 dilution to the protein sample.

2.4.8 Trypsin digest of recombinant His6 Sp-B and Sp-B

Trypsin digests were done on samples of His6 Sp-B with and without dithiothreitol (DTT) reduction. For the digest with DTT reduction, 1 M of DTT was added to the sample to give a final concentration of 1 mM. Concentrated samples of His6 Sp-B previously prepared from protein expression and purification were precipitated using the Methanol-Chloroform Precipitation Protocol. The concentration of the reagents used in the trypsin digest were recorded in Table 2.2.

Reagents	Concentration
Tris buffer (pH 7.8)	0.4 M
Urea (pH 7.8) (99.5-100.5% purity, CAS 57-13-6, AnalaR)	6 M
Trypsin from porcine pancreas (CAS 9002-07-7, Sigma Aldrich)	0.2 mg mL^{-1}
Alkylating reagent, pH 7.8	200 mM iodoacetamide (≥ 99% purity, CAS 144-48-9, Sigma Aldrich) dissolved in Tris buffer
Solution A	98% MilliQ-H_2O, 2% acetonitrile (CH_3CN) (> 99.9% purity, CAS 75-05-8, Sigma Aldrich),

	0.1% FA (96% purity, ACS reagent CAS 64-18-6, Sigma Aldrich)
Solution B	35% MilliQ-H_2O, 65% acetonitrile (CH_3CN) (> 99.9% purity, CAS 75-05-8, Sigma Aldrich), 0.1% FA (96% purity, ACS reagent CAS 64-18-6, Sigma Aldrich)
DTT solution in 1 M H_2O (CAS 3483-12-3, Sigma Aldrich)	1 M

Table 2.2: Reagents used in the trypsin digest protocol. Digestion was carried out at pH 7.8 for optimal hydrolysis.

100 µL of 6 M urea was added to the precipitated protein pellet and vortexed. The sample was vortexed with 20 µL of alkylating reagent and incubated at room temperature for 45 minutes. Subsequently, the concentration of urea was reduced by diluting with 775 µL ddH_2O. Trypsin solution was added in a 1:50 ratio to digest the protein overnight at 37˚C. The overnight reaction with trypsin was inhibited by lowering the pH of the solution below 6, using concentrated acetic acid. The Sep-Pak C18 Purification Protocol was used (University of Oxford, 2019). The C18 Sep-Pak Plus™ (Waters, SKU WAT020515) column was equilibrated with 5 mL of solution B, followed by 10 mL of solution A. The digested sample was added into the column and washed with 10 mL of solution A, and subsequently eluted with 2 mL of solution B. The purified peptides were dried down completely in a Speed-Vac on low heat, and resuspended in up to 20 µL solution A.

2.4.9 Gel electrophoresis of digested peptides on tricine SDS-PAGE

Table 2.3 shows the reagents prepared for use in the SDS-PAGE and staining of trypsin digested samples. Samples obtained from the trypsin digest with added tricine sample buffer were heat blocked at 85ºC, and run on the Novex™ 10-20% Tricine 1.0 mm Mini Protein Gel, using 1X tricine SDS buffer for 1 hour at 110 V. The Silver Staining Protocol was used to visualise the digested peptide fragments. The tricine gel was rinsed with ddH_2O, then incubated for 30 minutes with 5% glutaraldehyde solution, then rinsed again twice with ddH_2O. The gel was sensitised with 0.005% sodium thiosulfate solution for 30 minutes, rinsed with ddH_2O, and stained with silver nitrate developer until the bands were visible. Disodium EDTA solution was added once the gel was sufficiently stained.

Reagents	Concentration
Sodium thiosulfate ($Na_2S_2O_3$) (99-101% purity, CAS 7772-98-7, VWR Chemicals BDH®)	0.005%
Silver nitrate ($AgNO_3$) (63.5% Ag, CAS 7761-88-8, Johnson Matthey Chemicals Limited) solution	1%
Disodium EDTA (99.5% purity, CAS 6381-92-6, VWR Chemicals BDH®) solution	50 mM
Formaldehyde (37-41% analytical reagent grade, CAS 50-00-0, ThermoFisher Scientific) and potassium carbonate (99.99% purity, CAS 584-08-7, Merck) solution	0.036% CH_2O, 2% K_2CO_3
Glutaraldehyde (25% aq. solution, ThermoScientific Chemicals)	5%
Tricine SDS running buffer	1:10 dilution UltraPure Grade tricine (Amresco) with ddH_2O
Tricine sample buffer	46% ddH_2O, 20% 0.5 M Tris, 24% glycerol (≥ 99% purity, ThermoFisher Scientific), 10% SDS, bromothymol blue (LOT 9885600F, BDH Chemicals Ltd)

Table 2.3: Reagents used in the silver staining of tricine gel SDS-PAGE.

2.4.10 Mass spectrometry

Recombinant His6 Sp-B was expressed in BL21 (DE3) pLysS *E. coli* and purified in low imidazole Tris-based purification buffer. The sample was concentrated using an Amicon 3500 MWCO centrifugal filter (Sigma Aldrich). Purified His6 Sp-B was buffer exchanged into 20 mM ammonium acetate containing 0.1% formic acid to adjust the pH to 3. Protein concentrations were determined using a NanoDrop spectrophotometer (ThermoFisher Scientific). The samples were directly infused onto the OrbiTrap Eclipse Tribrid Mass Spectrometer (ThermoScientific) and electrosprayed in purification buffer. Sp-B fragments were computationally separated using protein deconvolution in the Bioaccord LC-MS system (Waters) for top-down native mass spectrometry analysis. The mass range was set between 500 to 2000 m/z for data acquisition.

2.5 Results

2.5.1 TEV protease cleavage conditions

Figure 2.10 shows the samples of His6 Sp-B dialysed with TEV protease at 4˚C and 16˚C respectively, at 1-hour intervals and overnight. It was observed that incubating the peptide with TEV protease for up to 6 hours at both temperatures showed no significant difference in the amount of His6 tags removed from Sp-B, due to the presence of a distinct band at approximately 27 kDa. Bands was also seen between 50 and 60 kDa at the 5th and 6th incubation hours at 16˚C. Overnight incubation with TEV at both temperatures showed a significant decrease in TEV protease present.

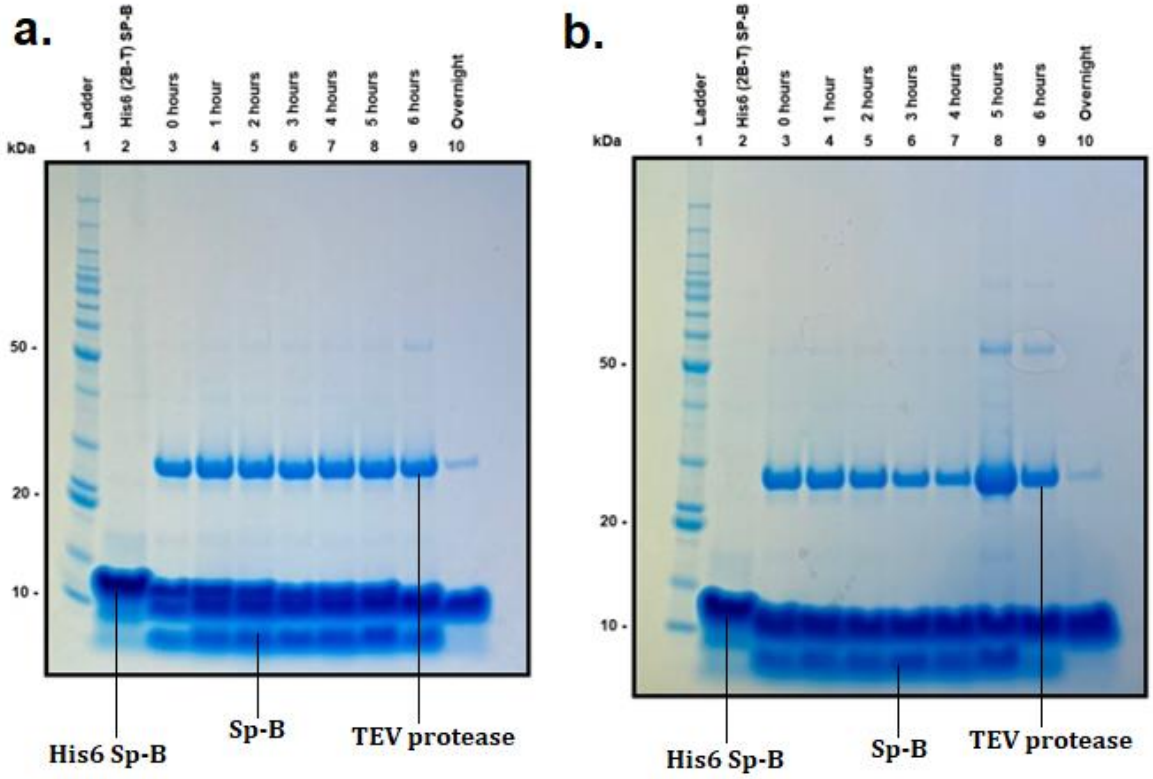


Figure 2.10: SDS-PAGE of His6 Sp-B dialysed with TEV protease at (a) 4˚C and (b) 16˚C, with samples taken and measured at 1-hour intervals from 0 to 6 hours, and overnight. **Lane 1:** BenchMark™ Protein Ladder, **Lane 2:** His6 Sp-B before incubation with TEV protease, **Lane 3-10:** Sp-B dialysed with TEV protease. The key bands in the gel are labelled accordingly.

2.5.2 SDS-PAGE and Western blot analysis of recombinant His6 Sp-B and Sp-B

The SDS-PAGE and Western blot shown in Figure 2.11 depict the samples of His6 Sp-B and Sp-B attained from affinity purification chromatography. In the Western blot, the membrane was incubated with a primary antibody specific to the His6 tag, and incubated with a secondary antibody which binds to an enzyme which allows the visualisation of protein bands on the membrane. Lane 2 shows His6 Sp-B cell extract prior to purification, where multiple bands in the SDS-PAGE were attributed to the presence of *E. coli* proteins and contaminants. In the Western blot, Lane 2 shows a singular band at around 11 kDa, indicative of the size of His6 Sp-B.

Lanes 3 to 6 show different volumes of His6 Sp-B samples loaded into the wells. In both the SDS-PAGE and Western blot, there is a strong band at around 11 kDa, indicating His6 Sp-B. In the SDS-PAGE, faint bands are seen at approximately 17 and 23 kDa. In the Western blot, faint bands are observed around 17 kDa, 25 kDa and 30 kDa.

Lanes 7 to 10 contain untagged Sp-B samples loaded at different volumes. In the SDS-PAGE gel, there is a singular band present in those lanes, right below the 10 kDa mark. This corresponds to the size of untagged Sp-B, which is 8.7 kDa. In the Western blot, there are no bands below the 10 kDa mark.

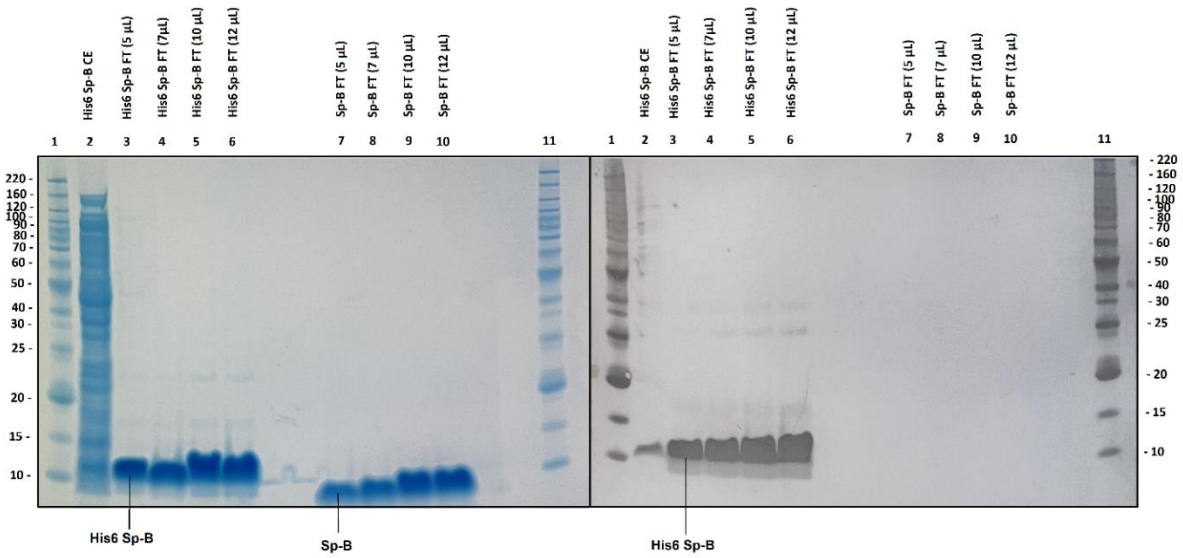


Figure 2.11: SDS-PAGE (left) and Western blot (right) of recombinant His6 Sp-B and Sp-B. The antibody used in the Western blot was anti-His6. **Lane 1, 11:** BenchMark™ Protein Ladder, **Lane 2:** His6 Sp-B cell extract, **Lane 3-6:** His6 Sp-B flowthrough, **Lane 7-10:** Sp-B flowthrough.

2.5.3 Stability of recombinant protein at different pH

The initial measured concentration of His6 Sp-B at pH 3 and 8 was 2.01 mg mL^{-1}. The concentration measured using the absorption at 280 nm, after overnight incubation of the protein at 4˚C was 1.82 mg mL^{-1} at pH 3, and mg mL^{-1} at pH 8. Figure 2.12 shows the SDS-PAGE gel run on samples of His6 Sp-B kept overnight in Tris-based buffer at pH 3 and 8. The bands present at approximately 11 kDa are indicative of His6 Sp-B.

There are multiple faint bands present at approximately 17 kDa, 23 kDa, 30 kDa and 80 kDa in the sample of His6 Sp-B incubated in the samples at pH 8. However, there appears to be fewer faint bands present in the pH 3 samples, suggesting that more contaminants are present at higher pH. The bands present in the pH 8 sample and not in the pH 3 sample are indicative of a contaminant in the sample and buffer used at pH 8, or the formation of oligomers of Sp-B which only forms at higher pH.

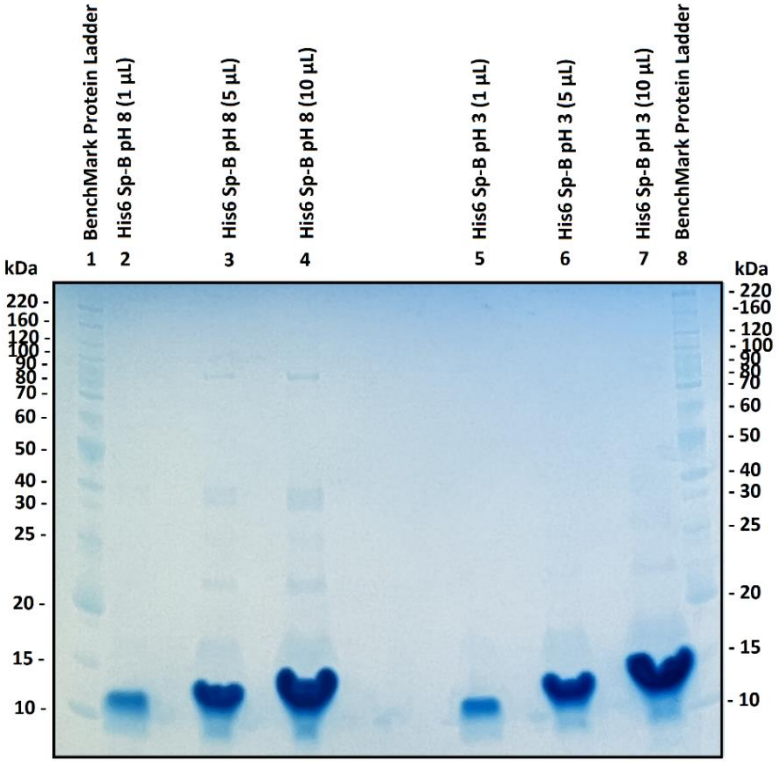


Figure 2.12: SDS-PAGE of His6 Sp-B purified at pH 8 and 3. **Lane 1, 8:** BenchMark™ Protein Ladder, **Lane 2-4:** His6 Sp-B purified at pH 8, **Lane 5-7:** His6 Sp-B purified at pH 3.

2.5.4 Trypsin digest of recombinant His6 Sp-B

Two trypsin digests were carried out on His6 Sp-B, with and without DTT reduction. Figure 2.13 shows the digest of His6 Sp-B reduced with DTT. A band at approximately 20 kDa in Lane 2 and another at 15 kDa in Lane 3 is observed. Two faint bands are observed around the 3 and 10 kDa mark in Lane 2, and a comparison to the data in Table 2.4 suggest that the two fragments represented by the faint bands may indicate the presence of uncleaved His6 Sp-B (11.1 kDa) and the 29-residue fragment (SSHHHHHHENLYFQSNI**FPIPLPYCWLCR**).

In the digest of His6 Sp-B without DTT reduction, the size of the bands on the tricine gel were approximated by calculating the logarithmic function, log10(Mw), of the distance travelled by protein marker bands divided by the distanced travelled by bromothymol blue (Figure 2.15).

A comparison of the theoretically digested protein fragment molecular weights to the molecular weights calculated from the gel in Figure 2.14 suggest that Band 1 may indicate the presence of undigested His6 Sp-B, which is approximately 11 kDa. The size of Bands 2 or 3 corresponds to a molecular weight of around 5 kDa, which may suggest the presence of the 29-residue fragment with a molecular weight of 3.6 kDa (Table 2.5). Bands 4 or 5, located below the 2 kDa mark, may indicate the presence of the 12-residue fragment with a molecular weight of 1.4 kDa (Table 2.5).

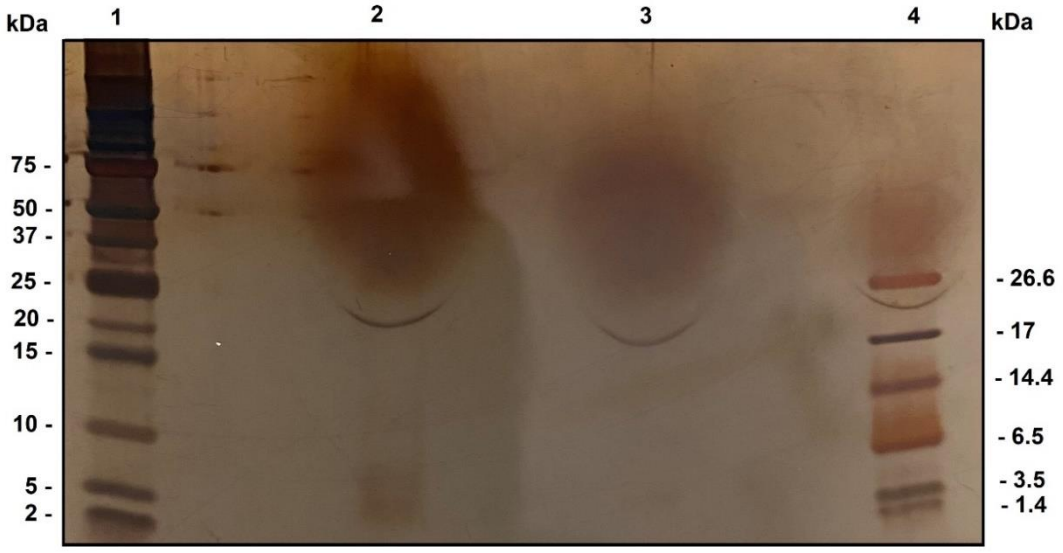


Figure 2.13: Trypsin digest of His6 Sp-B reduced with DTT, run on a Novex™ 10-20% Tricine 1.0 mm mini protein gel. **Lane 1:** BioRad Precision Plus Protein™ Dual Xtra standards + 1M DTT, **Lane 2:** trypsin-digested His6 Sp-B (8 µL), **Lane 3:** trypsin-digested His6 Sp-B (2 µL), Lane **4:** BioRad Polypeptide SDS-PAGE standard + 1M DTT.

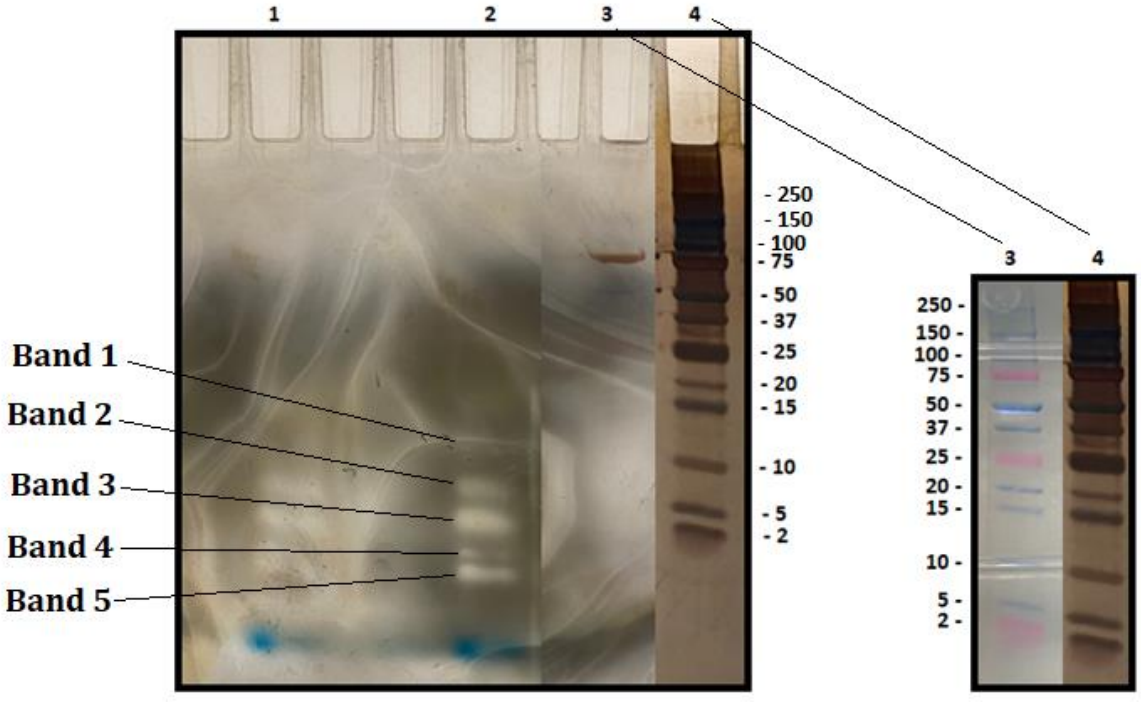


Figure 2.14: Trypsin digest of His6 Sp-B with no DTT added, run on a Novex™ 10-20% Tricine 1.0 mm mini protein gel. **Lane 1, 2:** trypsin-digested His6 Sp-B (20 µL added), **Lane 3:** BioRad Precision Plus Protein™ Dual Xtra Standard, **Lane 4:** BioRad Polypeptide SDS-PAGE standard + 1M DTT.

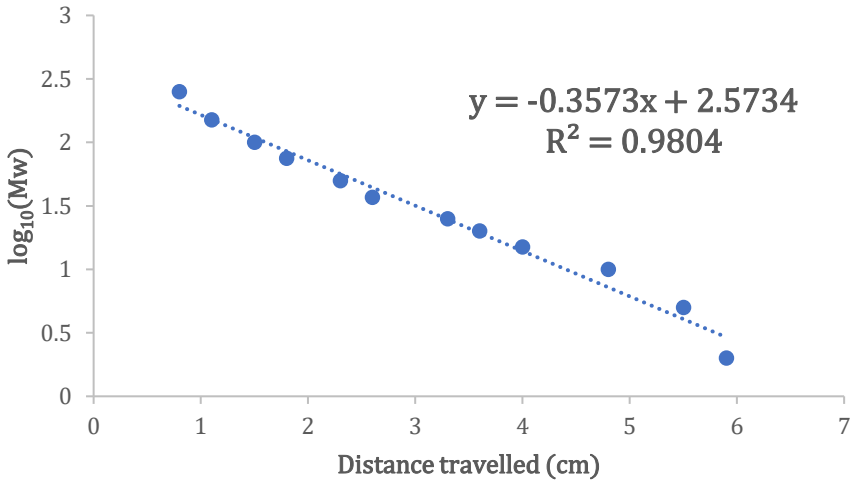


Figure 2.15: Molecular weight determination using the distance travelled by BioRad Precision Plus Protein™ Dual Xtra Standard markers on a tricine gel. The total distance moved by the protein standard was 6 cm. Logarithmic function of the molecular weight log10(Mw) in kDa was plotted against the distance travelled by the band (cm). The equation was used to calculate the molecular weight of the His6 Sp-B bands, shown in Figure 2.13 and Table 2.4.

Bands	Distance moved by His6 Sp-B (cm)	Distance moved by band / Total distance (X)	log10(kDa)	Mw (kDa)
Band 1 (faint)	4	0.513	1.1442	13.93798
Band 2	5.4	0.692	0.64398	4.40534
Band 3	5.9	0.756	0.46533	2.91964
Band 4	6.3	0.808	0.32241	2.10092
Band 5	6.6	0.846	0.21522	1.64142
Total distance	7.8	1		

Table 2.4: Molecular weight determination of His6 Sp-B bands found in Figure 2.13 (Lane 2), based on values obtained from Figure 2.14.

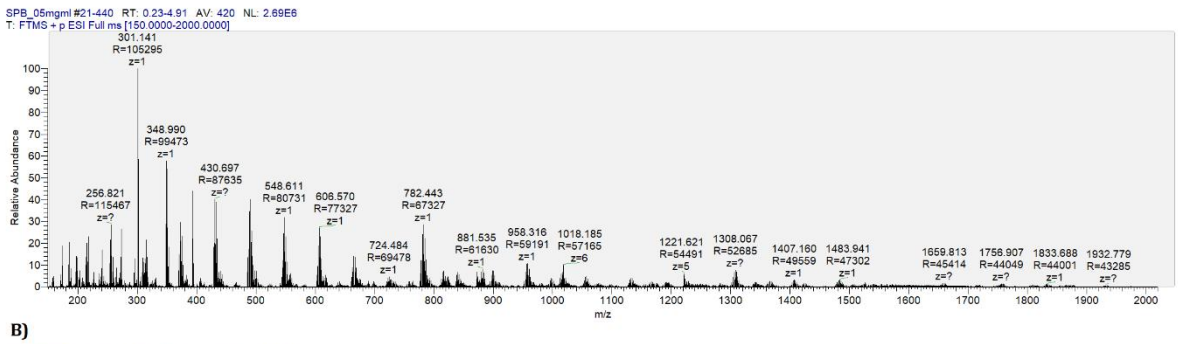
Fragments	No. of residues	Mw (kDa)
MKSSHHHHHHENLYFQSNIFPIPLPYCWLCR ALIKRIQAMIPKGALAVAVAQVCRVVPLVAGG ICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (undigested peptide)	98	11.10
SSHHHHHHENLYFQSNIFPIPLPYCWLCR	29	3.60
IQAMIPK	7	0.80
GALAVAVAQVCR	12	1.20
CVLQPLMR	8	0.90
GLLTDLLIVSYR	12	1.40
EALCQCIGGA	10	0.9

Table 2.5: Possible residue fragments generated from the trypsin digest of His6 Sp-B, where trypsin typically cleaves between lysine (K) and arginine (R) residues. His6 tag sequence is highlighted in blue.

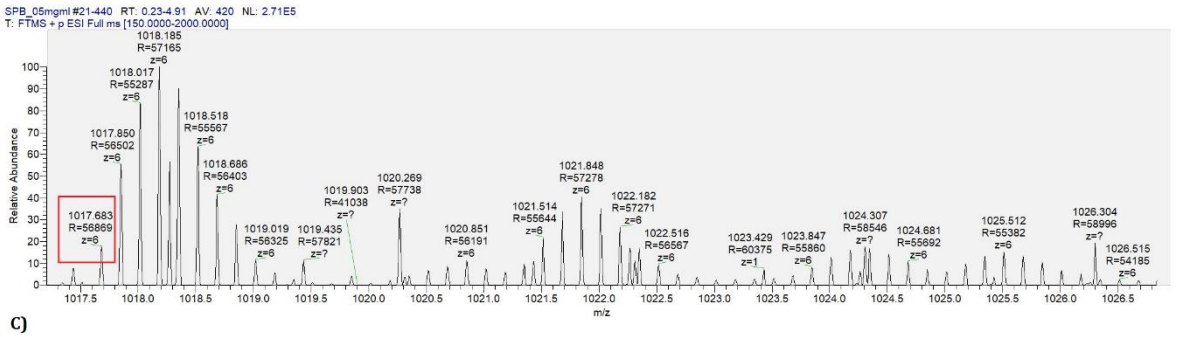
2.5.5 Direct infusion of His6 Sp-B

The full mass spectrum of His6 Sp-B revealed multiple signals with a mono-positive (+1) charge, which were not indicative of signals produced by peptides since their charge would always be more than 1. The peptide signals with multiple charges revealed the presence of fragments with a molecular weight of approximately 6.1 kDa, although this was not highly abundant. Zooming in to an m/z range of 1017.5 to 1026.5 revealed the presence of peptides with a +6 charge, and between an m/z ratio of 1222 to 1249, there were peptides with a +5 charge. Taking the first mass, 1017.683, of the isotopic pattern in Figure 2.16 (B) and multiplying that by 6 charges revealed the presence of a 6106.1 Da peptide fragment. For the peptide with +5 charges, the first peak of the isotopic pattern, 1221.019, multiplied by 5 and subtracted by 4 gives a molecular weight of 6101.1 Da. Taking the first isotopic peak as shown in Figure 2.16 (E), 1220.8, multiplied by the +5 charge and subtracting the mass of the hydrogens (5 x 1.0078) revealed a peptide fragment with mass 6099 Da. Digesting the peptide with trypsin resulted in a similar spectrum, where peptides with a +8 and +9 charge were detected. The respective molecular weights were calculated from the charge and m/z (679.123*9 and 763.888*8), again revealing the presence of a peptide fragment approximately 6.1 kDa.

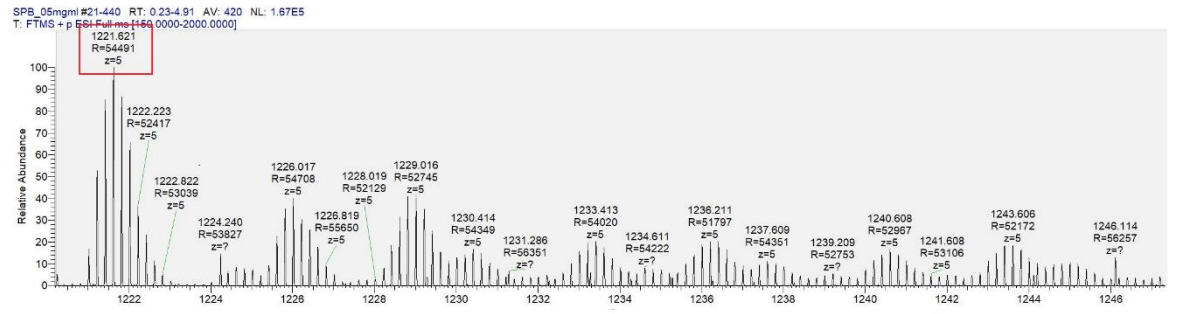
A)


SPB_05mgml #21-440 RT: 0.23-4.91 AV: 420 NL: 2.69E6
T: FTMS + p ESI Full ms [150.0000-2000.0000]
301.141 R=105295 z=1
348.990 R=99473 z=1
430.697 R=87635 z=?
256.821 R=115467 z=?
548.611 R=80731 z=1
606.570 R=77327 z=1
782.443 R=67327 z=1
724.484 R=69478 z=1
881.535 R=61630 z=1
958.316 R=59191 z=1
1018.185 R=57165 z=6
1221.621 R=54491 z=5
1308.067 R=52685 z=?
1407.160 R=49559 z=1
1483.941 R=47302 z=1
1659.813 R=45414 z=?
1756.907 R=44049 z=?
1833.688 R=44001 z=1
1932.779 R=43285 z=?
Relative Abundance
m/z

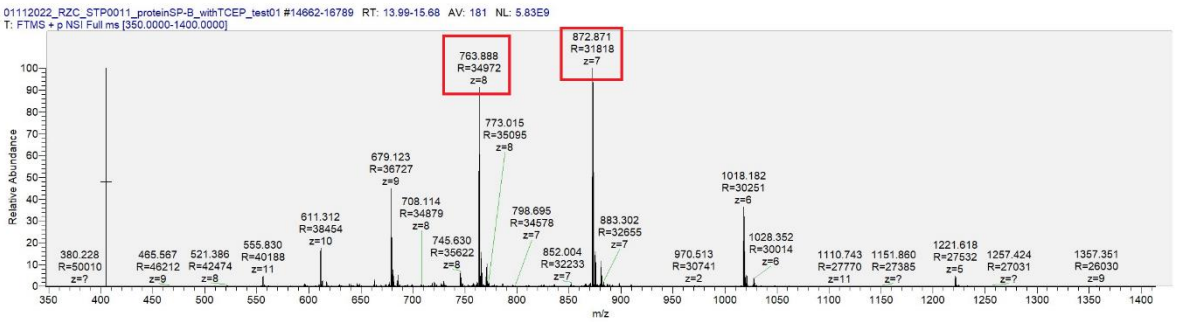
B)


SPB_05mgml #21-440 RT: 0.23-4.91 AV: 420 NL: 2.71E5
T: FTMS + p ESI Full ms [150.0000-2000.0000]
1018.185 R=57165 z=6
1018.017 R=55287 z=6
1017.850 R=56502 z=6
1018.518 R=55567 z=6
1018.686 R=56403 z=6
1017.683 R=56869 z=6
1019.903 R=41038 z=?
1020.269 R=57738 z=?
1019.019 R=56325 z=6
1019.435 R=57821 z=?
1020.851 R=56191 z=6
1021.514 R=55644 z=6
1021.848 R=57278 z=6
1022.182 R=57271 z=6
1022.516 R=56567 z=6
1023.429 R=60375 z=1
1023.847 R=55860 z=6
1024.307 R=58546 z=?
1024.681 R=55892 z=6
1025.512 R=55382 z=6
1026.304 R=58996 z=?
1026.515 R=54185 z=6
Relative Abundance
m/z

C)


SPB_05mgml #21-440 RT: 0.23-4.91 AV: 420 NL: 1.67E5
T: FTMS + p ESI Full ms [150.0000-2000.0000]
1221.621 R=54491 z=5
1222.223 R=52417 z=5
1222.822 R=53039 z=5
1224.240 R=53827 z=?
1226.017 R=54708 z=5
1228.019 R=52129 z=5
1229.016 R=52745 z=5
1226.819 R=55650 z=5
1230.414 R=54349 z=6
1231.286 R=56351 z=?
1233.413 R=54020 z=5
1234.611 R=54222 z=?
1236.211 R=51797 z=5
1237.609 R=54351 z=5
1239.209 R=52753 z=?
1240.608 R=52967 z=5
1241.608 R=53106 z=5
1243.606 R=52172 z=5
1246.114 R=56257 z=?
Relative Abundance

D)


01112022_RZC_STP0011_proteinSP-B_withTCEP_test01 #14662-16789 RT: 13.99-15.68 AV: 181 NL: 5.83E9
T: FTMS + p NSI Full ms [350.0000-1400.0000]
763.888 R=34972 z=8
872.871 R=31818 z=7
773.015 R=35095 z=8
679.123 R=36727 z=9
708.114 R=34879 z=8
798.695 R=34578 z=7
611.312 R=38454 z=10
745.630 R=35622 z=8
883.302 R=32655 z=7
852.004 R=32233 z=7
1018.182 R=30251 z=6
1028.352 R=30014 z=6
380.228 R=50010 z=?
485.567 R=46212 z=9
521.386 R=42474 z=8
555.830 R=40188 z=11
970.513 R=30741 z=2
1110.743 R=27770 z=11
1151.860 R=27385 z=?
1221.618 R=27532 z=5
1257.424 R=27031 z=?
1357.351 R=26030 z=9
Relative Abundance
m/z

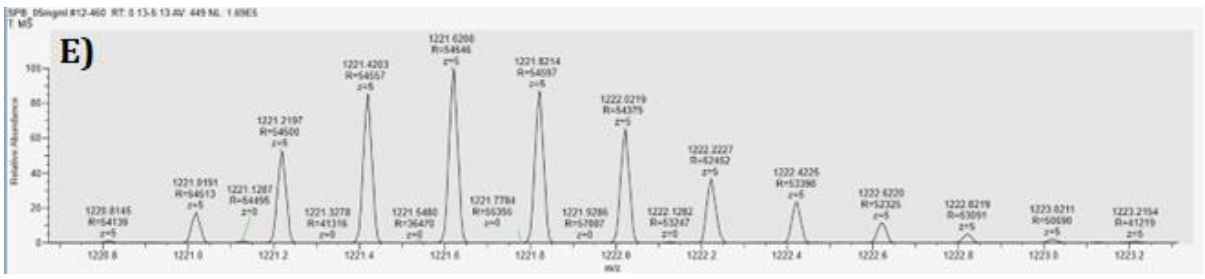

E)

Figure 2.16: Mass spectrometry data obtained from the direct infusion of Sp-B. **A)** Full spectrum of His6 Sp-B directly infused and electrosprayed with 0.1% formic acid, from an m/z range of 200 to 2000. **B)** The same spectrum (A) zoomed in to an m/z range of 1017.5 to 1026.5, showing peptides with a +6 charge. **C)** Spectrum (A) zoomed in to an m/z range of 1222 to 1246, showing peptides with a +5 charge. **D)** Spectrum on trypsin digested His6 Sp-B from m/z 350 to 1400. **E)** Spectrum zoomed in to an m/z range of 1220 to 1223.

2.5.6 Mass deconvolution of Sp-B

Mass deconvolution computationally separates ionised fragments in the sample, creating a pure spectrum for the components observed in the mass spectrometer. The charged species produced from the ionisation of Sp-B was deconvoluted to single masses. A spectrum was plotted reflecting the intensity of the observed fragments against their mass, measured in Da, as reflected in Figure 2.17, where a peak at 4051 Da (4.05 kDa) was observed. Results from Table 2.16 illustrates the possible fragments generated from the ionisation of His6 Sp-B and Sp-B. The 6.1 kDa peak recorded in the mass spectrum of His6 Sp-B matches up to the molecular weight of the N-terminal fragment cleaved between Valine 53 and Cysteine 54.

Cleaving the N-terminal fragment of Sp-B between Valine 37 and Valine 38 produces a peptide fragment of 4.1 kDa. Alternatively, cleaving the C-terminal fragment of Sp-B between Glycine 43 and 44 generates a 4.08 kDa peptide fragment. The C-terminal fragment cleaved between Glycine 44 and Isoleucine 45 produces a fragment that is 4.02 kDa. It is suggested that these were possible fragments accounting for the peak at 4.05 kDa.

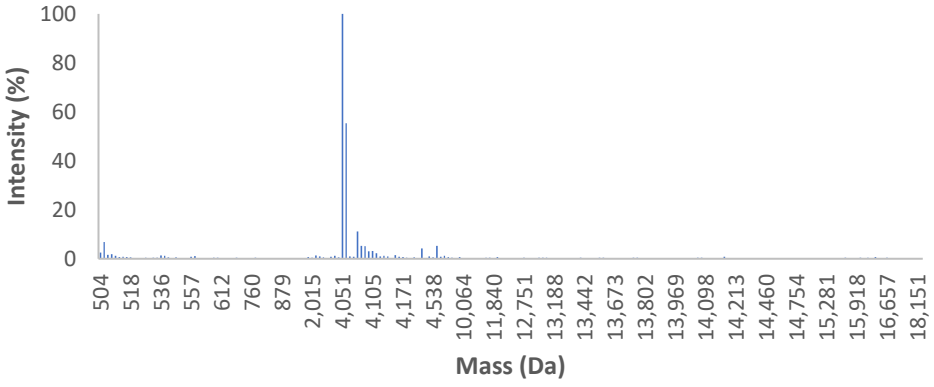


Figure 2.17: Mass deconvolution of Sp-B, with intensity plotted against mass.

Sequence	Number of residues	Mw (kDa)	pI
MKSSHHHHHHENLYFQSNIFPIPLPYCWLCR ALIKRIQAMIPKGALAVAVAQVCRVVPLVAGG ICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (His6 Sp-B sequence)	98	11.10	9.36
FPIPLPYCWLCR ALIKRIQAMIPKGALAVAVAQVCRVVPLVAGG ICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (Sp-B sequence)	79	8.71	9.44
MKSSHHHHHHENLYFQSNIFPIPLPYCWLCR ALIKRIQAMIPKGALAVAVAQV (N-terminal His6 Sp-B fragment)	53	6.1	9.59
FPIPLPYCWLCR ALIKRIQAMIPKGALAVAVAQVCRV (N-terminal Sp-B fragment	37	4.1	9.89
PKGALAVAVAQVCRVVPLVAGG ICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (C-terminal His6 Sp-B fragment)	57	6.08	9.14
GICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (C-terminal Sp-B fragment)	36	4.08	8.55
ICQCLAERYSVILLDTLLGRMLPQLVCRLVLR CSM (C-terminal Sp-B fragment)	35	4.02	8.55
SSHHHHHHENLYFQSNIFPIPLPYCWLCR	29	3.60	6.98
IQAMIPK	7	0.80	8.75
GALAVAVAQVCR	12	1.20	8.25
CVLQPLMR	8	0.90	8.25
GLLTDLLIVSYR	12	1.40	5.84
EALCQCIGGA	10	0.9	4.00

Table 2.6: Possible fragments generated from the ionisation of His6 Sp-B.

2.6 Discussion

TEV protease is commonly used in the removal of the His6 affinity tag from the recombinant protein. Its highly specific nature enables the cleavage of the amino acid recognition sequence (ENLYFQGS), between the glutamine (Q) and glycine (G) residues. TEV protease binds to the sequence on the His6-tagged protein, effectively removing the affinity tag. As a result, the His6-TEV protease can be removed by affinity chromatography, with the untagged protein eluting in the flowthrough. Results suggest that an overnight incubation of His6 Sp-B with TEV protease would remove most affinity tags from the protein.

At hours 5 and 6 of incubation at 16ºC, the presence of a band at approximately 55 kDa indicated the presence of impurities or the dimerisation of TEV protease. Due to its structure, TEV protease functions efficiently as a monomer, as shown in Figure 2.10.

Moreover, membrane proteins like Sp-B are highly sensitive to changes in temperature. Lower temperatures may therefore improve the stability, solubility, and conformation of highly hydrophobic proteins like Sp-B especially when expressed outside of their native lipid environment. Lower temperatures may also lead to a decrease in hydrophobic interactions, which may reduce protein aggregation. Hence, it was determined that incubating His6 Sp-B overnight at 4˚C would lead to the most favourable outcome.

SDS-PAGE data primarily shows information on the molecular weights of the proteins in the sample. By comparing the molecular weights of the bands present in the gel to the molecular weight of the target protein, the bands can indicate the presence of the target protein or contaminants in the sample. For example, multiple bands seen in the SDS-PAGE of unpurified His6 Sp-B cell extract clearly show the presence of *E. coli* proteins and contaminants. Results shown in Figure 2.12 suggest that it is likely that oligomers of His6 Sp-B form at pH 8, which do not form at pH 3. It is possible that contaminant proteins precipitate at lower pH since precipitated protein must be resolubilised before showing up on an electrophoresis gel.

Data from the SDS-PAGE and Western blot confirmed that TEV protease cleaved all His6 tags from Sp-B. The presence of the bands below 10 kDa in the SDS-PAGE wells containing Sp-B, which were not observed in the Western blot show that the bands most likely indicate the presence of untagged Sp-B which could not be detected by the antibodies.

The presence of the His6 tag improved the stability of the recombinant protein; however, both tagged and untagged recombinant Sp-B could not be concentrated past 2 mg mL^{-1}, which may be the solubility limit of Sp-B, since further concentration resulted in precipitation and sample loss. The hydrophobic nature of Sp-B is emphasised in high concentrations when intermolecular interactions increase. Increasing the number of interactions between Sp-B molecules increase the chances of interactions between hydrophobic regions that are normally buried within the structure to minimise contact with the aqueous environment. As a result, forcing the hydrophobic regions of Sp-B closer together causes more precipitation to occur.

Incubating concentrated His6 Sp-B overnight at 4˚C at pH 8 usually resulted in some precipitation. When the pH of the buffer is lowered to 3, no visible precipitation was observed, and the concentration of the sample remained stable. Similarly, Sp-B was seen to precipitate less when kept at lower pH. This suggests that the presence of hydrophobic interactions strongly influence the folding of Sp-B, where a lowered pH can strengthen hydrophobic interactions within the protein, improving its structural stability. A consequence of protein precipitation is partial unfolding, exposing more hydrophobic regions to the aqueous environment. Sp-B possibly becomes less soluble at lower pH, reducing aggregation and structural unfolding, resulting in increased stability. The charges on Sp-B at pH 8 possibly resulted in reduced electrostatic repulsion between the molecules, causing precipitation.

The size, hydrophobicity and presence of 3 intramolecular disulfide bonds in the structure of His6 Sp-B as shown in Figure 2.15 and 2.16, may hinder the ability of trypsin to cleave the protein at lysine and arginine residues. Membrane proteins like Sp-B also typically have domains surrounded by lipids, which may inhibit digestion by trypsin. In theory, the addition of DTT would reduce the disulfide bonds formed between cysteine residues in His6 Sp-B. Additionally, the usage of denaturants like urea in future experiments may increase the exposure of cleavage sites on His6 Sp-B. It was observed that not reducing the protein with DTT resulted in fragments observed on the gel. This may be attributed to His6 Sp-B's reliance on the intermolecular disulfide bonds, which help to maintain its conformation.

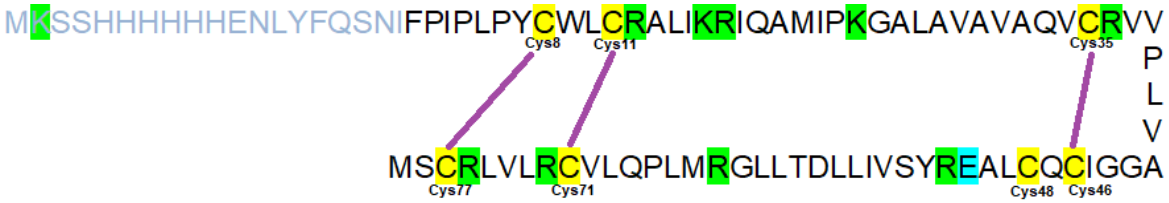


Figure 2.18: Diagram showing the amino acid residues, highlighted in green, where trypsin cleaves. Cysteine residues are highlighted in yellow.

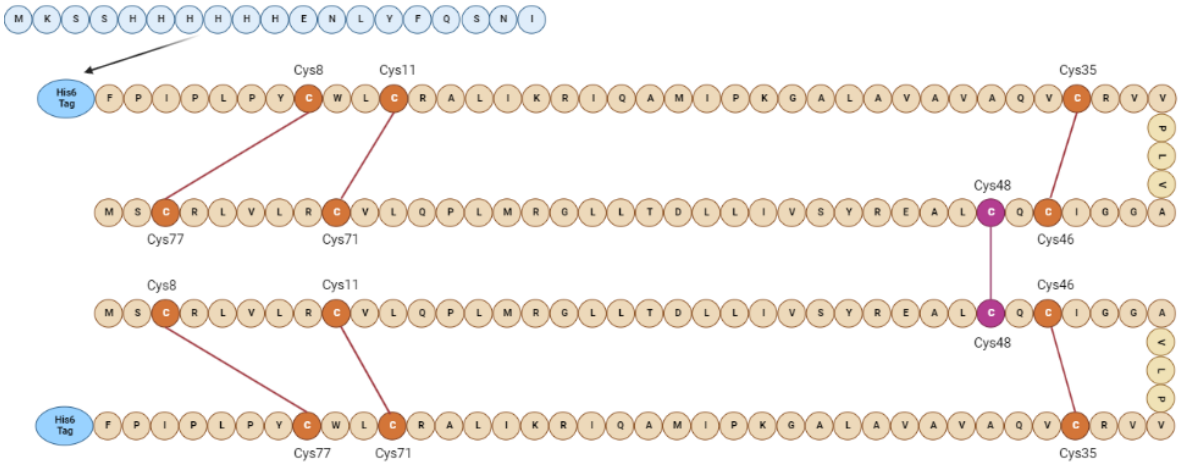


Figure 2.19: Diagram showing the formation of intermolecular (purple) and intramolecular (dark orange) disulfide bonds between cysteine residues in His6 Sp-B. Sp-B is known to dimerise by forming a disulfide bond at the Cys48 residue.

Mass spectrometry results from the fragmentation of His6 Sp-B showed that there is a peak at 6.1 kDa, while a peak at 4.1 kDa was observed in the Sp-B sample. A 53-residue N-terminal fragment of His6 Sp-B, cleaved between Valine 53 and Cysteine 54, has a molecular weight of 6.1 kDa which could account for the peak observed at 6.1 kDa. However, there was no peak observed in the Sp-B sample at 3.75 kDa, which is the expected molecular weight of the same N-terminal fragment without the His6 tag. Without this information, the peak at 6.1 kDa cannot be confirmed to be the N-terminal fragment. In both samples, there was no signal observed at the expected molecular weight of full-length His6 Sp-B or Sp-B. The lack of fragments observed may indicate difficulties in ionising samples of the proteins, due to their hydrophobicity and tendency to interact with lipids. Despite precipitating the protein to remove bound lipids, it is likely that many peptides were still lipid-bound, rendering them unable to 'fly' in the mass spectrometer.

2.7 Conclusions

The results suggest that a pure sample of recombinant His6 Sp-B was obtained from affinity purification using a HisTrap column, based on SDS-PAGE analysis with Coomassie staining and Western blotting. It was concluded that size exclusion chromatography mainly resulted in a large loss of sample and was not necessary in obtaining a pure sample of recombinant protein, since most contaminants and *E. coli* proteins were observed to be removed in affinity purification. Digestion of the sample with TEV successfully removed the His-tag from the Sp-B but the resulting protein was less stable. To supplement SDS-PAGE data, comparing the intensity of the protein bands to pre-determined standards such as bovine serum albumin (BSA) may help in quantitatively determining the concentration of His6 Sp-B present. SDS-PAGE data alone cannot provide structural information on the expressed protein. Methods like CD spectroscopy may therefore shed some insight into the secondary structural elements of recombinant His6 Sp-B and Sp-B. In the next chapter, such methods will be explored to elucidate the secondary structure and intermolecular disulfide bonds that are formed within the expressed protein.

Given the size of His6 Sp-B, the resolution of the tricine SDS-PAGE gel provides a qualitative estimate of the molecular weight of expected fragments that can be seen from a trypsin digest. Although it was discovered that His6 Sp-B remained more stable at lower pH, trypsin optimally hydrolyses protein at pH 7.8 and overnight at 37˚C (Chelulei Cheison et al., 2011). Further optimising the parameters to accommodate trypsin activity and the stability of Sp-B may be necessary for mass spectrometry analysis.

The ambiguity of the trypsin digests suggests that both peptides were very resistant to trypsin digestion, possibly due to the hydrophobicity and presence of multiple intramolecular disulfide bonds. As a result, the top-down mass spectrometry approach was used to ionise the samples. Although no conclusive data was obtained regarding the fragmentation patterns of full length His6 Sp-B and Sp-B, the main objective of this research focuses on getting a stable buffer system for His6 Sp-B and Sp-B. Ultimately, the nature of such membrane proteins makes it challenging to express and purify them in large quantities without resulting in aggregation or precipitation, limiting their analysis in mass spectrometry.

Chapter 3 – Structural Analysis of Expressed Surfactant Protein B using Circular Dichroism Spectroscopy

3.1 Introduction

3.1.1 Circular dichroism spectroscopy

Circular dichroism (CD) is a light absorption spectroscopy technique used to structurally characterise proteins via the differential absorption of left and right-handed circularly polarised light. Polarised light refers to how light wave oscillations are orientated, which can be broadly defined as linearly or circularly polarised light. Light waves in linearly polarised light oscillate in a single plane, going up or down at a specific angle in that plane. In circularly polarised light, light waves rotate in a circular helical pattern, continuously changing the electric field (Figure 3.1).

In CD spectroscopy, left-handed polarised light is produced when the electric field rotates counterclockwise, and clockwise to produce right-handed polarised light. As shown in Equation 3.1, the CD signal is a measure of differential absorbance, which is the difference or change in absorbance (ΔA) between left (A_{LCP}) and right (A_{RCP}) circularly polarised light. The absorptions of left and right circularly polarised light are calculated by multiplying the extinction coefficient of left circularly polarised light (ε_{LCP} or ε_{RCP}) by the concentration and the pathlength used. The extinction coefficient of the CD signal is therefore measured as the difference in extinction coefficients, $\Delta\varepsilon$, between left and right circularly polarised light (Equation 3.2).

$$CD_{signal} = \Delta A = A_{LCP} - A_{RCP} \quad (3.1)$$

$$CD = \Delta\varepsilon = \varepsilon_{LCP} - \varepsilon_{RCP} \quad (3.2)$$

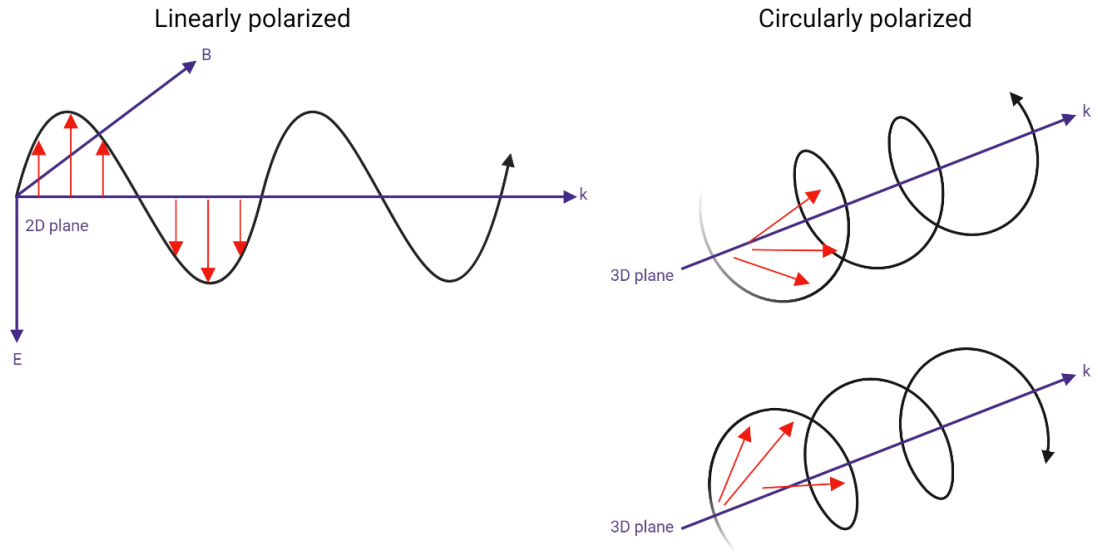


Figure 3.1: Diagram showing how linearly and circularly polarised light travels. The electric field in linearly polarised light oscillates in a uniform line, whereas in circularly polarised light it rotates clockwise (right-handed polarised light) or counterclockwise (left-handed polarised light). Created in BioRender.

3.1.2 Applications of circular dichroism spectroscopy

CD spectroscopy allows the analysis of folding pathways, dynamics, and interactions between proteins. CD spectra measured between 260 nm to about 180 nm, where an n → π* at approximately 222 nm and at approximately 208 nm and 190 nm are the two main types that occur and give rise to the CD signal (Miles and Wallace, 2016; Miles, Ramalli and Wallace, 2022). The spectra are analysed to detect various secondary structural elements, namely alpha helices, parallel and anti-parallel beta sheets, turns and others.

CD spectroscopy is relevant in the analysis of membrane protein data, which often require different methods to maintain their structure and function in an amphipathic environment. The crystal structure of such proteins is often challenging to solve; hence CD spectral data may provide insight into fold recognition studies and in structural genomics. In a study by Wallace, membrane proteins have different spectral characteristics compared to soluble proteins, with the effects of these characteristics on the calculation of secondary structures of membrane proteins using reference databases of soluble protein spectra being analysed. It was found that inaccurate results were produced when these databases were used to analyse the CD spectra of membrane proteins (Wallace, 2003). There are many possible reasons to account for the discrepancy

in results, with the solvent shift phenomenon being a likely reason. Ideally, using a reference database with a range of secondary structural elements determined from membrane proteins could improve the analysis of membrane protein CD spectra.

3.1.3 Estimating the secondary structure of Sp-B using preexisting CD spectra

The shape of a CD spectra can reveal information on a protein's secondary structure, due to characteristic peaks produced in each spectrum. Figure 3.2 illustrates the typical CD spectra obtained from proteins each containing purely alpha-helical, anti-parallel beta sheet, turns or other folds respectively, measured from approximately 180 nm to 250 nm. The spectrum of the alpha-helical protein shows a sharp peak at 190 nm, with 2 dips at 208 nm and 225 nm respectively. For the anti-parallel beta sheet protein spectrum, a peak is observed at 195 nm, with a small dip at around 213 nm. A protein containing mostly turns in its structure produces a spectrum with a smaller peak at 205 nm, and 2 dips at 190 nm and 225 nm respectively. Proteins containing predominantly other secondary structures result in a spectrum characterised by a large dip at 198 nm, and two small peaks at 184 nm and 218 nm respectively.

A limitation of CD spectroscopy is that in general, it can only be used for samples that absorb weakly. If the absorbance is too high, then hardly any light would be reaching the detector, since virtually all the light would have been absorbed. This results in only noise being measured in the CD spectra; hence the results are not meaningful. When recording the spectra, the voltage, also known as the high tension (HT), is required to detect the signal. A high HT value indicates a low signal, where in these experiments, a HT value of above 600 V would indicate that the CD signal is mainly noise and hence, not considered meaningful.

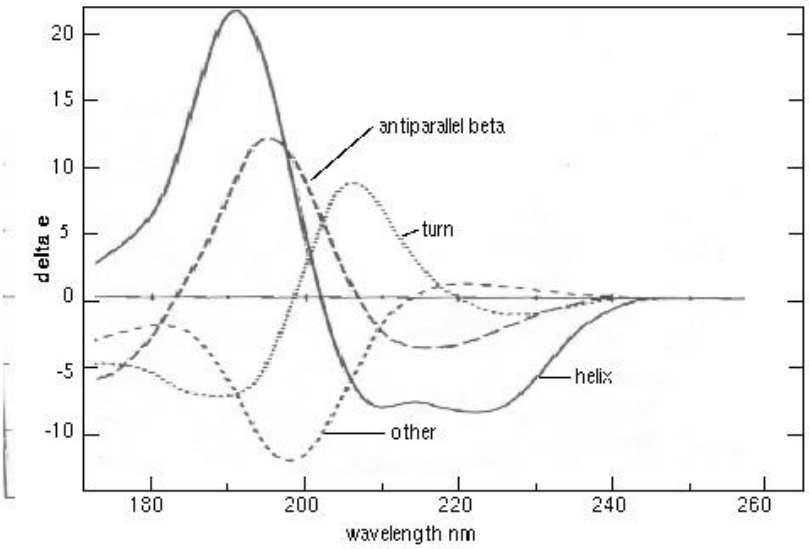


Figure 3.2: CD spectra measured using 'pure' secondary structures; alpha helix, antiparallel beta sheets, turns and other. Delta e (ΔƐ) stands for the difference in the extinction coefficient for left and right circularly polarised light. Adapted from (Brahms and Brahms, 1980).

As mentioned in previous chapters, the hydrophobic nature of Sp-B makes it susceptible to precipitation, especially in aqueous solutions. It was established in Chapter 2 that the expression of Sp-B did not yield high amounts of protein, so recording the CD spectra of His6 Sp-B and Sp-B in trifluoroethanol (TFE) was a reasonable alternative to analysing the CD spectra of Sp-B in micelles or lipid vesicles. TFE is therefore a particularly useful solvent in stabilising the alpha helical structure of Sp-B and minimising the formation of aggregates. TFE has been thought to mimic a lipid membrane environment; it is a strong hydrogen bond acceptor which interacts with the peptide bonds in proteins, to promote the formation of alpha helical secondary structures. TFE stabilises helical regions in proteins by inducing helical folds in regions that are not as well-defined in aqueous solutions. Hence, it is hypothesised that adding increasing concentrations of TFE to Sp-B would lead to an increase in alpha helicity. Additionally, Super Mini-B and Mini-B, which are peptide mimics of Sp-B, contain two intramolecular disulfide bonds between Cysteine 8 and 34, and Cysteine 11 and 40 (Waring *et al.*, 2023) (Figure 3.3). It is shown that the experimentally determined structure of Mini-B has a predominantly alpha helical structure; Cys1 to Ile15 forming the N-terminal helix, and Leu22 to Leu31 forming the second helix nearer to the C-terminal, leaving only 8 residues that are not in any helical structure (Figure 3.4).

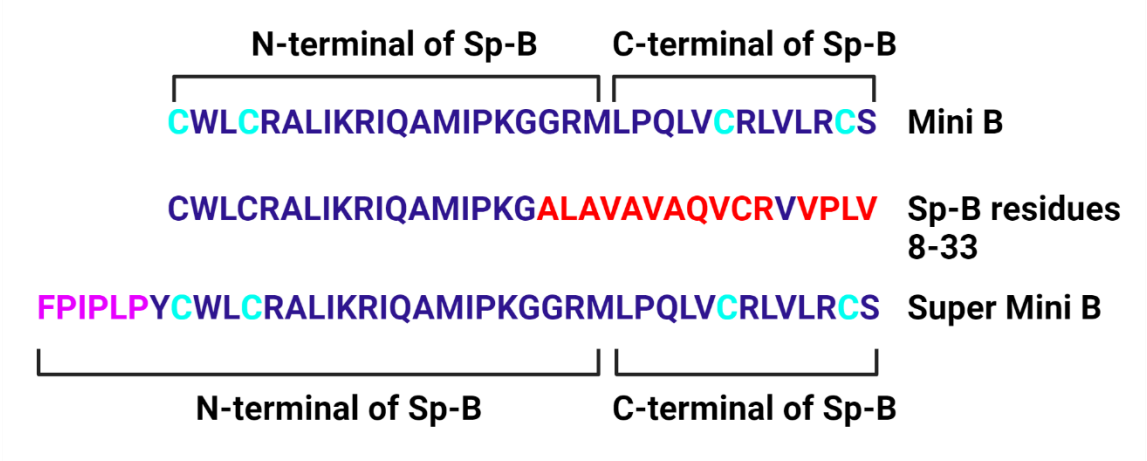


Figure 3.3: Amino acid sequences for synthetic Sp-B peptide constructs. Mini-B (top) comprises 34 residues total, made up of residues 8-25 from the N-terminal end of Sp-B and residues 63-78 from the C-terminal end of Sp-B. Middle: sequence of Sp-B from residues 8-33. Bottom: Super Mini-B (41 residues total) is made up of residues 1-29 from the N-terminal end of Sp-B and residues 63-78 of the C-terminal end of Sp-B. Made in BioRender.

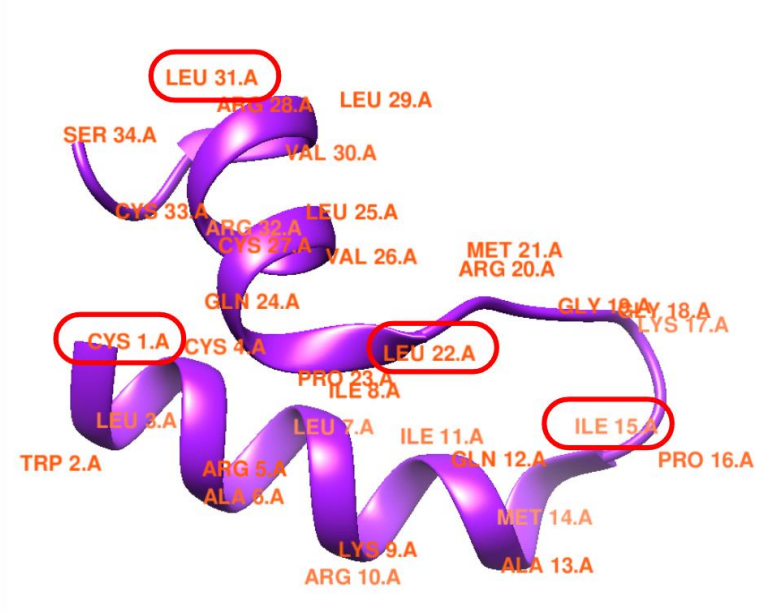


Figure 3.4: Experimentally determined structure of the peptide mimic Mini-B. Residues Cys1 to Ile15 form the N-terminal helix, and residues Leu22 to Leu31 form the C-terminal helix. Made in Chimera and BioRender.

3.1.4 Analysis programmes used in CD spectroscopy

The base curves obtained from multi-linear regression, singular value decompositions or neural networks of a CD spectral set may vary greatly depending on the type of proteins used as standards in the database due to differences in the CD spectra resulting from the presence of disulfide bridges, aromatic amino acids, or rare conformations (Kuwajima, 1996). Hence, when choosing an optimal analysis programme to determine the conformation of an unknown protein, the protein standards used should have similar spectral characteristics to that of the unknown protein. In DichroWeb, the available methods include VARSLC, SELCON3, CONTIN, CDSSTR and K2D (Whitmore and Wallace, 2007). Generally, such programs use algorithms to deconvolute the CD spectra and quantitatively determine the amounts of secondary structure types are present in a protein by comparing the spectra to experimental CD spectra with known protein structures in reference datasets.

The variable selection (VARSLC) program allows the significant improvement of protein conformation estimation by combining them with singular value decompositions. An initial protein database containing known secondary structures and spectra are selected, with the spectra processed to be systematically eliminated to create new databases containing fewer standards. Singular value decomposition is subsequently used on the newly reduced databases to elucidate the conformation of the unknown protein sample. The results are compared and analysed, where the datasets fulfilling the selection parameters for an optimal fit are averaged. However, this method is not optimal for analysing data for Sp-B within the 300 to 180 nm range, since the buffers that Sp-B were purified in absorbed strongly at wavelengths below 190 to 200 nm.

The SELCON3 program is a self-consistent method is a modification of the variable selection method (VARSLC program), having an advantage over VARSLC due to its improved accuracy and speed. The program arranges the protein database in an order of increasing root mean squared difference from the analysed CD spectrum, and spectra which have the largest difference from the spectrum of interest are systematically deleted to minimise the time taken to find the best solutions (van Stokkum *et al.*, 1990). SELCON3 is also based on the finding that prediction of the unknown protein improves if it is included in the basis set to minimise bias towards the test protein structure. SELCON3 makes an initial guess of the unknown protein's structure and includes that conformation

in the database to be deconvoluted using singular value decomposition. A limitation in using the program lies in the fact that the currently available database of reference spectra poorly predicts polypeptide structures with high beta sheet content, as SELCON3 grossly overestimates the proportion of alpha helices while underestimating the proportion of beta sheets present in the structure (Greenfield, 1996). SELCON is recommended especially for the estimation of globular protein structures in solution, but not for polypeptides with high beta-pleated sheet content.

The ridge regression analysis (CONTIN) program is based on the theory that the CD spectra of unknown proteins could be directly fit using a linear combination of a large database of protein spectra with known conformations. CONTIN uses a variation of the least squares method, like the mathematical ridge regression method. The contribution of the unknown protein and reference protein spectra are significant only when both contribute to a good agreement between the raw data and the theoretical best fit curve (Greenfield, 1996). Compared to using the convex constraint algorithm, singular value decomposition or simple multiple linear regression, CONTIN gives a better estimate of beta turns in the protein structure. Prediction of the protein conformation is unaffected even with data truncation at 200 nm. However, the fit of the data determined by CONTIN is heavily dependent on the proteins used in the standard protein database, which may be countered and improved with the inclusion of denatured proteins as references for random conformations (Venyaminov *et al.*, 1993). This program can be used in conjunction with SELCON to optimise the overall estimates of both polypeptide and protein conformations.

CDSSTR is a computational method that combines features of other analysis programmes, where a minimal number of reference proteins are needed for good data analysis. Proteins from the reference set are randomly selected for the analysis of the CD spectrum, where many combinations can be made from the various reference proteins due to the flexible nature of the method. Solutions of each combination are obtained from the singular value decomposition algorithm (Grund, 1979), however, the random selection of proteins from the dataset may result in instability of the solutions. Generally, if the solution meets the requirements of three selection parameters, namely the sum of fraction value being between 0.95 to 1.05; with each fraction being larger than -0.03; and the root mean squared deviation (RMSD) of the reconstructed and experimental CD

spectra being less than 0.25 ΔЄ, which is the difference in extinction coefficient for left and right circularly polarised light.

A fixed minimum number of admissible solutions are subsequently subjected to another selection parameter based on the helical content, determined from the estimation of the helix fraction from the full reference set and the helix fractions of admissible solutions (Johnson, 1999; Sreerama and Woody, 2000).

K2D refers to a neural network computer program which detect both correlations and patterns in data. Neural networks are computational techniques comprised of three types of units, namely the input units which receive external signals from sources to be sent to the other units; the output units which receive the signals from the input units and send signals to the environment; and hidden units which receive inputs from other units and send output signals to other units without directly receiving the data or the final results of the output signals (Greenfield, 1996). These units can be organised into layers, with neuronal connections between the units located in different layers, which are numerically weighted for the input data to be correctly output. In the analysis of CD data, CD spectra are input patterns, and the weights of the secondary structures are output patterns. In the learning phase of neural networks, connections are made between the CD spectra points and the secondary structure standards. The weights of the connections are adjusted to a point where the value of the error between the calculated and actual secondary structures is as low as possible. The recall phase of neural networks utilises unused data from the learning phase to calculate the corresponding output using the adjusted weights. In a study by Böhm *et al.*, the K2D neural network used had an input layer with 83 units, corresponding to 83 wavelengths between 260 and 178 nm in the CD spectrum. It also possesses a hidden layer with 45 neurons, and an output layer with 5 neurons corresponding to the alpha helix, beta turns, antiparallel and parallel beta sheets, and remaining structures. A very good prediction of alpha helical and antiparallel beta sheets, with correlation coefficients of 1.0 and 0.91 obtained. Despite truncation of the wavelength data between 250 and 200 nm, prediction of the proportion of alpha helices remained accurate while beta sheet prediction resulted in a negative correlation coefficient (Böhm, Muhr and Jaenicke, 1992). K2D significantly improves the correlation between calculated and observed secondary structures and is good at estimating beta

sheet structures in polypeptides, between a spectrum of 240 and 200 nm, but fails to estimate the proportion of beta turns in the structure.

In a study by Sreerama *et al.*, various protein reference sets also containing denatured proteins were analysed by the CDSSTR, SELCON3 and CONTIN programs. While CDSSTR performed best in the 29-protein reference set, with SELCON3 performing the worst. However, CDSSTR performed the worst in a 37-protein reference set whereas both SELCON3 and CONTIN performed better. In a 43-protein reference set, all three methods had a similar NRMSD value and had similar performance levels. It was seen that with larger wavelength range and smaller reference sets, the performance of CDSSTR was the best, and with smaller wavelength range and larger reference sets, CONTIN was the best. A comparison of the performance indices (root mean squared differences and correlation coefficients of X-ray and CD-predicted values) of individual secondary structures showed that no method was superior in providing the best performance indices for all secondary structures, as each method gives insignificant differences for any one of the analysed secondary structures. Generally, a larger protein reference set resulted in an improved performance compared to smaller sets, due to a bigger representation of variations in the structure and spectra of beta sheets and turns. The combined usage of all analysis programmes is therefore recommended to reliably analyse the secondary structure of an unknown protein sample (Sreerama and Woody, 2000).

3.1.5 Protein reference sets used in the analysis of CD spectra

The soluble protein 175 cutoff (SP175) reference set is created based on bioinformatical principles to include fold and secondary structure space. It exclusively comprises 98 soluble proteins and outputs the proteins' predicted spectra in the 175 – 260 nm wavelength range. The proteins were collected in 150 mM NaCl/NaF and 50 mM phosphate buffer (pH 7) or under crystallisation buffer conditions to examine the environmental sensitivity of the spectra (Lees et al., 2006).

The CATH (Class, Architecture, Topology, Homology superfamily) distribution of the SP175 dataset revealed a higher proportion of mainly-beta and mixed alpha/beta proteins, compared to mainly alpha helical proteins, as seen in Figure 3.5. This is an advantageous feature for CD reference datasets due to mainly-beta structures being more

spectrally diverse compared to mainly alpha helical structures (Wallace *et al.*, 2004), and allows for improvement in predicting the proportions of secondary structures in an unknown protein sample.

The SP175 dataset encompasses 72 proteins having different CATH classifications, 19 of which have different architectures and 55 of which have different topologies. 61 out of 72 proteins are single-domain proteins, possessing only a single CATH fold type. The spectral diversity of the superfolds are demonstrated in Figure 3.6, where the native nine superfolds of the single-domain proteins are shown. To optimise the analysis of secondary structures, the maximal wavelength range (170 nm) should be used to obtain decent prediction results of the protein's helical content. SP175 can still predict secondary structural content even with data truncation at higher wavelengths (200 nm). The identification of features exclusive to lower wavelength data, by cluster analysis or neural networks may value the lower wavelength data more so compared to the higher wavelength data, further highlighting the importance of SP175 being a broad reference database.

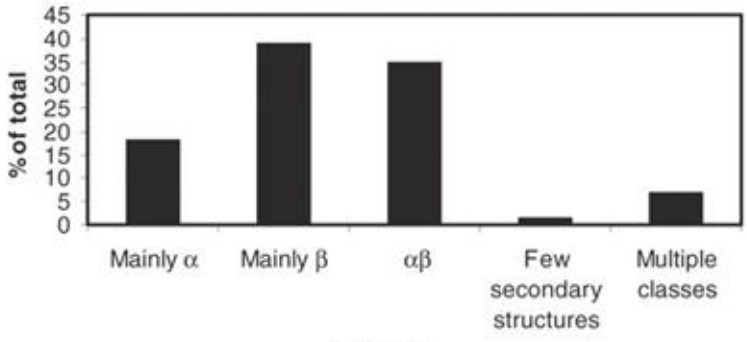


Figure 3.5: Secondary structure distribution of the SP175 dataset proteins (Lees et al., 2006).

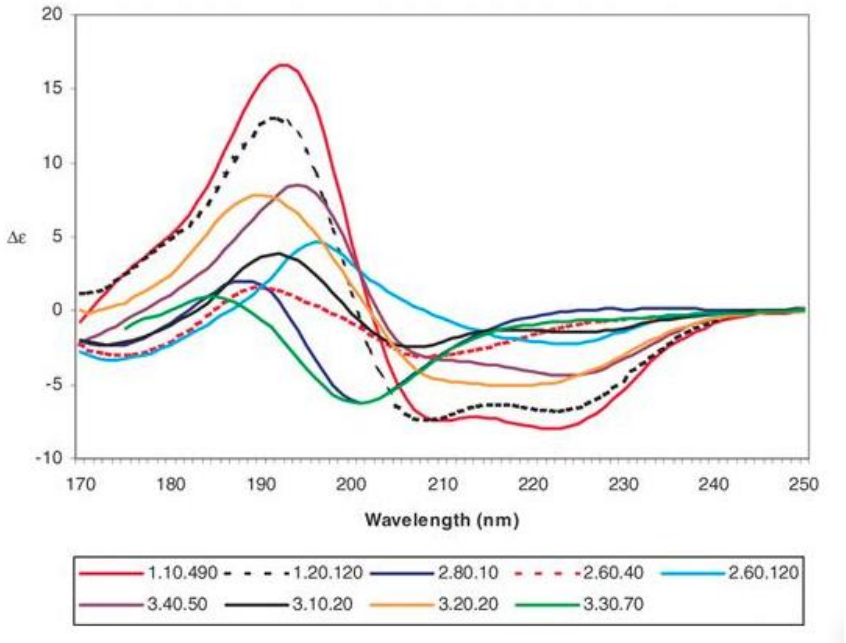


Figure 3.6: The nine CATH superfolds of single-domain proteins of the SP175 reference database. The numbers in the box represent the CATH code of proteins with a specific superfamily classification. For example, 1.10.490 refers to proteins belonging to the mainly alpha class (1), orthogonal bundle architecture (1.10), with globin-like topology (1.10.490) (Lees et al., 2006).

SMP180 contains all proteins in the SP175 reference set, with 30 additional membrane proteins to improve the overall prediction of both membrane and soluble proteins (Table 3.1). This reference set outputs the proteins' predicted spectra in the 180 – 260 nm range. SMP180 improves analysis of membrane proteins with large ectodomains, and prediction of both soluble and membrane protein secondary structures are greatly improved. A study by Abdul-Gader *et al.* employed the 'leave one out' cross-validation method, to determine the accuracy of calculating secondary structures, on both the principal and alternative secondary structure classification schemes. All datasets of the principal scheme had high correlations for helical structures, with the SMP180 dataset giving best overall results. The alternative scheme had high correlations for all secondary structures except for turns. The study tested the mixed soluble/membrane protein SMP180 dataset against the soluble protein $SP175_{t180}$ dataset, and results of the alternative scheme (Table 3.3) yielded significant improvements compared to that of the principal scheme (Table 3.2). Compared to MP180 and $SP175_{t180}$, SMP180 generally produced great results for all secondary structures in both alternative and principal schemes, suggesting that it was

therefore preferable in the analysis of membrane and soluble proteins (Abdul-Gader, Miles and Wallace, 2011).

Name : **SMP180**		Range (nm) : **180-240**		Classifications: **αR αD βR βD T U**	
SMP180 includes all the proteins of SP175 plus the proteins which have these PDB codes:					
1qhj	2fed	2cfq	2nu9	2nop	1hzx
1l7v	1rh5	2dyr	1j95	2oar	2gif
1pcr	2wjn	1be3	1t5s	1nek	1xl4
2j58	3jqo	1uyn	1aos	2fcp	1fep
1nqh	2vdf	2jwy	1h2s	2a65	1nkz

Table 3.1: Protein reference sets of soluble SMP180 proteins respectively (Lees et al., 2006; Abdul-Gader, Miles and Wallace, 2011).

Dataset	Test set	Parameter	Secondary structure type		
			helix (*H* +*G*)	β -sheet (*E*)	*O*
$SP175_{t180}$	MP180	*r*	0.920	0.923	0.465
		δ	0.125	0.114	0.093
		ζ	2.035	2.056	0.876
SMP180	MP180	*r*	0.931	0.937	0.595
		δ	0.118	0.092	0.067
		ζ	2.166	2.547	1.224
SMP180	$SP175_{t180}$	*r*	0.940	0.879	0.745
		δ	0.074	0.088	0.067
		ζ	2.838	1.824	1.487

Table 3.2 Predicted performances of SMP180 dataset for the $SP175_{t180}$ soluble proteins, and the $SP175_{t180}$ dataset for the MP180 membrane proteins using principal secondary structure categories (helix (H + G), sheet (E) and 'other' (O)) (Abdul-Gader, Miles and Wallace, 2011).

Dataset	Test set	Parameter	Secondary structure type					
			α_R	α_D	β_R	β_D	Turn	*U*
$SP175_{t180}$	MP180	*r*	0.862	0.770	0.890	0.805	0.173	0.434
		δ	0.102	0.049	0.118	0.030	0.028	0.074
		ζ	1.972	1.203	1.665	1.442	0.945	1.003
SMP180	MP180	*r*	0.928	0.871	0.940	0.896	0.227	0.642
		δ	0.077	0.032	0.090	0.021	0.030	0.060
		ζ	2.702	1.886	2.224	2.022	0.945	1.302
SMP180	$SP175_{t180}$	*r*	0.949	0.813	0.839	0.880	0.365	0.713
		δ	0.050	0.035	0.064	0.027	0.043	0.061
		ζ	3.145	1.693	1.852	1.820	0.991	1.405

Table 3.3: Predicted performances of SMP180 dataset for the $SP175_{t180}$ soluble proteins, and the $SP175_{t180}$ dataset for the MP180 membrane proteins using alternative secondary structure categories (regular helix (α_R), distorted helix (α_D), regular β-sheet (β_R), distorted β -sheet (β_D), turns and unordered (U)) (Abdul-Gader, Miles and Wallace, 2011).

The results yielded from the various analysis programmes used will depend on the proteins included in the base reference set. Different reference datasets have been made over the years, containing spectra of different proteins which in turn, produce different results for calculated secondary structures that depend on the characteristics of the proteins in the reference set, and that of the unknown protein. An unknown protein sample with relatively standard structure will usually result in all analysis programmes and reference databases producing similar results. However, a protein possessing unusual structural elements, such as 3_{10} or polyproline II helices, or the presence of disulphide bridges or aromatic sidechains, may impact the ability of these algorithms to produce and replicate an experimental spectrum of the unknown protein. This is a common problem with membrane proteins, as all existing protein reference databases were obtained from soluble protein spectra, with peaks shifted relative to that of membrane proteins.

The development of databases with broad coverage of protein secondary structure will lead to improvements in the accuracy of these analysis methods. To determine which

algorithm and database yields results that best fit the experimental spectrum, it is recommended to test out different combinations of reference sets and analysis programmes used. Protein reference sets 1 and 2 accept a wavelength range of 178 to 260 nm, proteins in sets 3 and 6 have a wavelength range of 185 to 240 nm and proteins in set 5 have the broadest wavelength range of 178 – 260 nm. As seen in table 3.4, reference sets 4 and 7 consist of proteins in the 190 to 240 nm wavelength range, with 6 output secondary structure classifications, namely αR (regular alpha helix/helix 1), αD (distorted alpha helix/helix 2), βR (regular beta sheet/sheet 1), βD (distorted beta sheet/sheet 2), T (turns) and U (unordered) (Sreerama and Woody, 2000). Reference sets 1, 3 and 6 also have the same secondary structural classification outputs, however, reference sets 2 and 5 have different secondary structure definitions which needs data to cover the 178 – 260 nm wavelength range, namely α (alpha helix), β (beta sheet), 3/10 (3_{10} helix), P2 (polyproline II helix), T (turns) and U (unordered). Only reference set 2 can be used for the analysis of 3_{10} helices, while both set 2 and 5 can be used for analysing polyproline II helices (Lees et al., 2006).

Reference Set 4 (190 – 240 nm)	**Reference Set 7 (190 – 240 nm)**
A-Bungarotoxin (ABNG)	A-Bungarotoxin (ABNG)
Alcohol dehydrogenase (ADH)	**Apo-cytochrome C (5C) denatured (ACY5)**
Adenylate kinase (ADK)*	**Apo-cytochrome C (90C) denatured (ACY9)**
Azurin (AZU)	Alcohol dehydrogenase (ADH)
B-lactoglobulin (BLAC)	Adenylate kinase (ADK)*
Bence Jones Protein (BNJN)	Azurin (AZU)
Bovine pancreatic trypsin inhibitor (BPTI)*	B-lactoglobulin (BLAC)
Carbonic anhydrase (CANH)	Bence Jones Protein (BNJN)
CGA	Bovine pancreatic trypsin inhibitor (BPTI)*
A-chymotrypsin (CHYT)	Carbonic anhydrase (CANH)
Colicin A (COLA)	CGA
Concanavalin A (CONA)	A-chymotrypsin (CHYT)
Carboxypepsidase A (CPAS)*	Colicin A (COLA)
Cytochrome C (CYTC)	Concanavalin A (CONA)
EcoR1 Endonuclease (ECOR)	Carboxypepsidase A (CPAS)*
Elastase (ELAS)	Cytochrome C (CYTC)

Flavodoxin (FLVD)	EcoR1 Endonuclease (ECOR)
G-crystalin (GCR)	Elastase (ELAS)
Green fluorescent protein (GFP)	Flavodoxin (FLVD)
Glyceraldehyde 3-p dehydrogenase (GPD)	G-crystalin (GCR)
Glutathione reductase (GRS)	Green fluorescent protein (GFP)
Haemoglobin (HBN)	Glyceraldehyde 3-p dehydrogenase (GPD)
Hemerythrin (HMRT)	Glutathione reductase (GRS)
Rat intestinal fatty acid binding protein (IFBP)	Haemoglobin (HBN)
Insulin (INSL)*	Hemerythrin (HMRT)
Lactate dehydrogenase (LDH)	Rat intestinal fatty acid binding protein (IFBP)
Lysozyme (LYSM)	Insulin (INSL)*
Myoglobin (MGLB)	Lactate dehydrogenase (LDH)
Nuclease (NUCL)*	Lysozyme (LYSM)
Papain (PAPN)	Myoglobin (MGLB)
Parvalbumin (PARV)*	Nuclease (NUCL)*
Phosphoglycerate kinase (PGK)	**Ribonuclease (20C) denatured (OX20)**
Pepsinogen (PPSN)	Papain (PAPN)
Prealbumin (PRAL)	Parvalbumin (PARV)*
Rhodanase (RHD)	Phosphoglycerate kinase (PGK)
Ribonuclease A (RNAS)	Pepsinogen (PPSN)
Substilin BPN (SUBB)	Prealbumin (PRAL)
Substilin novo (SUBN)	Rhodanase (RHD)
Superoxide dismutase (SUDS)	Ribonuclease A (RNAS)
T4 lysozyme (T4LS)	**Staphylococcal nuclease (6C) denatured (SN06)**
Thermolysin (THML)	**Staphylococcal nuclease (70C) denatured (SN70)**
Tumour necrosis factor (TNF)	Substilin BPN (SUBB)
Triose phosphate isomerase (TPI)	Substilin novo (SUBN)
	Superoxide dismutase (SUDS)
	T4 lysozyme (T4LS)
	Thermolysin (THML)
	Tumour necrosis factor (TNF)
	Triose phosphate isomerase (TPI)

Table 3.4: Proteins included in Reference Sets 4 and 7. Proteins denoted with * are proteins used exclusively for Sets 4 and 7, while proteins in bold are proteins that are included in Set 7 but not Set 4 (Wallace, 2003).

3.2 Methods

3.2.1 Materials

His6 Sp-B and Sp-B protein samples used in CD spectroscopic measurements were obtained from expression in competent *E. coli* cells as previously described in Chapter 2. The samples were purified in Tris-based buffer containing NaCl and buffer exchanged via overnight dialysis at 4˚C into phosphate-based buffer containing NaF (25 mM sodium phosphate, 300 mM NaF, 10% glycerol) at pH 8. This was done to exchange highly absorbing Cl^- present in the Tris-based buffer to an equivalent buffer containing F^- ions, which do not absorb as strongly at the measured wavelengths. TFE (99.5% NMR grade, CAS 75-89-8, Alfa Aesar) was added to the protein samples at concentrations of 10-50%. All samples were centrifuged at 12,800*g* for 15 minutes to remove insoluble material and precipitate.

Sample concentrations were determined using the absorption measured at 280 nm using a NanoDrop™ spectrophotometer. The extinction coefficients of His6 Sp-B and Sp-B were calculated to be 0.900 and 0.974 $(mg\ mL^{-1})^{-1}\ cm^{-1}$ respectively, assuming all cysteine residues were reduced, by inputting the amino acid sequences into the ProtParam ExPASy tool (Gasteiger et al., 2003). To calculate the mean residue weight in Daltons, the molecular weight of the protein sample was divided by the total number of residues subtracted by 1. The mean residue weight of His6 Sp-B was calculated to be 114.18 Da, and 111.67 Da for Sp-B.

3.2.2 CD spectroscopy measurements

CD spectroscopy measurements were conducted in the Birkbeck Biophysics lab, using the Jasco J-20 Spectropolarimeter. The protein samples were placed in a quartz Suprasil QS High Precision Cell (Hellma UK Ltd) with a 1 mm pathlength to optimise the CD signal. The CD spectra were measured at wavelength range between 300 nm to 185 nm at 20˚C, and the HT value threshold was set at 600 V. The signal value was averaged across 12 runs, and data processing was carried out using DichroWeb and CDSSTR reference set proteins.

3.2.3 Selecting the analysis program method

The SELCON3, CONTIN, CDSSTR and K2D analysis programmes were used alongside four reference set proteins; Set 4 (optimised for 190 – 240 nm), Set 7 (optimised for 190 – 240 nm), SP175 (optimised for 190-240 nm) and SMP180 (optimised for 190-240 nm). Reference set 4 comprises 43 proteins while set 7 consists of 48 proteins. Set 7 includes all the reference proteins used in Set 4, with additional 5 denatured proteins (ACY5, ACY9, OX20, SN06 and SN70). However, a comparison of the NRMSD values (Table 3.5) across all reference sets and analysis programmes used show that Set 4 fits the experimental data of His6 Sp-B better than Set 7. This indicates the possibility of more denatured proteins in the experimental Sp-B sample, where a more well-fitted spectrum is inclusive of unfolded/denatured proteins. Generally, the usage of reference sets containing more proteins is also preferable. However, if the experimental spectra had an equally good fit using both Sets 4 and 7, this would suggest that the reference proteins in Set 4 were sufficient to achieve a good fit to the experimental data.

	SELCON3	CONTIN	CDSSTR	K2D
Set 4 (190 to 240 nm)	0.122	0.158	0.040	0.763
Set 7 (190 to 240 nm)	0.103	0.158	0.043	0.763
SP175 (190 to 240 nm)	0.188	0.117	0.067	0.763
SMP180 (190 – 240 nm)	-	0.071	0.042	0.763

Table 3.5: The NRMSD values of the spectra measured by four analysis programmes (SELCON3, CONTIN, CDSSTR, K2D) and the corresponding protein reference sets used (Set 4, Set 7, SP175, SMP180).

DichroWeb includes a broad range of algorithms as well as protein reference databases to accept data produced using any conventional CD or synchrotron radiation CD instrument, subsequently calculating a single-fit parameter and the normalised root mean squared deviation (NRMSD) to allow comparisons between methods and databases (Miles, Ramalli and Wallace, 2022). Protein spectra reference datasets are used in

combination with least squares methods, principal component analysis or machine learning techniques to calculate a spectrum of the protein secondary structure which best fits the experimental spectrum (Miles and Wallace, 2016). The NRMSD is a measure of how well the experimental spectrum corresponds to the calculated best fit spectrum. Generally, a lower NRMSD value correlates with a better fit of the spectra, where an NRMSD value of less than 0.1, alongside other parameters, is essential for a reliable conclusion of the secondary structural data. As seen in Table 3.5, five datasets have an NRMSD value below 0.1, indicating that these combinations of analysis programmes and reference sets allowed both the experimental and calculated spectra to fit well. For the four reference sets used, CDSSTR performed the best, with Set 4 reference proteins (190 – 240 nm) having the lowest NRMSD value of 0.040. K2D however, performed the worst with all four reference sets having an NRMSD value of 0.763. The CONTIN program performed moderately well with the various protein reference sets, with NRMSD values ranging from 0.071 to 0.158. The SELCON3 program produced moderately good results with Set 4 (190 – 240 nm) and Set 7 (190 – 240 nm) reference protein datasets, however with SP175 (190 – 240 nm) and SMP180 (190 – 240 nm) reference sets the program failed to find a valid solution. This was likely due to a bad fit of the experimental and reference datasets, or analysis of the protein sample by SELCON3 resulted in only low-quality solutions. Hence, the optimal combination to analyse the CD spectra of Sp-B would be to use the CDSSTR analysis programme with Set 4 reference set proteins.

3.3 Results

3.3.1 CD spectra of diluted and undiluted His6 Sp-B

Figure 3.7 shows the CD spectra of diluted and undiluted His6 Sp-B. For all CD measurements, the signal produced by the buffer is subtracted from the total value, to give the signal of the protein alone. The His6 Sp-B spectra shows a dip at 220 nm, and the HT value increases beyond 600 V at shorter wavelengths. As explained earlier, data obtained when the HT value is above 600 V should be treated with caution due to high noise levels. The more concentrated sample has no reliable data beyond about ~225 nm. In the diluted sample, two dips are observed at 203 nm and 224 nm, with the HT value increasing past the threshold at wavelengths below 205 nm. The signal value of Sp-B shows two dips at 207 nm and 224 nm as shown in Figure 3.8, with the threshold of the HT value exceeded at a wavelength of below 210 nm.

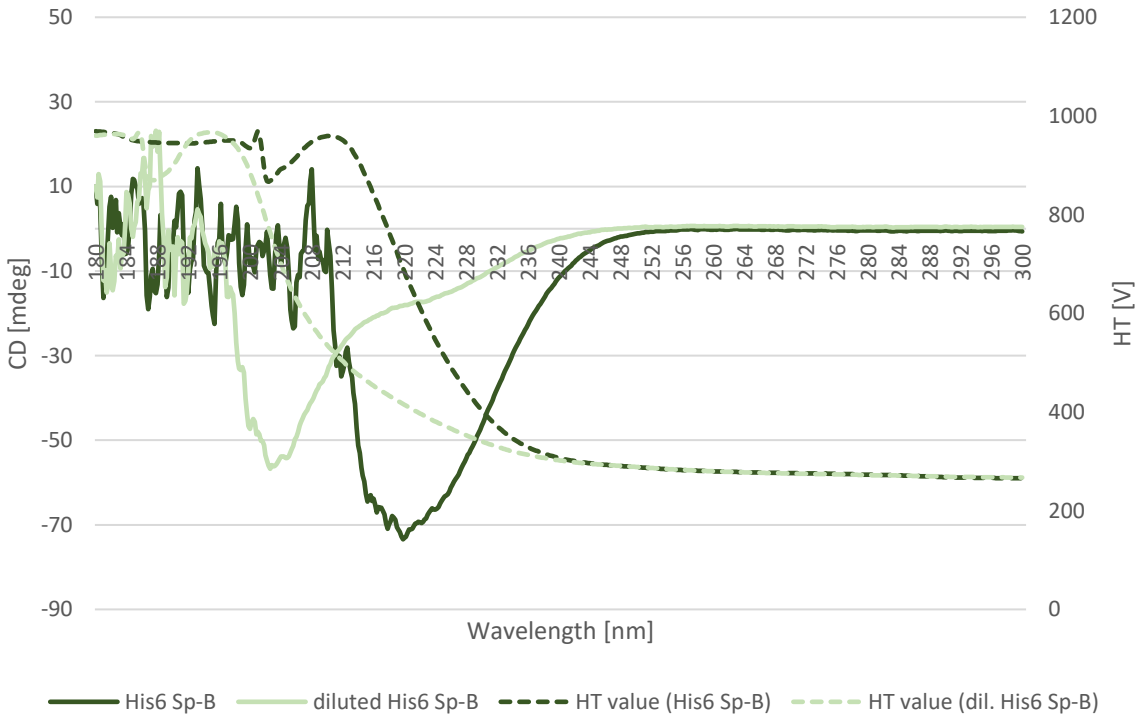


Figure 3.7: CD spectra of diluted His6 Sp-B [0.103 mg mL^{-1}] and undiluted His6 Sp-B [0.428 mg mL^{-1}] purified in NaF-$NaPO_4$ buffer. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

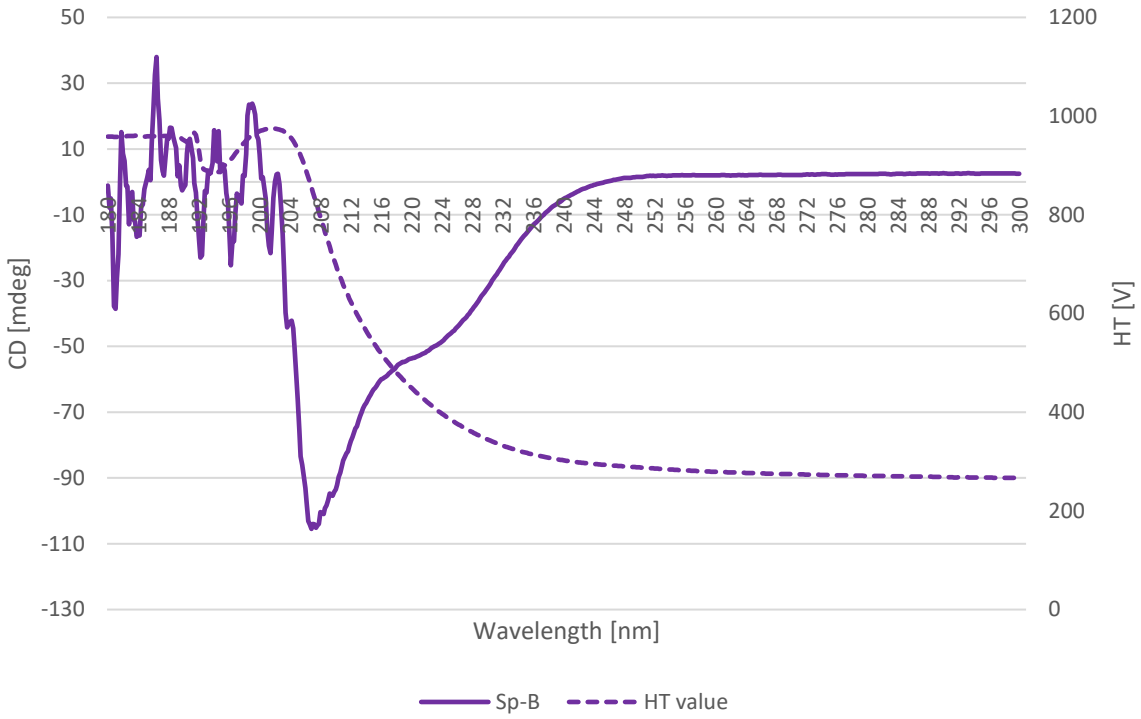


Figure 3.8: CD spectra of Sp-B [0.061 mg mL^{-1}] purified in NaF-$NaPO_4$ buffer. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

The CD data of diluted His6 Sp-B and Sp-B were analysed using DichroWeb, with the NRMSD values implying a good fit of the experimental to reconstructed data (Figure 3.9). Using Set 4 reference proteins, CDSSTR detected two alpha helices (Helix 1, 2) and two beta sheets (Strand 1, 2) in His6 Sp-B and Sp-B (Table 3.6). In 100 residues of His6 Sp-B, there were 4.980 helix segments found with an average helix length per segment of 12.798 in the structure of His6 Sp-B. This means that 64% of its residues were found within the two detected alpha helices. There were 2.504 strand segments per 100 residues in the structure, with an average strand length per segment of 4.802, resulting in 12% of its residues residing in the two beta sheets. Approximately 8% of residues were in turns, and 16% were unfolded.

In 100 residues of Sp-B, there were 2.332 helix segments with an average helix length per segment of 22.322, resulting in a total of 52% of its residues detected within the two alpha helices. There were 5.045 strand segments per 100 residues with an average strand length per segment of 5.034, where 25% of Sp-B residues were found in beta sheet strands. 5% of the residues in Sp-B were in turns, and 19% were unfolded.

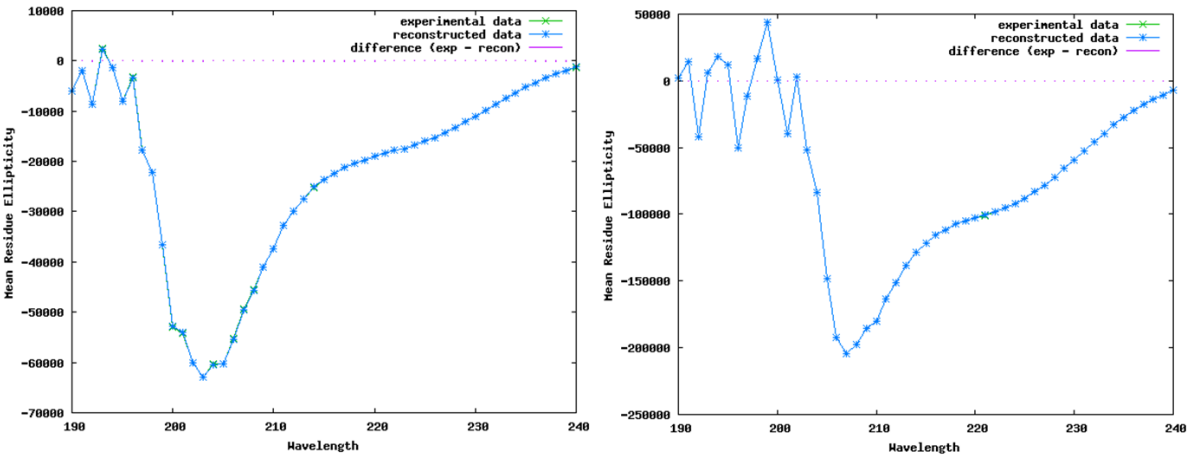


Figure 3.9: Graphical output of diluted His6 Sp-B (left) and Sp-B (right) generated in DichroWeb (NRMSD $\leq$ 0.001). Experimental (input) data are plotted in green, and reconstructed data determined from the calculated output secondary structure are plotted in blue. Differences in input and reconstructed data are shown in purple.

	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unfolded
His6 Sp-B	0.44	0.20	0.07	0.05	0.08	0.16
Sp-B	0.43	0.09	0.15	0.10	0.05	0.19

Table 3.6: The calculated secondary structure results obtained for the closest matching solutions for His6 Sp-B and Sp-B.

3.3.2 Addition of trifluoroethanol to His6 Sp-B and Sp-B

In another set of experiments, the CD spectra of His6 Sp-B and Sp-B with and without the addition of 40% TFE, to mimic the membrane environment, was recorded as seen in Figure 3.10 and 3.11. Two dips at 204 nm and 220 nm were observed in the spectra for His6 Sp-B, and at 206 nm and 223 nm for His6 Sp-B with 40% TFE. The HT value reached the 600 V threshold for the samples at approximately 206 nm without TFE and at 215 nm with 40% TFE.

The spectra for Sp-B without TFE added had two dips at 201 nm and 221 nm, while Sp-B with 40% TFE had dips at 205 nm and 220 nm. There is a peak observed at around 190 nm for the sample containing 40% TFE, however spectral data obtained below the HT value threshold of 199 nm implies that there is too much noise around the peak.

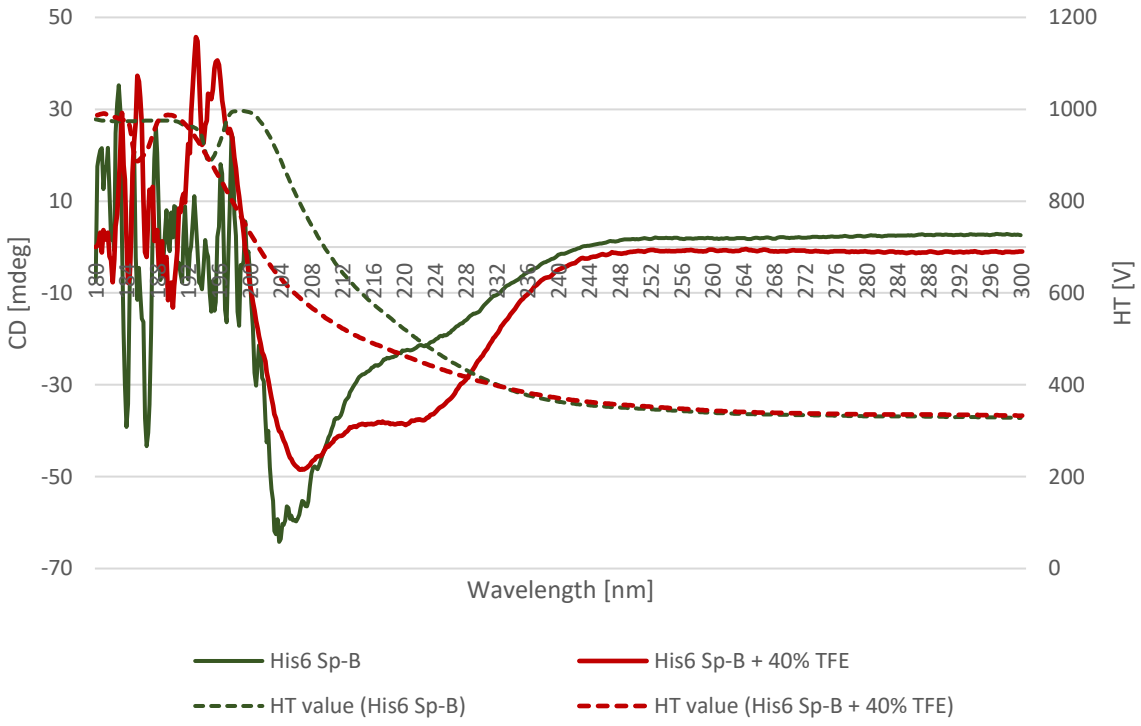


Figure 3.10: CD spectra of His6 Sp-B [0.167 mg mL^{-1}] and His6 Sp-B with 40% TFE [0.132 mg mL^{-1}] purified in NaF-NaPO$_4$ buffer. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

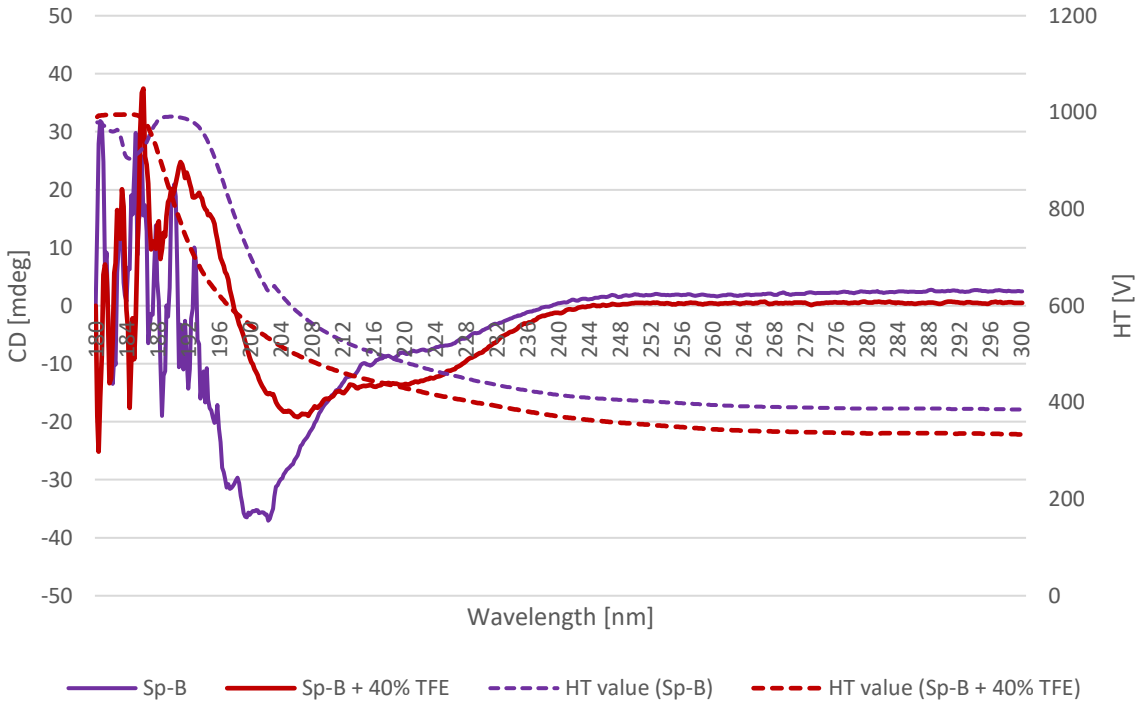


Figure 3.11: CD spectra of Sp-B [0.022 mg mL^{-1}] and Sp-B with 40% TFE [0.028 mg mL^{-1}] purified in NaF-NaPO$_4$ buffer. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

The spectral data were subsequently plotted in DichroWeb, where all datasets had a good correspondence between experimental and reconstructed data (Figure 3.12). There were 4.517 helix segments per 100 residues of His6 Sp-B with an average helix length per segment of 14.248, resulting in 64% of its residues residing in the two detected alpha helices (Helix 1 and 2). There were 3.06 strand segments per 100 residues with an average strand length per segment of 5.1, resulting in a total of 15% of its residues in the two beta sheets (Strand 1 and 2). 6% and 15% of His6 Sp-B residues were found in turns and unfolded structures respectively.

In the sample of His6 Sp-B with 40% TFE, there were 5.148 helix segments per 100 residues with an average helix length per segment of 13.174, resulting in approximately 68% of all residues in the two alpha helices. 2.166 strand segments per 100 residues with an average strand length per segment of 4.894, meaning 10% of its residues were found in the beta sheets. Like the sample without TFE added, 6% and 15% of its residues were found in turns and unfolded structures respectively.

In 100 residues of Sp-B, there were 1.588 helix segments with an average helix length per segment of 27.007. This means that approximately 43% of its residues were a part of alpha helical structures. There were 5.847 strand segments with an average strand length per segment of 4.978, showing that 29% of its residues were in beta sheets. In Sp-B, 7% of its residues resided in turns, with 21% of residues being in unfolded structures. In the sample of Sp-B with 40% TFE, approximately 72% of residues were in alpha helices, 11% of residues in beta sheets, 4% of residues in turns and 12% of residues in unfolded structures.

Table 3.7 shows the proportion of secondary structures in the samples, where the addition of 40% TFE showed an increase in alpha helical content for both samples. The total alpha helical content of His6 Sp-B increased from 64% to 68%, while Sp-B was observed to have a significant increase in alpha helical content from 43% to 72%. Moreover, the percentage of unfolded residues in Sp-B decreased from 21% to 12% with the addition of TFE.

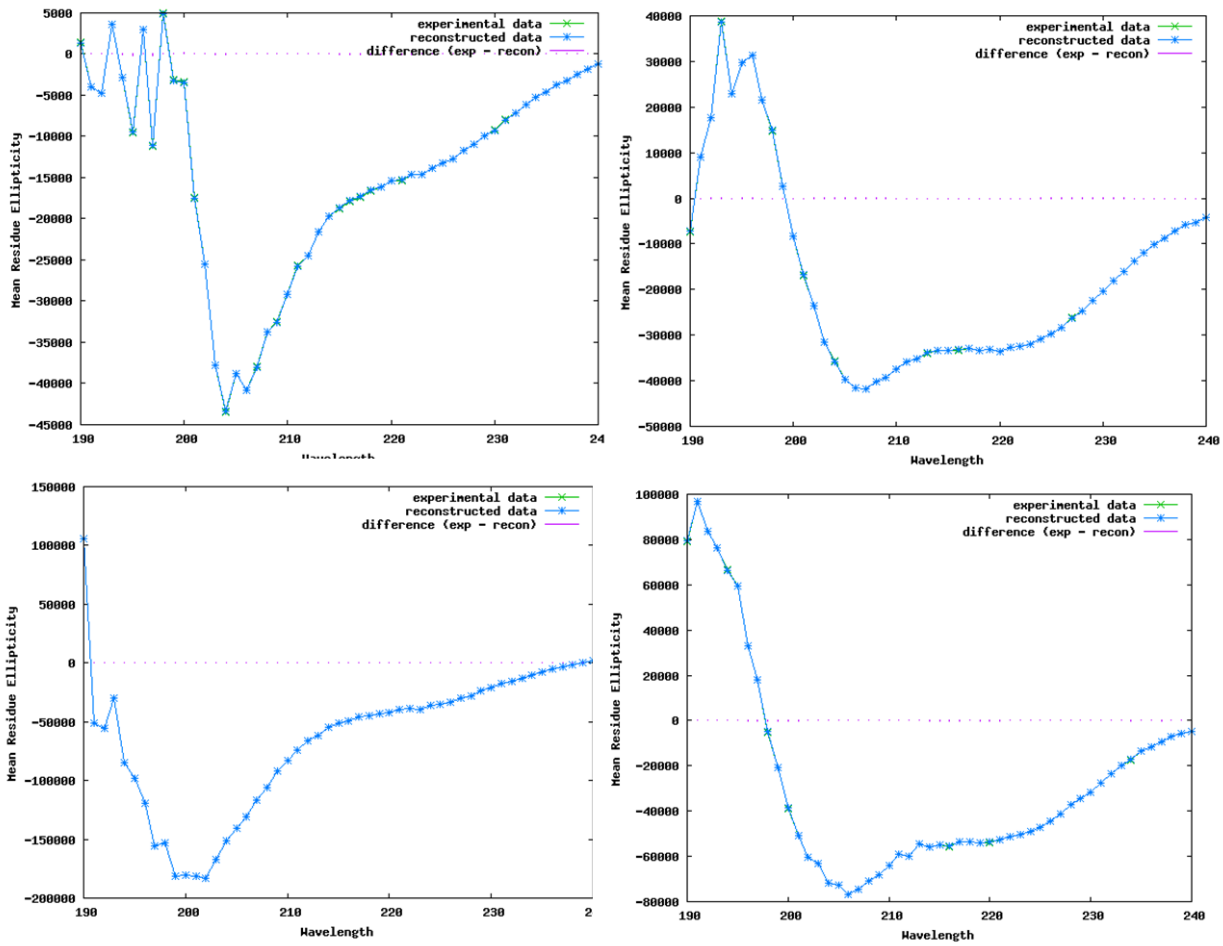


Figure 3.12: Graphical output of His6 Sp-B (top left), His6 Sp-B with 40% TFE (top right), Sp-B (bottom left) and Sp-B with 40% TFE (bottom right) generated in DichroWeb (NRMSD $\leq$ 0.003). Experimental (input) data are plotted in green, and reconstructed data determined from the calculated output secondary structure are plotted in blue. Differences in input and reconstructed data are shown in purple.

	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unfolded
His6 Sp-B	0.46	0.18	0.09	0.06	0.06	0.15
40% TFE	0.47	0.21	0.06	0.04	0.06	0.15
Sp-B	0.37	0.06	0.17	0.12	0.07	0.21
40% TFE	0.52	0.20	0.06	0.05	0.04	0.12

Table 3.7: The calculated secondary structure results obtained for the closest matching solutions for His6 Sp-B and Sp-B with and without the addition of 40% TFE.

TFE was added to samples of His6 Sp-B and Sp-B at concentrations varying from 0-50%, as shown in Figure 3.13 and 3.14. Two dips characteristic of an alpha helical spectrum was seen in all six recorded CD spectra of His6 Sp-B, with the first dip ranging from 202 nm to 207 nm, and the second dip observed at approximately 222 nm. Similar results were seen in Sp-B with the first dip ranging from 201 nm to 206 nm, and the second at approximately 220 nm. At higher concentrations of TFE in both protein samples, the characteristic alpha helical peak at approximately 190 nm was also observed.

Table 3.8 and 3.9 show the extended results generated from DichroWeb, using the CDSSTR analysis programme and set 4 reference proteins. The addition of TFE to both protein samples generally resulted in an increase in the total percentage of helix segments per 100 residues, but there was no significant change in the number of strand segments per 100 residues (Table 3.8).

The total helical content of His6 Sp-B and Sp-B without TFE added was 53% and 42% respectively. The addition of 50% TFE to the samples resulted in an increase in helical content to 71% in both His6 Sp-B and Sp-B. The total beta strand content was seen to decrease from 22% to 12% in His6 Sp-B, and 30% to 10% in Sp-B with the addition of 50% TFE. However, no significant change was observed in the percentage of turns and unfolded structures in both protein samples when increasing concentrations of TFE was added (Table 3.9).

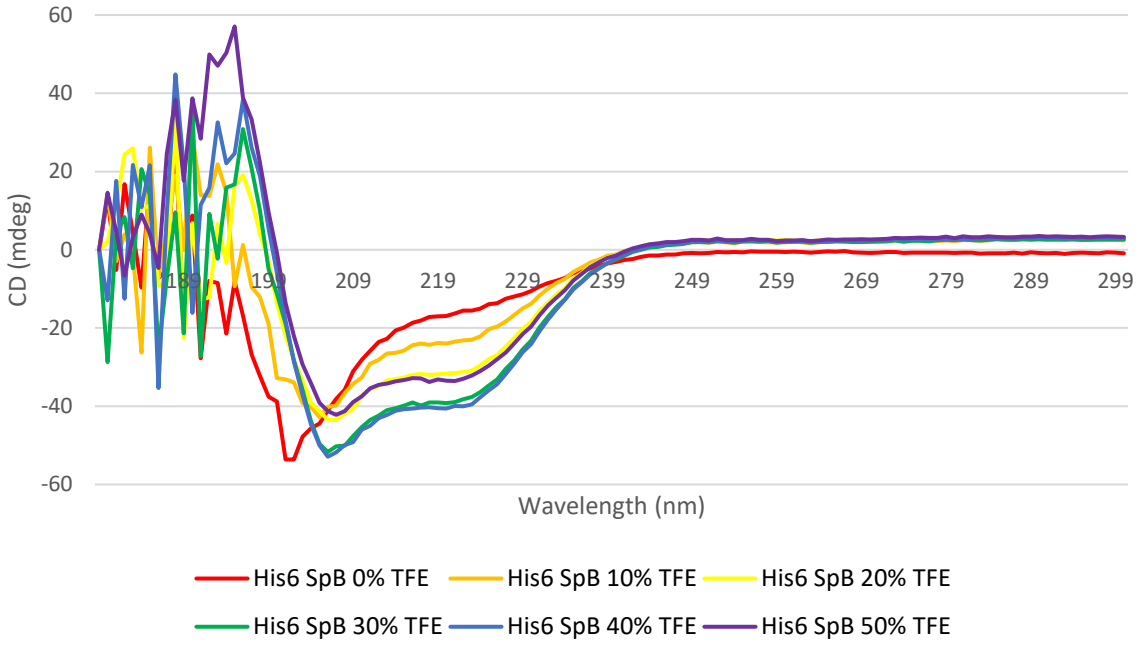


Figure 3.13: CD spectra of His6 Sp-B [0.1 mg mL^{-1}] purified in NaF-$NaPO_4$ buffer, with different concentrations of TFE added, ranging from 0-50%. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

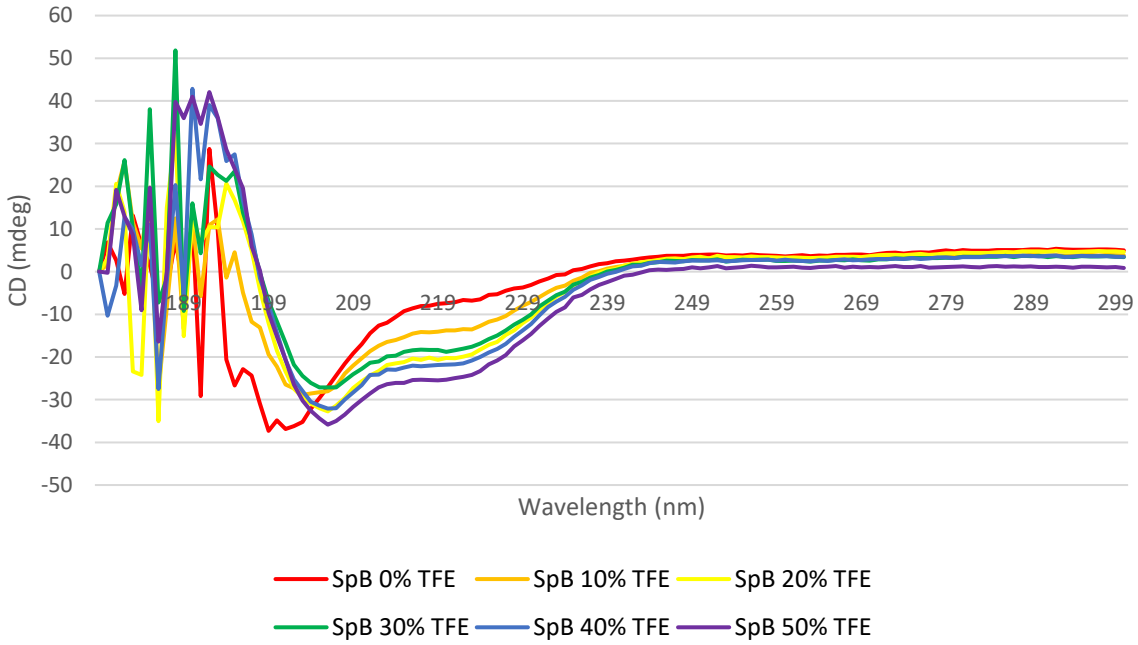


Figure 3.14: CD spectra of Sp-B [0.022 mg mL^{-1}] purified in NaF-$NaPO_4$ buffer, with different concentrations of TFE added, ranging from 0-50%. Measurements were made in a 1 mm pathlength cuvette, and recorded from 300 to 180 nm. Data obtained at wavelengths when the HT value was above 600 V should be treated with caution.

His6 Sp-B	Helix segments per 100 residues	Avg helix length per segment / Å	Strand segments per 100 residues	Avg strand length per segment / Å
0% TFE	2.726	19.347	4.438	4.834
10% TFE	3.319	18.088	4.162	3.780
20% TFE	5.053	14.728	2.506	2.773
30% TFE	2.887	18.798	3.919	4.966
40% TFE	3.160	19.578	2.951	4.833
50% TFE	4.612	15.582	2.801	4.266
Sp-B				
0% TFE	1.525	27.645	5.998	4.965
10% TFE	3.606	19.098	3.448	3.062
20% TFE	3.540	17.112	3.570	4.248
30% TFE	2.976	18.807	4.136	4.349
40% TFE	2.844	20.107	4.164	4.409
50% TFE	3.931	18.035	3.424	2.967

Table 3.8: DichroWeb results of His6 Sp-B and Sp-B at varying concentrations of TFE.

His6 Sp-B	Helix 1	Helix 2	Strand 1	Strand 2	Turns	Unfolded
0% TFE	0.42	0.11	0.13	0.09	0.06	0.20
10% TFE	0.47	0.13	0.07	0.08	0.07	0.17
20% TFE	0.54	0.20	0.02	0.05	0.07	0.10
30% TFE	0.43	0.12	0.12	0.08	0.06	0.20
40% TFE	0.49	0.13	0.08	0.06	0.07	0.17
50% TFE	0.53	0.18	0.06	0.06	0.04	0.12
Sp-B						
0% TFE	0.36	0.06	0.18	0.12	0.07	0.21
10% TFE	0.54	0.14	0.04	0.07	0.03	0.17
20% TFE	0.46	0.14	0.08	0.07	0.05	0.19
30% TFE	0.44	0.12	0.10	0.08	0.05	0.21
40% TFE	0.46	0.11	0.10	0.08	0.05	0.20
50% TFE	0.55	0.16	0.03	0.07	0.03	0.17

Table 3.9: The calculated percentages of secondary structure elements obtained for the closest matching solutions for His6 Sp-B and Sp-B with increasing concentrations of TFE added.

3.4 Discussion

Overloading the cuvette with high concentrations of protein may increase light scattering or excessive light absorption, which can interfere with the CD signal. The initial concentration used, 0.428 mg mL^{-1} was too strongly absorbing to provide useful data in the required wavelength range, as the HT exceeded 600 V below 222 nm. Diluting the concentration of His6 Sp-B from 0.428 to 0.103 mg mL^{-1} improved the situation, as the HT exceeded 600 V below wavelengths of 205 nm. However, Cl^- ions strongly absorb in the UV region and so buffers containing high concentrations of Cl^- ions are not suitable for the analysis of proteins by CD spectroscopy. For this reason, the original buffer was replaced with a buffer containing NaF, instead of NaCl.

The CD spectra of His6 Sp-B and Sp-B are obtained by subtracting the signal of the buffer from the total signal. Previous work had revealed that Sp-B was not stable in buffers with low salt concentrations, as it precipitates out. It was therefore necessary to buffer exchange the Sp-B, replacing the original Tris-based buffer containing Cl^- ions, which strongly absorb at the far UV wavelength range of 190 to 250 nm (Rodger, 2013), with one containing phosphate and NaF instead. 25 mM $NaPO_4$ and 300 mM of NaF allowed for the visualisation of the signal value down to approximately 200 nm. Further lowering the ionic strength may reduce the HT value threshold, where less noise at wavelengths below 200 nm would allow for the determination of the characteristic alpha helical peak at 190 nm in all samples.

Typically, alpha helices found in hydrophobic membrane proteins are longer than the helices found in soluble proteins. For instance, an alpha helix spanning a membrane thickness of 30-50 nm would require approximately 7.5 turns, where each helical turn contains 3.6 residues (Robinson, Afzal, and Leader, 2014), resulting in 27 residues in the helical structure. On the other hand, soluble proteins have shorter helices made up of 12 residues or less (Miles and Wallace, 2016).

As explained earlier, the addition of TFE mimics the environment of a lipid membrane. Results from the CD spectra for His6 Sp-B and Sp-B showed that the addition of increasing concentrations of TFE resulted in a more alpha helical structure. The addition of TFE increased the percentage of alpha helical structures and decreased the percentage of strands in both His6 Sp-B and Sp-B. This agrees with the mechanism by which TFE increases alpha helicity, where it promotes the formation of intramolecular hydrogen

bonds between the carbonyl oxygen and the amide hydrogen on neighbouring residues, which is essential for alpha helix formation in a protein. The presence of TFE enhances hydrophobicity of membrane proteins, by reducing protein interaction with water molecules. This increase in intramolecular hydrogen bonding results in a more compact, alpha-helical structure. Alternatively, high concentrations of TFE can cause protein denaturation, causing the alpha helical structure to convert to an anti-parallel beta sheet aggregate (Shiraki, Nishikawa and Goto, 1995). However, dissolving His6 Sp-B and Sp-B in increasing concentrations of TFE appeared to enhance the formation of alpha helical structures. Addition of 50% TFE to samples of His6 Sp-B and Sp-B resulted in a total percentage helicity of 71% respectively. This agrees with a study done by Cruz *et al.*, where the addition of 70% TFE to Sp-B isolated from porcine lungs caused Sp-B's alpha helicity to increase up to 60% (Cruz, Casals, and Perez-Gil, 1995).

The difference in physical environments can result in the solvent shift effect. Since membrane proteins are usually embedded in lipid membranes, the characteristics of an aqueous environment can influence the CD spectra. Solvents that do not have the optimal parameters for Sp-B to maintain its native structure may cause structural unfolding or misfolding, in turn affecting the CD spectral results.

DichroWeb is still the most comprehensive analysis tool for CD spectra and aims to provide a summary of the best deconvolution using algorithms and protein reference datasets. The NRMSD assesses the difference between the experimental input spectra with the reconstructed data. The NRMSD value of a good fit differs between the analysis programmes; for example, it would be 0.070 with SELCON3, 0.03 for CONTIN and 0.007 for CDSSTR. A low NRMSD score is a requirement but not a determining factor in concluding that the analysis has generated a reliable result. This means that a high NRMSD value shows that the calculated secondary structure of the experimental and reconstructed data is inaccurate, but using a more appropriate protein reference set may improve results of the data (Whitmore and Wallace, 2004). High NRMSD values (>0.1) may also be caused by errors in the concentration or cell pathlength measurements used.

3.5 Conclusions

CD spectroscopy is particularly useful in overcoming the challenges faced in determining the secondary structure of hydrophobic membrane proteins, which can aid developments in the full-length 3D structure of Sp-B. The data from this chapter confirmed that both His6 Sp-B and Sp-B had a predominantly alpha helical structure. This agrees with data in the PDB, where both the structures of Sp-B_{1-25} (1DFW) and Super Mini-B (1SSZ) have been solved in a membrane, revealing a predominantly helical structure (Gordon et al., 2000; Waring et al., 2005).

The CD experiments were successful in showing that both the His6 Sp-B and Sp-B samples were predominantly folded in aqueous solution, in the presence of 300 mM NaF, and that the helical content increased upon the addition of TFE to the samples, suggesting that in a lipid membrane Sp-B would have a more helical structure. Previously, there was no data that experimentally determined whether His6 Sp-B or Sp-B formed helical structures in a membrane environment. Since TFE aimed to mimic a lipid membrane environment, the experiments on His6 Sp-B and SP-B agreed with previously experimentally determined data of peptide mimics from the PDB, which were mapped in similar membrane conditions.

Although techniques like protein crystallisation may provide much more detailed information on the 3D structure of these proteins, the stability and solubility of Sp-B becomes problematic when such traditional crystallisation methods are used. The hydrophobic nature of Sp-B prevents it from freely existing in a solution, where the protein was seen to precipitate at higher pH and low salt concentrations, causing them to lose their 3D structure. It was also proposed that Sp-B molecules coexisted in different conformational states and interacts with lipids and detergents in solution. Hence, its structural heterogeneity likely hinders the ability of Sp-B to form a uniform crystal structure.

Chapter 4 – Molecular Dynamics Simulations of Sp-B_{1-25} in Lipid Monolayers at the Air-Water Interface with and without Oxidised Lipids

4.1 Introduction to Molecular Dynamics

Molecular dynamics (MD) is a computational tool used to simulate and predict the movement, interactions, and motions of molecular systems overtime. In this research, it generates information on the interactions of the peptide Sp-B_{1-25}, in various lipid monolayers. MD simulations can provide a more realistic view of these interactions compared to looking at static snapshots. MD simulations compute the physical, time-evolution trajectorial movement of a set of interacting atoms and molecules. These particles interact over a set timeframe, simulating the motion of the protein according to classical dynamics. Fundamentally, the trajectories of the particles are measured by the integration of Newton's equation of motion for all particles using the following equations:

$$\boldsymbol{F_i = m_i a_i} \qquad (4.1)$$

where F_i refers to the exerted force on a particle i, m_i refers to the mass of particle i and a_i refers to the acceleration of particle i. Alternatively, the force F_i can be expressed as the gradient (∇_i) or rate of change of the potential energy (V) with respect to their position, r_i, as shown in Equation (4.2):

$$\boldsymbol{F_i = \nabla_i V = -\frac{dV}{dr_i}} \qquad (4.2)$$

The acceleration a_i can be expressed as Equation (4.3):

$$\boldsymbol{a_i = \frac{d^2 r_i}{dt^2}} \qquad (4.3)$$

The derivative of the potential energy (V) can be related to the change in position as a function of time by combining Equations (4.2) and (4.3) to form Equation (4.4). Integrating these equations of motion results in a trajectory describing the accelerations, positions, and velocities of the particle in discrete timesteps, allowing the simulation and prediction of the system's state at any time, up to the nanosecond timescale.

$$\boldsymbol{-\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2}} \qquad (4.4)$$

4.1.1 GROMACS

The GROningen MAchine for Chemical Simulations (GROMACS) (Abraham *et al.*, 2015) is an MD software package primarily used to simulate biomolecules such as proteins, lipids, and nucleic acids. It allows for the simulation of the motion and many interactions of molecules down to the nanosecond scale. The intensity of computation and length of simulation time has a direct impact on the quality of the simulation.

The sampling time of a simulation considers two factors – longer simulations result in more sampling but is less economical compared to shorter simulations both of which are constrained by the timescale of the dynamics (Hernández, Anderson and Lemaster, 2012; Cunha *et al.*, 2015). Longer simulations can result in the generation of numerical errors (round-off, truncation errors) due to numerical integration (Brooks, 1989; Arizumi and Bond, 2012), which can be alleviated by computational methods such as specialised integration algorithms (Skeel, 1999; Tuckerman and Martyna, 2000).

Like any experimental technique, MD simulations are prone to limitations, with error sources determining the extent of correct sampling from a thermodynamic ensemble in the simulation (Romo and Grossfield, 2014). The equation used to derive the interacting forces on all atoms in the system determines the accuracy of the parameterised forcefield used, which influences the degree of error in the simulation (Bradley and Radhakrishnan, 2013). Errors can also result from heterogenous distributions of conformational states sampled by the simulation, and how much conformational sampling was obtained from it (Beck, Armen, and Daggett, 2005).

4.1.2 Forcefields

The forces between the atoms of the simulated molecules must be estimated using forcefields, which is a computational tool integral to the function of MD simulations. It has a set of parameters which calculates the potential energy of systems and can be used in both atomistic and coarse-grained (CG) MD simulations which are described later. Atomistic MD simulations allow for more detailed studies of biological systems by accounting for all atoms and forces present, including hydrogen. CG simulations on the other hand, are less chemically detailed and are used in longer simulations of

macromolecules to maintain computational efficiency, by using pseudo-atoms to represent atom groups.

In all forcefields, the functional form of potential energy calculated in MD simulations comprises bonded and non-bonded interactions, shown in Equation 4.5. Bonded interactions include atomic interactions accounting for covalent bonds, bond stretching, dihedral rotation and valence angles. Non-bonded interactions account for long range non-covalent bonds, electrostatic and Van der Waals forces. The total energy will therefore be comprised of bonded, non-bonded and other interactions which includes terms specific to the forcefield being used.

$$\boldsymbol{E_{total}} = \boldsymbol{E_{bonded}} + \boldsymbol{E_{nonbonded}} + \boldsymbol{E_{other}} \quad (4.5)$$

Out of the bonded interactions, improper dihedrals may maintain the chirality or planarity of aromatic rings or at an atom centre. The inclusion of such parameters is dependent on the forcefield; in this research, the atomistic Chemistry at Harvard Macromolecular Mechanics (CHARMM) 36m forcefield was employed. The CHARMM forcefields include an additional term for improper dihedral energy with a quadratic dependence on its value, accounting for out-of-plane bending in any set of four non-successively bonded atoms. Additionally, the CHARMM forcefield has an Urey-Bradly angle term, treating two terminal atoms in an angle with a quadratic term that is dependent on the distance between the atoms (Guvench and MacKerell, 2008). These differences allow the CHARMM forcefield to have more degrees of freedom to parameterise vibrational spectra.

With regards to the non-bonded, Van der Waals interactions, the Lennard-Jones (LJ) potential fundamentally states that atoms or molecules will not interact at infinite distances, but they will attract each other at a moderate distance and repel each other at close distances. The CHARMM forcefield uses the geometric mean to calculate the LJ potentials for the depth of potential energy ε_{ab}, where *a* and *b* are the atoms associated with these parameters. This allows for differentiation of a hydrogen atom bonded to oxygen, compared to a hydrogen atom bonded to an aromatic carbon atom. CHARMM uses the arithmetic mean to calculate the atomic positions (R) of atom *a* and *b*, in the equation:

$$\frac{\boldsymbol{1}}{\boldsymbol{2}}(\boldsymbol{R_{min,a}} + \boldsymbol{R_{min,b}}) \quad (4.6)$$

The CHARMM forcefield models saturated and unsaturated phospholipids (Schlenkrich *et al.*, 1996; Feller *et al.*, 1997), with recent improvements made to the improper dihedral energy and LJ parameters which includes polyunsaturated chains to accurately model a diverse range of lipids (Feller and MacKerell, 2000; Feller, Gawrisch and MacKerell, 2002; Jen Chen, Yin and MacKerell, 2002), making it especially suitable for the purpose of simulating Sp-$B_{1\text{-}25}$ in various lipid monolayer systems.

4.1.3 Periodic boundary conditions

Another factor to consider during MD simulations would be to preserve thermodynamic properties such as density, pressure and temperature while avoiding any effects that may hinder how the proteins and lipids may be simulated in a realistic environment. For example, simulating the surfactant lipids in a small volume of solvent (the simulation box), may render them susceptible to surface effects, such as evaporation, since a disproportionate number of molecules would interact with the edge of the simulation box. Periodic boundary conditions (PBCs) allow for the lipids, assumed to be in a periodic box (unit cell), to be simulated in an infinite and homogenous system by taking that unit cell and repeating it in 3D space. If the trajectory of a molecule diffuses outside of the boundaries of the unit cell, it would simultaneously reappear and enter on the opposite side via lattice symmetry to minimise surface effects in GROMACS.

4.1.4 Ensembles

In the real world, the laws of thermodynamics are governed through a multitude of processes, through the transfer of energy between matter. In MD simulations, this is simplified by separating the system that Sp-B and surfactant lipids are simulated in, from the surroundings. A statistical ensemble allows for molecules to be simulated in a controlled environment, where all the possible states of a real system are represented.

There are two main types of ensembles used in MD simulations The isothermal-isobaric (NPT) ensemble outlines a system in contact with a thermostat at temperature T and a barostat at pressure P. In this ensemble, three variables are kept constant (amount of substance N, pressure P and temperature T) in an MD simulation to maintain conditions mimicking laboratory conditions, open to ambient temperature and pressure. In a

canonical (NVT) ensemble, three variables in the system (number of particles N, volume V and temperature T) are kept constant. Endothermic and exothermic processes exchanged using various thermostat algorithms can add or remove energy at the boundaries of an MD simulation.

In these simulations, a variety of thermostats were used to control the temperature of the systems. The Nosé-Hoover thermostat (Nosé and Nose, 1984; Hoover, 1985) achieves realistic constant temperature conditions in canonical MD simulations, by considering the heat bath as an integral part of the system. The Berendsen thermostat is an alternative algorithm that can be used to control temperature, by rescaling the particle velocities in MD simulations. It suppresses fluctuations in kinetic energy and volume, while quickly relaxing a system to the target temperature. However, regulating the temperature using the Nosé-Hoover thermostat becomes more advantageous once the system has reached an equilibrium, since it becomes more crucial to generate an accurate canonical ensemble.

4.1.5 Coarse-grained vs. atomistic MD simulations

Atomistic MD simulations allow for the detailed study of biological systems using statistical mechanics which predict atomic trajectories, by accounting for all forces present. Atomistic simulations provide high-resolution details of individual atomic structures and molecular interactions. CG simulations on the other hand, represent a more simplified computational model used to study complex biological systems. This system is based on the usage of pseudo-atoms that represent atom groups; the folding and structure of proteins can be predicted by using a few pseudo-atoms per amino acid (Kmiecik *et al.*, 2016). This also provides an alternative for maintaining computational efficiency in larger-scale simulations that require numerous timesteps due to long timescales, that may be costly to run using atomistic simulations (Tuffery and Derreumaux, 2012). However, CG simulations face issues with inaccurate models, non-transferability, and neglect of non-native interactions (Takada, 2012).

Atomistic simulations are preferable in the study of hydrogen bonding between atoms in a simulation since they accurately capture their positioning and interactions. For example, a study by Ollila *et al.* concluded that atomistic simulations of a lipid bilayer had

a detailed representation of the order parameters of acyl chain structures and dynamics. The simulations gave a realistic view of the saturated and unsaturated acyl chains in PC lipid bilayers at 300 K, where it was observed in both experimental data and simulations that the order parameter of the lipid chains increased with increasing cholesterol concentration (Ollila and Pabst, 2016).

Additionally, understanding the geometry of hydrogen bonds requires a higher accuracy of the bond lengths, bond angles, and dihedral angles. For instance, atomistic simulations can differentiate between types of hydrogen bonds, such as donor-acceptor pairs involving nitrogen or oxygen, which may reveal bonding patterns in complex molecular systems. For instance, hydrogen bonding between water and PC lipids were analysed using atomistic MD simulations by Pasenkiewicz-Gierula *et al.*, where it was discovered that 70% of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) molecules are linked by water molecules to form clusters of 2-7 DMPC molecules. It was also found that approximately 70% of intermolecular water bridges are formed between non-ester phosphate oxygens. The remaining H-bonds were formed between non-ester phosphate and carbonyl oxygens, showing the average geometry of H-bonding to be planar trigonal. When oxygen atoms form three H-bonds with non-ester phosphates, there is a steric tetragonal geometry (Pasenkiewicz-Gierula et al., 1997). This may not have been achievable with the use of CG MD simulations.

In this research, atomistic simulations were performed on lipid monolayers containing Sp-$B_{1\text{-}25}$. This provides a detailed view of each atom in the molecular system, to study lipid-lipid and lipid-peptide interactions. Atomistic simulations were favoured over CG simulations, as the simulated monolayer system and peptide was small, eliminating the downside of high computational cost.

4.1.6 Molecular dynamics simulations of lipid monolayers at the air-water interface

To study the behaviour at the air-water interface of the alveoli, the lung surfactant monolayer must be simulated. MD simulations of two CG models of DPPC monolayer systems performed by Baoukina *et al.* were done with the GROMACS v.3.3.1 simulation

package (Lindahl, Hess and van der Spoel, 2001) using the MARTINI forcefield. The simulation adopted a canonical (NVT) ensemble, where temperature was maintained using a Berendsen heat bath to determine the pressure-area isotherm of a DPPC-based monolayer. A small system (for a coarse grain study) with 128 DPPC and 1746 water molecules, and a large system with 8192 DPPC and 111,744 water molecules were simulated with box sizes of 6.8 × 6.8 × 20 nm and 54 × 54 × 75 nm respectively. The initial structure of the small system was compressed using positive pressure or expanded using negative pressure using pressures of -60 bars up to +60 bars for 100 ns. The initial structure of the large system was compressed using lateral pressure of 5 bars for 40 ns, obtaining a set of isotherms from different timeframes during the compression run. The small system was subsequently simulated for 1 μs using an equilibration run of 200 ns and a production run of 800 ns, while the large system was simulated for 400 ns using an equilibration run of 200 ns and a production run of 200 ns. To ensure that the box size remained constant normal to the membrane upon compression and expansion of the monolayer, the systems were simulated using the Berendsen coupling scheme, using a coupling constant of 4 ps with a lateral pressure of 5×10^{-5} bars. To analyse the structural and dynamic properties of lipids, surface tension was calculated at the vacuum-water interface using 33,248 water molecules in a 20 × 20 × 20 nm water cube inside a 20 × 20 × 30 nm simulation box, and at the lipid-vacuum interface, two layers containing 128 lipids respectively were bound by vacuum slabs on each side. At the lipid-water interface, 1593 water molecules were added to the lipids.

In this study, the CG model did not significantly affect the properties of the compressed monolayer. The tension-area isotherm could be compared to experimental pressure-area isotherms using corrections to the surface tension at the vacuum-water interface. The surface tension calculated at the lipid chain-water and lipid chain-vacuum interfaces were in good agreement with the experimental values for bulk alkanes, however, the surface tension at the vacuum-water interface is much lower than the experimental value of the surface tension at the air-water interface. Overall, the CG simulations of the DPPC systems produced a realistic model, which provided a characterisation of the geometry and structure of the monolayer which is in good agreement with the experimental data (Baoukina et al., 2007).

A study by Khatami *et al.* ran atomistic MD simulations on Sp-B, using the optimised potentials for liquid simulations all atom forcefield (OPLS-AA) forcefield to understand how its structure and mechanism affected lipid reorganisation. This was to obtain detailed insight into the structural properties of Sp-B which allowed it to interact with lipid bilayers, which would otherwise be difficult with experimental studies alone. Sp-B was generated in the open and bent (V-shaped) conformations and placed at three different positions in the lipid bilayer – inside, half inside and outside of the bilayer, giving a total of six initial system configurations (Khatami, Saika-Voivod and Booth, 2016). The simulations resulted in a total of three main Sp-B/lipid topologies. One consisted of Sp-B's helices inside the bilayer plane, while another had its helices inclined with respect to the lipid bilayer, and a closed Sp-B structure with a peripheral interaction with the bilayer. Simulations of Sp-B showed a 35% alpha helicity in its structure, like experimental data (Olmeda, García-Álvarez and Pérez-Gil, 2013). The simulations also showed the rapid collapse of Sp-B into a compact, closed structure which peripherally interacts with the lipid bilayer when it is positioned outside of the bilayer; this mechanism of Sp-B binding adjacent lipid layers was also shown in previous studies (Serrano and Pérez-Gil, 2006). 2–2.5 μs all-atom simulations of two energetically realistic conformations of Sp-B was expressed by placing Sp-B partially or completely inside the bilayer. An open Sp-B structure with all helices parallel to the bilayer and the backbone located deeply in one leaflet of the bilayer corresponds to the FTIR experimental results indicating that the protein can exist in a parallel orientation (Vandenbussche *et al.*, 1992). The other conformation consists of some or all helices taking on inclined orientations not in the plane of the bilayer. These orientations associate with bilayer disruption which stabilises pore-like defects which are caused by the presence of water molecules and lipid headgroups in the membrane interior. This correlates with the experimental observation of lipid permeability in the presence of Sp-B (Parra *et al.*, 2013).

4.1.7 MD simulation set-up

If the non-stationary trajectories are included in the calculations of equilibrium, it will lead to experimental errors. Hence, the simulated system must first undergo energy minimisation to remove any excess forces and likelihoods of fluctuations in energy during

the simulation. This was done using the steepest descent algorithm, starting from the coordinates in the opposite direction to the starting point (negative gradient). The step size is adjusted in an iterative process, which repeats until a sufficient energy minimum is reached in all directions for each atom.

Subsequently, the system was coupled to the relevant thermostats; in this study, equilibration was done in the NPT ensemble and coupled to the Berendsen thermostat with added position restraints, to stabilise and maintain constant pressure and temperature in the system. The position restraints are released before the final production run, using the NVT ensemble coupled to the Nosé-Hoover thermostat, maintaining a constant volume and temperature. The NVT assembly was chosen so that simulations could be performed at different constant, areas per lipid.

4.1.8 Simulating Sp-$B_{1\text{-}25}$ in a double monolayer system

MD simulations provide insight into the biophysical function of surfactant lipids and Sp-$B_{1\text{-}25}$, where changes in the environment, such as with pollutants, can alter the structure and phase behaviour of lipids in lung surfactant. Previous experiments studying the activities of lung surfactant are commonly simulated using a symmetric model, which is to simulate a slab of water with two monolayers on opposite sides of the water box, and a vacuum above and below the monolayers (Stachowicz-Kuśnierz *et al.*, 2022).

In Figure 4.1, Sp-$B_{1\text{-}25}$ was simulated in a double monolayer system with the polar lipid headgroups facing the inner region towards the water molecules, and the acyl chain tails faces the vacuum. The presence of water molecules sandwiched between the two monolayers keeps the lipids and peptides correctly in place, as previous studies revealed that a single monolayer system would result in water molecules seeping into the lipid layer containing the peptide and outwards into the vacuum layer above.

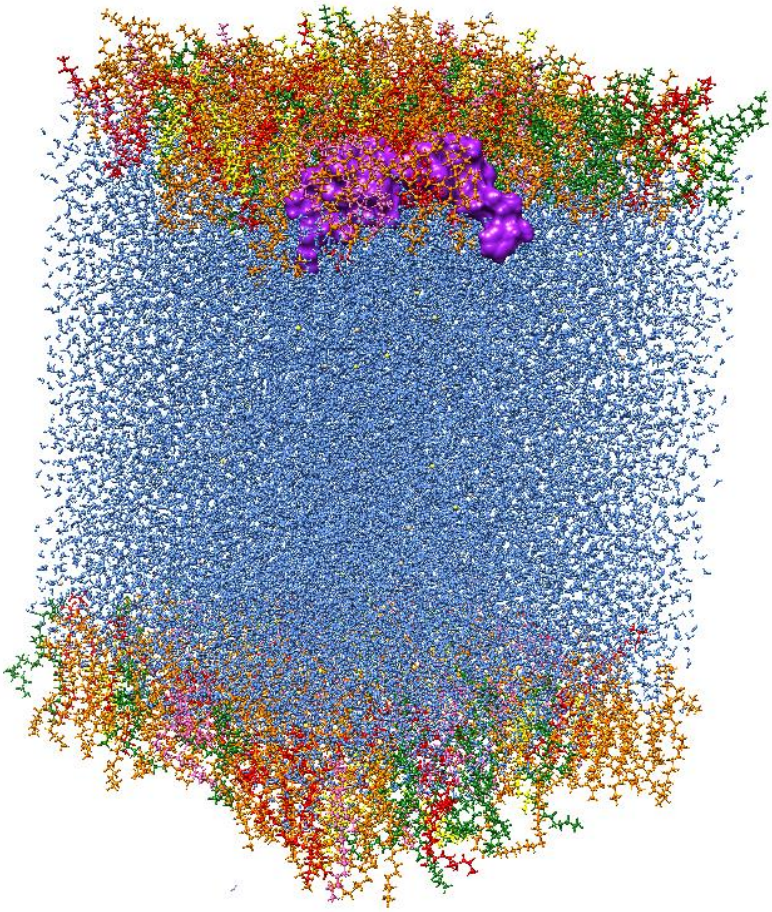


Figure 4.1: A double monolayer system showing Sp-$B_{1\text{-}25}$ (purple) in the upper leaflet, composed of DPPC (orange), POPG (green), POPC (pink), PLPC (red), cholesterol (yellow). Water molecules were simulated as TIP3P (blue), containing 0.15 M Na^+ and Cl^- respectively. There are a total of 138 lipids in the upper leaflet containing Sp-$B_{1\text{-}25}$, and 128 lipids in the lower leaflet.

4.2 Methods

The MD simulations reported in this thesis made use of time on HPC granted via the UK High-End Computing Consortium for Biomolecular Simulation, HECBioSim (https://www.hecbiosim.ac.uk/), supported by EPSRC grant no. EP/R029407/1 and no. EP/X035603/1. In this study, the open-source software package GROMACS 2021.1 (Lindahl *et al.*, 2021) was used for all simulations. The simulations were run on an SSH server for the Tier 2 supercomputer Bede, hosted by the Northern Intensive Computing Environment at Durham University. The periodic boundary conditions were visualised in the Visual Molecular Dynamics (VMD) software (Humphrey, Dalke and Schulten, 1996), and all other simulation parameters were visualised in Chimera (Pettersen *et al.*, 2004).

4.2.1 Structure of Sp-$B_{1\text{-}25}$

For all simulations, the structure of Sp-$B_{1\text{-}25}$ (1DFW) was taken from the Research Collaboratory for Structural Bioinformatics (RCSB) Protein Data Bank (PDB). Full length human Sp-B was not simulated in the monolayers because an experimental structure of the peptide has yet to be determined. From the available entries in the PDB, the N-terminal segment of Sp-B, i.e., Sp-$B_{1\text{-}25}$, was chosen as the starting structure for MD simulations. Furthermore, it was shown in the literature that Sp-$B_{1\text{-}25}$ maintains and possesses the important biophysical properties and functions of the full-length protein (Gordon et al., 2000).

Human Sp-B contained a total of 2 negatively charged residues (1 Asp, 1 Glu), highlighted in green, and 9 positively charged residues (7 Arg, 2 Lys), highlighted in red in the figure below. Sp-$B_{1\text{-}25}$ was formed by attaching an amine (NH_2) group to the N-terminal end of the peptide. In Sp-$B_{1\text{-}25}$, there are no negatively charged residues and 4 positively charged residues (2 Arg, 2 Lys) (Figure 4.2).

MAESHLLQWLLLLLPTLCGPGTAA *WTTSSLACAQGPEFWCQSLEQALQCRALGHCLQEVWGHVGAD*
DLCQECEDIVHILNKMAKEAIFQDTMRKFLEQECNVLPLKLLMPQCNQVLDDYFPLVIDYFQNQTDSN
*GICMHLGLCKSRQPEPEQEPGMSDPLPKPLRDPLPDPLLDKLVLPVLPGALQARPGPHTQDLSEQQ*FPI
PLPYCWLCRALIKRIQAMIPKGALAVAVAQVCRVVPLVAGGICQCLAERYSVILLDTLLGRMLPQLVCRL
VLRCSM*DDSAGPRSPTGEWLPRDSECHLCMSVTTQAGNSSEQAIPQAMLQACVGSWLDREKCKQFV*
EQHTPQLLTLVPRGWDAHTTCQALGVCGTMSSPLQCIHSPDL

FPIPLPYCWLCRALIKRIQAMIPKGALAVAVAQVCRVVPLVAGGICQCLAERYSVILLDTLLGRMLPQLVC
RLVLRCSM (Full-length human Sp-B)

NH_2-FPIPLPYCWLCRALIKRIQAMIPKG (Sp-B_{1-25})

Figure 4.2: Top: Amino acid sequence of full length Sp-B. Residues 1-24 represent the signal peptide (purple), residues 25-200: Sp-B propeptide 1 (orange), residues 201-279: mature Sp-B peptide chain (blue), residues 280-381: Sp-B propeptide 2 (orange). **Middle:** Amino acid sequence of the mature Sp-B peptide chain. Negatively charged residues are highlighted in green, while positively charged residues are highlighted in red. **Bottom:** Amino acid sequence of Sp-B_{1-25}. Positively charged residues are highlighted in red.

4.2.2 Structure of the surfactant lipids

The composition of lung surfactant revealed DPPC as the most abundant lipid present, followed by unsaturated PC lipids, PG lipids, and cholesterol (CHL) (Han and Mallampalli, 2015). For the purpose of this study, only the four most abundant surfactant lipids and CHL were used in the simulations with Sp-B_{1-25} (Figure 4.3).

To simulate oxidative damage in some lipid monolayers, the carbon double bond between Carbon 8 and 9 on the SN_2 chain of POPG was broken and the lipid tail terminated with an aldehydic group on Carbon 8, to give palmitoyl-2-(9'-oxo-nonanoyl)-sn-glycero-3-phospho-(1'-rac-glycerol), a known oxidation product of POPG (National Center for Biotechnology Information, 2024)., oxidised and replaced with an aldehydic group on Carbon 8. This was done manually in the University of California San Francisco (UCSF) Chimera software, with the oxidised POPG lipid renamed as PXPG (Figure 4.4). The monolayers containing the oxidised lipids were then parameterised using SwissParam (Zoete et al., 2011).

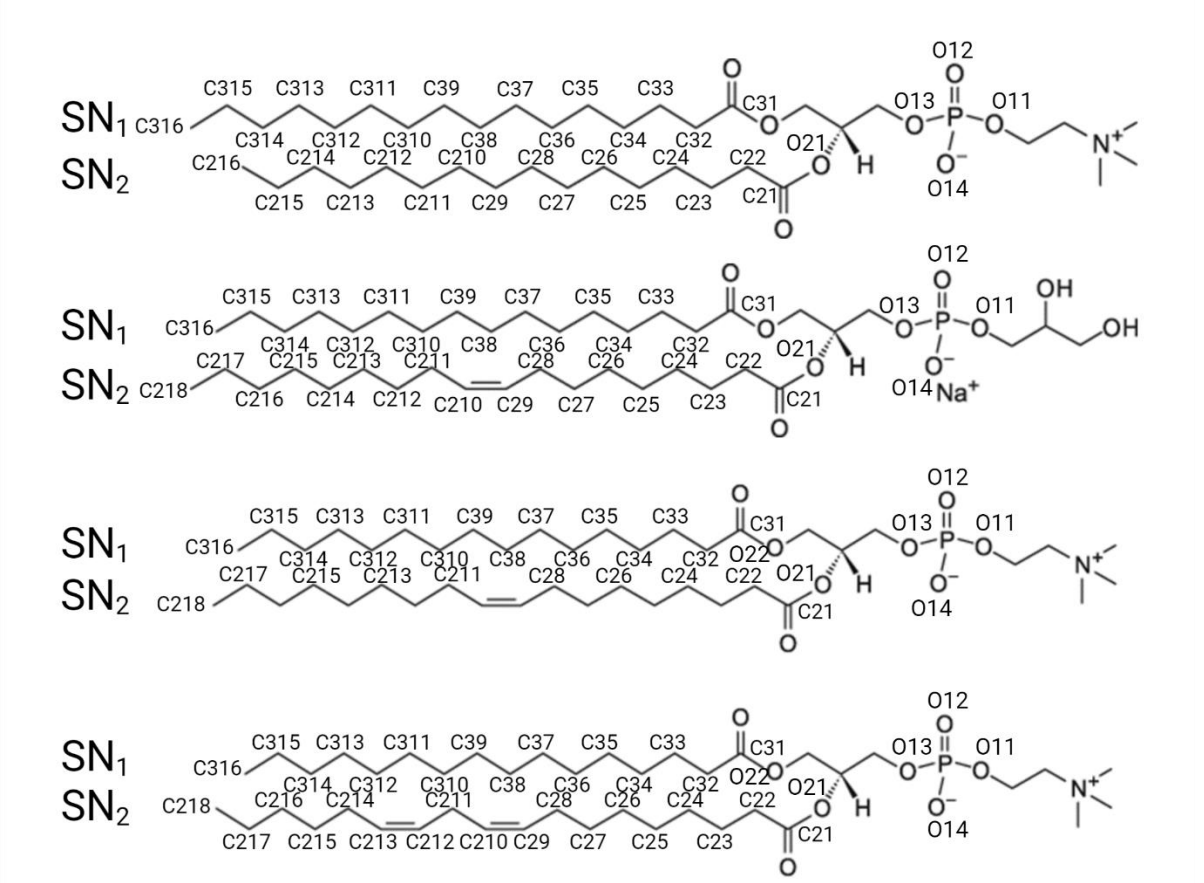


Figure 4.3, top to bottom: Structure of 1, 2-dipalmitoyl-sn-glycero-3-phosphocholine **(DPPC)**, 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt) **(POPG),** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine **(POPC),** 1 palmitoyl-2-linoleoyl-sn-glycero-3-phosphocholine **(PLPC)**. The carbon atoms are labelled according to the order parameters index groups, where the SN_1 and SN_2 chains are as shown. The indexed oxygen atoms for the lipid headgroups are shown.

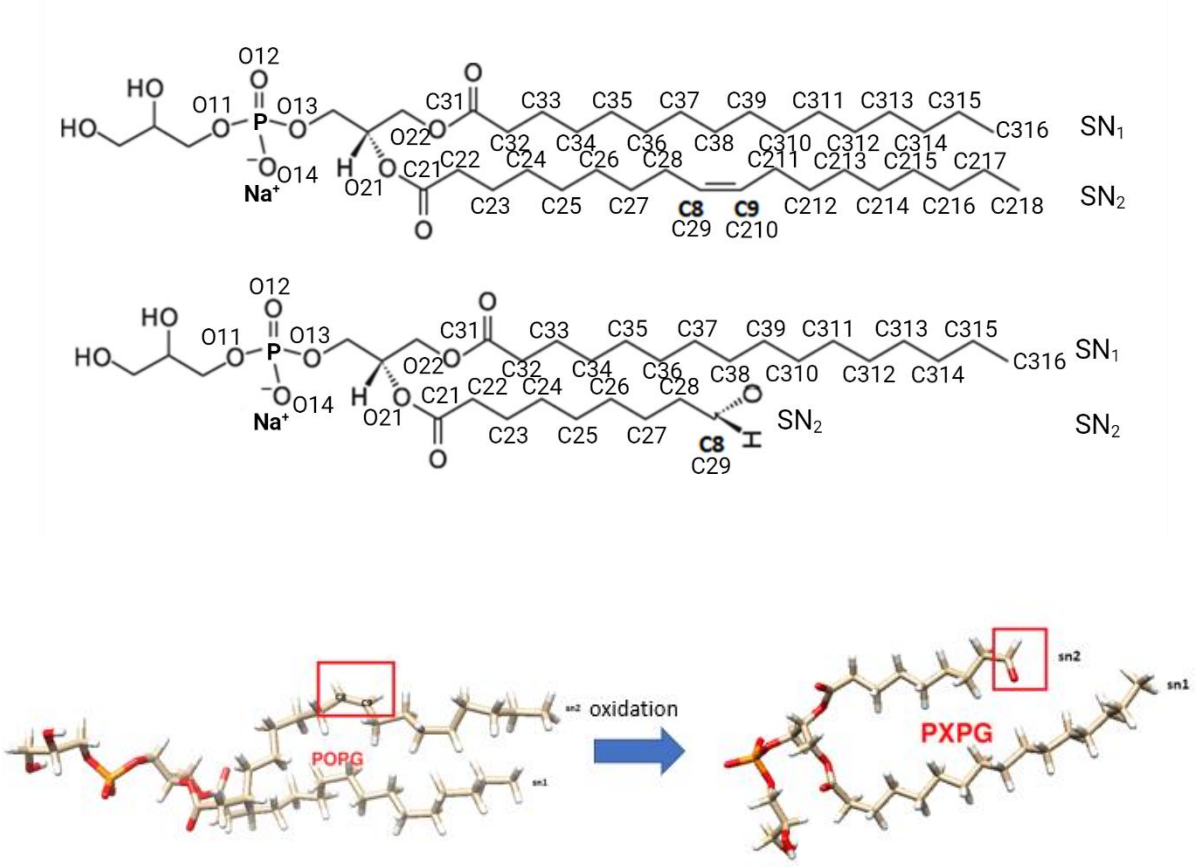


Figure 4.4: Oxidation on Carbons 8 and 9 of the SN_2 chain of POPG to produce an aldehydic form of the lipid known as PXPG.

4.2.3 Constructing the monolayers for GROMACS

The CHARMM-GUI Monolayer Builder was used to create the double monolayer systems. This study aimed to construct five monolayers, one containing only DPPC, another containing DPPC and POPG or PXPG in an 80:20 mol% ratio, and monolayers DPPC, POPG or PXPG, POPC, PLPC and CHL. The number of each lipid type has been represented in Table 4.1.

All five monolayers utilised Sp-$B_{1\text{-}25}$ (1DFW) in the Protein/Monolayer system. They were constructed in the CHARMM-GUI Monolayer Builder using the CHARMM36m forcefield. Sp-$B_{1\text{-}25}$ was aligned in the upper monolayer along the first principal axis along Z. The peptide was rotated at an angle of 60° with respect to the X axis, 90° with respect to the Y axis and was translated along the Z axis at -12 A. Lipid-like pseudo atoms were used in the replacement method to build the appropriate lipids around Sp-$B_{1\text{-}25}$. To neutralise the system, the Monolayer Builder included 0.15 M of Na^+ and Cl^- ions. The

input structures for all monolayers were exported and the periodic boundary conditions were checked in Chimera and VMD before starting the MD simulation steps.

	Upper leaflet	**Lower leaflet**
Sp-B_{1-25}	1	0
100% DPPC	137 DPPC	128 DPPC
80% DPPC 20% POPG	110 DPPC 27 POPG	102 DPPC 26 POPG
80% DPPC 20% PXPG	110 DPPC 27 PXPG	102 DPPC 26 PXPG
50% DPPC **10% POPG** **16%POPC** **14% PLPC** **10% Cholesterol**	69 DPPC 14 POPG 22 POPC 19 PLPC 14 CHOL	64 DPPC 13 POPG 20 POPC 18 PLPC 13 CHOL
50% DPPC **10% PXPG** **16%POPC** **14% PLPC** **10% Cholesterol**	69 DPPC 14 POPG 22 POPC 19 PLPC 14 CHOL	64 DPPC 13 POPG 20 POPC 18 PLPC 13 CHOL

Table 4.1: Number of peptides and lipids simulated in each leaflet. The lower leaflet contains lipids only, whilst the upper leaflet has the peptide plus lipids.

4.2.4 Initial simulation set-up

All simulations were done on Bede using the GROMACS 2021.1 module. The five monolayers generated from CHARMM-GUI were energy minimised using a maximum of 5000 steps of steepest descent, with position restraints for the lipids (DPOSRES_FC_LIPID = 1000.0), backbone (DPOSRES_FC_BB = 4000.0), side chains (DPOSRES_FC_SC = 2000.0) and dihedral angle restraints for the peptide (DDIHRES_FC = 1000.0). The gmx energy command was then used to generate the graph to check the structure and parameters of the monolayer (Figure 4.5).

In the first equilibration phase, the system was equilibrated in the NPT ensemble using the Berendsen thermostat, for 500 ps with a temperature time constant of 5 ps and semi-isotropic pressure coupling to achieve a pressure of 1.0 bar. There were position restraints on the backbone (DPOSRES_FC_BB = 50.0) and hydrogen bond constraints, achieving a constant temperature of 310 K. If the peptide becomes uncentred after running simulations on the monolayer, the command gmx trjconv was used to re-centre the peptide (Figure 4.6).

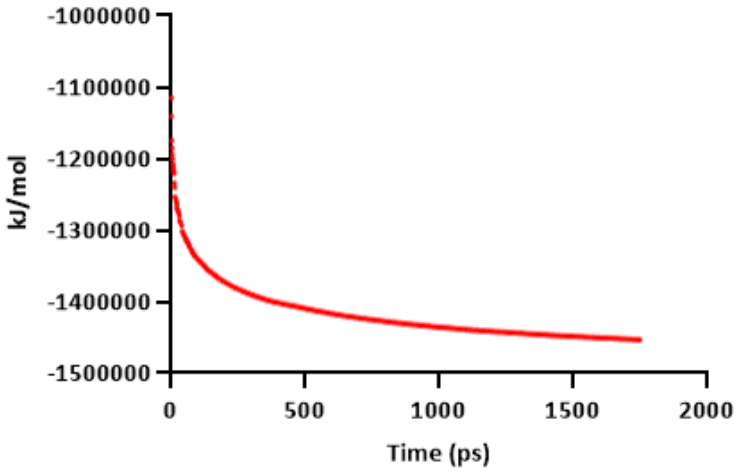


Figure 4.5: Process of energy minimisation representing the decrease in potential energy (kJ/mol) against time (ps) using GROMACS 2021.1.

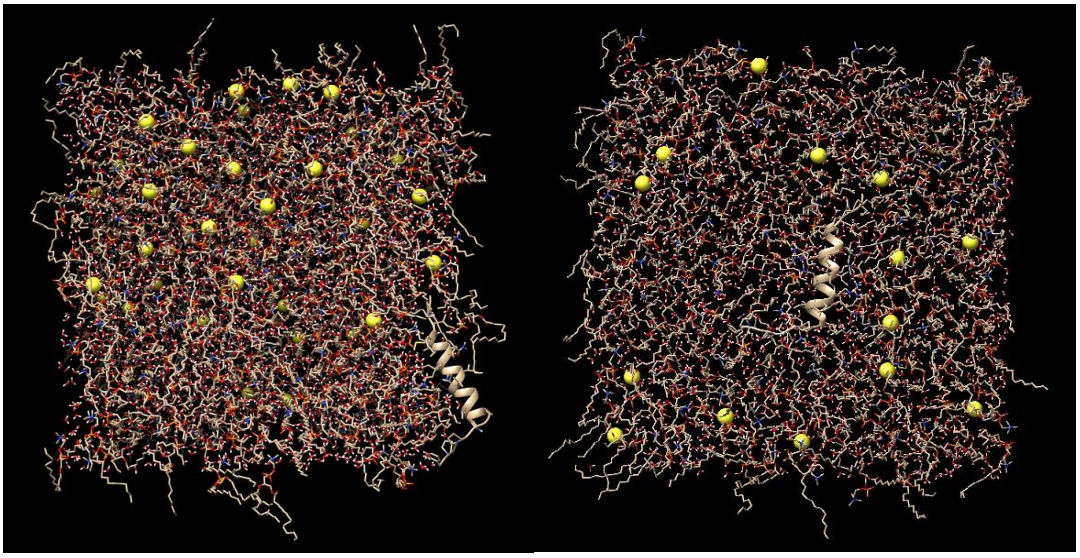


Figure 4.6: Sp-$B_{1\text{-}25}$ (beige) in a DPPC/POPG (80:20 mol%) monolayer before (left) and after (right) centring. Sodium ions are depicted in yellow.

4.2.5 Adjusting the box size of the system

In the second equilibration phase, the box size and APL were adjusted for each system to simulate the surface area of lung surfactant expanding and compressing during inhalation and exhalation. For all systems, the monolayers were equilibrated to approximately 47, 55 and 72 $Å^2$. This was done using the same ensemble, thermostat, and temperature time constant as the first equilibration, and hydrogen bond constraints were used.

In the DPPC monolayer, the system was equilibrated for 8 ns using semi-isotropic pressure coupling, at a pressure of 1 bar (ref_p) to achieve an APL of 47.3 $Å^2$ (Table 4.2). The compressibility was set at 4.5×10^{-5} bar^{-1} in both the xy and z directions. The system was equilibrated for 12 ns using the same parameters, at a pressure of -10 bar (ref_p), then for 6 ns at -25 bar to expand the monolayer to an APL of 55.4 $Å^2$ (Table 4.3). Lastly, the system was equilibrated for 26 ns at -25 bar (ref_p), then for 2 ns at -35 bar, and 2 ns at -25 bar to expand the monolayer to an APL of 72 $Å^2$ (Table 4.4).

In the DPPC:POPG (80:20 mol%) monolayer, the system was equilibrated for 12 ns at a pressure of 1 bar (ref_p) in the xy direction and 0 bar in the z direction, to shrink the APL to 55.4 $Å^2$ (Table 4.5). The monolayer was equilibrated for another 22 ns to shrink the APL to 51.8 $Å^2$, before adjusting the pressure in the z direction to 1 bar and equilibrated for 42 ns, further shrinking the APL to 47.5 $Å^2$ (Table 4.6). The final equilibration of this

system involved expanding the monolayer for 36 ns using a pressure of -25 bar in the xy direction, then 6 ns at -35 bar, and for 4 ns at -25 bar to achieve an APL of 71.9 Å^2 (Table 4.7).

Similarly, the DPPC:PXPG (80:20 mol%) monolayer was equilibrated for 2 ns at a pressure of 1 bar, shrinking the APL to 54.6 Å^2. The APL was brought down to 47.4 Å^2 by equilibrating the system for another 14 ns at the same pressure (Table 4.8). To expand the APL of the system to 72 Å^2, the system was first equilibrated for 10 ns at -25 bar, followed by 2 ns at -20 bar (Table 4.9).

The mixed monolayer containing unoxidised POPG was shrunk to 47.2 Å^2 by equilibrating the system for 6 ns using a pressure of 1 bar (Table 4.10). Equilibrating the system for 2 ns at -25 bar achieved an APL of 59.2 Å^2. Expanding the APL to 73 Å^2 required a further 12 ns at -25 bar, and subsequently 4 ns at -30 bar (Table 4.11).

In the final mixed monolayer containing oxidised PXPG, the system was equilibrated for 8 ns at 1 bar to shrink the APL to 47.1 Å^2 (Table 4.12). The file generated from the first equilibration (step6.6_equilibration.gro) had an APL of 57 Å^2, which did not require further equilibration. The system was then expanded to an APL of 73 Å^2, by equilibrating for 6 ns at a pressure of -25 bar (Table 4.13).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area ($Å^2$)	APL (upper leaflet) ($Å^2$)
ref_p = 1.0 bar:				
2	8.38000	70.224	7022.4	51.258
4	8.26234	68.266	6826.6	49.829
6	8.10392	65.674	6567.4	47.937
8	8.05281	64.848	6484.8	**47.334**

Table 4.2: Shrinking the DPPC monolayer to 47 $Å^2$. The APL was calculated by multiplying the box lengths (Box X * Box Y) in $Å^2$, divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area ($Å^2$)	APL (upper leaflet) ($Å^2$)
ref_p = -10.0 bar:				
2	8.14055	66.2685	6626.85	48.3712
4	8.18255	66.9533	6695.33	48.8710
6	8.16522	66.6708	6667.08	48.6648
8	8.13800	66.2270	6622.70	48.3409
10	8.12533	66.0210	6602.10	48.1905
12	8.14135	66.2816	6628.16	48.3807
ref_p = -25.0 bar:				
2	8.45225	71.4405	7144.05	52.1464
4	8.64620	74.6841	7468.41	54.5139
6	8.71303	75.9169	7591.69	**55.4138**

Table 4.3: Expanding the DPPC monolayer to 55 $Å^2$. The APL was calculated by multiplying the box lengths (Box X * Box Y) in $Å^2$, divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = -25.0 bar:				
0	8.71303	75.9169	7591.69	55.4138
2	8.86330	78.5581	7855.81	57.3417
4	8.95057	80.1127	8011.27	58.4764
6	8.93244	80.2351	8023.51	58.5658
8	8.98962	80.8133	8081.33	58.9878
10	9.12804	83.3211	8332.11	60.8183
12	9.18060	84.2834	8428.34	61.5207
14	9.19530	84.5535	8455.35	61.7179
16	9.27496	86.0249	8602.49	62.7919
18	9.33639	87.1682	8716.82	63.6264
20	9.43043	88.9330	8893.30	64.9146
22	9.47206	89.7199	8971.99	65.4890
24	9.55558	91.3091	9130.91	66.6490
26	9.54468	91.1009	9110.09	66.4970
ref_p = -35.0 bar:				
2	9.97496	99.4998	9949.98	72.6276
ref_p = -25.0 bar:				
2	9.93091	98.6229	9862.29	**71.9875**

Table 4.4: Expanding the DPPC monolayer to 72 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = 1.0 bar:				
2	9.29136	86.3294	8632.94	63.0141
4	9.29136	86.3294	8632.94	63.0141
6	8.85245	78.3659	7836.59	57.2014
8	8.75931	76.7255	7672.55	56.0040
10	8.75157	76.5900	7659.00	55.9051
12	8.71039	75.8709	7587.09	**55.3602**

Table 4.5: Shrinking the DPPC/POPG (80:20 mol%) monolayer to 55 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area ($Å^2$)	APL (upper leaflet) ($Å^2$)
ref_p (xy) = 1.0 bar, ref_p (z) = 0 bar:				
2	8.69190	75.5491	7554.91	55.1453
4	8.63921	74.6359	7463.59	54.4788
6	8.59747	73.9165	7391.65	53.9536
8	8.55786	73.2369	7323.69	53.4576
10	8.52033	72.5960	7259.60	52.9898
12	8.49081	72.0939	7209.39	52.6233
14	8.46578	71.6694	7166.94	52.3135
16	8.42881	71.0448	7104.48	51.8575
18	8.42951	71.0566	7105.66	51.8662
20	8.43788	71.1978	7119.78	51.9692
22	8.42753	71.0233	7102.33	51.8418
ref_p (xy) = 1.0 bar, ref_p (z) = 1.0 bar:				
2	8.42258	70.9399	7093.99	51.7809
4	8.40311	70.6123	7061.23	51.5418
6	8.36088	69.9043	6990.43	51.0250
8	8.33711	69.5074	6950.74	50.7353
10	8.33312	69.4409	6944.09	50.6868
12	8.31924	69.2098	6920.98	50.5181
14	8.30096	68.9059	6890.59	50.2963
16	8.27837	68.5314	6851.14	50.0229
18	8.26499	68.3101	6831.01	49.8614
20	8.26519	68.3134	6831.34	49.8638
22	8.27615	68.4947	6849.47	49.9961
24	8.26098	68.2438	6824.38	49.8130
26	8.27542	68.4826	6848.26	49.9873
28	8.25250	68.1038	6810.38	49.7108
30	8.22037	67.5745	6757.45	49.3244
32	8.23796	67.8640	6786.40	49.5358
34	8.22418	67.6371	6761.71	49.3702
36	8.21309	67.4548	6745.48	49.2371
38	8.16761	66.7099	6670.99	48.6933
40	8.11621	65.8729	6587.29	48.0824
42	8.06294	65.0110	6501.10	**47.4533**

Table 4.6: Shrinking the DPPC/POPG (80:20 mol%) monolayer to 47 Å^2. The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å^2, divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area (Å^2)	APL (upper leaflet) (Å^2)
ref_p (xy) = -25.0 bar:				
2	8.85564	78.4224	7842.24	57.2426
4	8.96245	80.3255	8032.55	58.6318
6	9.04928	81.8895	8188.95	59.7733
8	9.11588	83.0993	8309.93	60.6564
10	9.17378	84.1582	8415.82	61.4294
12	9.21949	84.9990	8499.90	62.0431
14	9.27361	85.9998	8599.98	62.7736
16	9.31190	86.7115	8671.15	63.2931
18	9.36175	87.6424	8764.24	63.9725
20	9.39757	88.3143	8831.43	64.4630
22	9.45014	89.3051	8930.51	65.1862
24	9.49250	90.1076	9010.76	65.7719
26	9.52988	90.8186	9081.86	66.2910
28	9.55860	91.3668	9136.68	66.6911
30	9.56493	91.4879	9148.79	66.7795
32	9.56184	91.4288	9142.88	66.7363
34	9.57416	91.6645	9166.45	66.9084
36	9.60137	92.1863	9218.63	67.2893
ref_p (xy) = -35.0 bar:				
2	9.75066	95.0754	9507.54	69.3981
4	9.87946	97.6037	9760.37	71.2436
6	9.98187	99.6377	9963.77	72.7283
ref_p (xy) = -25.0 bar:				
2	9.96453	99.2919	9929.19	72.4758
4	9.92383	98.4824	9848.24	**71.8850**

Table 4.7: Expanding the DPPC/POPG (80:20 mol%) monolayer to 72 Å^2. The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å^2, divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = 1 bar:				
0	8.93734	79.87604	7984.60	58.2818
2	8.65391	74.89016	7489.02	**54.6643**
4	8.45526	71.49142	7149.14	52.1835
6	8.33085	69.40306	6940.31	50.6592
8	8.22003	67.56889	6756.89	49.3204
10	8.16811	66.71802	6671.80	48.6993
12	8.15384	66.48511	6648.51	48.5293
14	8.13200	66.12942	6612.94	48.2697
16	8.05918	64.95038	6495.04	**47.4090**

Table 4.8: Shrinking the DPPC/POPG (80:20 mol%) monolayer to 55 and 47 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = -25.0 bar:				
0	9.39078	88.18675	8818.68	64.3699
2	9.66908	93.49111	9349.11	68.2417
4	9.89991	98.00822	9800.82	71.5388
6	9.83059	96.64050	9664.05	70.5405
8	9.79811	96.00296	9600.30	70.0752
10	10.05452	101.09337	10109.34	73.7908
ref_p = -20 bar:				
2	9.93118	98.62834	9862.83	**71.9915**

Table 4.9: Expanding the DPPC/POPG (80:20 mol%) monolayer to 72 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (137 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = 1 bar:				
0	8.62963	74.42392	7442.39	53.93038
2	8.35691	69.83794	6983.79	50.60721
4	8.20098	67.25607	6725.61	48.73628
6	8.07695	65.23712	6523.71	**47.27327**

Table 4.10: Shrinking the mixed monolayer with POPG to 47 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (138 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	APL (upper leaflet) (Å²)
ref_p = -25 bar:				
2	9.04285	81.77314	8177.31	**59.25590**
4	9.26827	85.74760	8574.76	62.13594
6	9.40930	88.53493	8853.49	64.15574
8	9.54524	91.11161	9111.16	66.02290
10	9.60324	92.22222	9222.22	66.82769
12	9.69395	93.97267	9397.27	68.09614
14	9.72230	94.52312	9452.31	68.49501
Changed ref_p to -30.0 bar in xy direction:				
2	9.9639	98.13656	9813.66	71.11345
4	10.03940	100.78955	10078.79	**73.03591**

Table 4.11: Expanding the mixed monolayer with POPG to 59 and 73 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (138 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area ($Å^2$)	APL (upper leaflet) ($Å^2$)
ref_p = 1 bar:				
0	8.47323	71.7956	7179.56	52.0258
2	8.25833	68.2000	6820.00	49.4203
4	8.22380	67.6309	6763.09	49.0079
6	8.16744	66.7071	6670.71	48.3385
8	8.06526	65.0484	6504.84	**47.1365**

Table 4.12: Shrinking the mixed monolayer with PXPG to 47 $Å^2$. The APL was calculated by multiplying the box lengths (Box X * Box Y) in $Å^2$, divided by the number of lipids in the upper leaflet (138 lipids).

Equilibration time (ns)	Box length (nm)	Box area (nm^2)	Box area ($Å^2$)	APL (upper leaflet) ($Å^2$)
ref_p = -25 bar:				
2	9.35150	87.4506	8745.06	63.3700
4	9.75315	95.1239	9512.39	68.9304
6	10.06016	101.2068	10120.68	**73.3383**

Table 4.13: Expanding the mixed monolayer with PXPG to 73 $Å^2$. The APL was calculated by multiplying the box lengths (Box X * Box Y) in $Å^2$, divided by the number of lipids in the upper leaflet (138 lipids).

4.2.6 Production MD

Production simulation of the monolayer at all areas per lipids were run for a total of 30 steps (300 ns) with each step being 10 ns, using the NVT ensemble with the Nose-Hoover thermostat, and a temperature time constant of 5 ps. Hydrogen bond constraints were added. Semi-isotropic Parinello-Rahman pressure coupling was used to achieve a pressure of 1.0 bar, with the compressibility set at 0 bar^{-1} in the xy direction for a fixed box size, and 4.5×10^{-5} bar^{-1} in the z direction.

4.2.7 Indexing the surfactant lipids

Prior to analysing the output from the simulations, index files of the headgroups and carbons on each acyl chain of the lipid were created, using the gmx make_ndx command. The SN_1 and SN_2 chains were indexed separately. An index file for the headgroup and acyl chains was created with the following atoms shown below. For membrane density analysis, the other groups present in the index file were not deleted, and the SN_1 and SN_2 chains were named separately and included in the index file.

The POPG, PXPG, POPC and PLPC lipid acyl chains and headgroups were indexed similarly, using the gmx make_ndx command with separate index files for the SN_1, SN_2 chains and headgroup. The unsaturated acyl chains were indexed as the SN_2 chain while the saturated acyl chains were indexed as the SN_1 chain. In the mixed monolayers, the lipids in the upper and lower leaflet were indexed separately, for example the POPG lipids in the upper leaflet remained labelled as POPG, while the POPG lipids in the lower leaflet were labelled as PXPG (Figure 4.7).

DPPC SN_1 chain:

a C31 | a C32 | a C33 | a C34 | a C35 | a C36 | a C37 | a C38 | a C39 | a C310 | a C311 | a C312 | a C313 | a C314 | a C315 | a C316

DPPC SN_2 chain:

a C21 | a C22 | a C23 | a C24 | a C25 | a C26 | a C27 | a C28 | a C29 | a C210 | a C211 | a C212 | a C213 | a C214 | a C215 | a C216

DPPC headgroup:

a O21 | a C2 | a HA | a C3 | a O31 | a HA | a C1 | a HB | a HA | a O11 | a O14 | a P | a O13 | a O12 | a C11 | a H11A | a H11B | a C12 | a H12A | a H12B | a N | a C13 | a H13A | a H13B | a H13C | a C14 | a H14A | a H14B | a H14C | a C15 | a H15A | a H15B | a H15C

POPG SN_1 chain:

a C31 | a C32 | a C33 | a C34 | a C35 | a C36 | a C37 | a C38 | a C39 | a C310 | a C311 | a C312 | a C313 | a C314 | a C315 | a C316

POPG SN_2 chain:

a C21 | a C22 | a C23 | a C24 | a C25 | a C26 | a C27 | a C28 | a C29 | a C210 | a C211 | a C212 | a C213 | a C214 | a C215 | a C216 | a C217 | a C218

POPG headgroup:

r POPG & a O22 | a O21 | a C2 | a HS | a C1 | a HB | a HA | a O11 | a O14 | a P | a O13 | a O12 | a C11 | a H11A | a H11B | a C12 | a H12A | a HO2 | a OC2 | a C13 | a H13A | a H13B | a HO3 | a OC3 | a C3 | a HX | a HY | a O31 | a O32

PXPG SN_1 chain:

a C31 | a C32 | a C33 | a C34 | a C35 | a C36 | a C37 | a C38 | a C39 | a C310 | a C311 | a C312 | a C313 | a C314 | a C315 | a C316

PXPG SN_2 chain:

a C21 | a C22 | a C23 | a C24 | a C25 | a C26 | a C27 | a C28 | a C29

PXPG Headgroup:

r PXPG & a O22 | a O21 | a C2 | a HS | a C1 | a HB | a HA | a O11 | a O14 | a P | a O13 | a O12 | a C11 | a H11A | a H11B | a C12 | a H12A | a HO2 | a OC2 | a C13 | a H13A | a H13B | a HO3 | a OC3 | a C3 | a HX | a HY | a O31 | a O32

POPC SN_1 chain:

a C31 | a C32 | a C33 | a C34 | a C35 | a C36 | a C37 | a C38 | a C39 | a C310 | a C311 | a C312 | a C313 | a C314 | a C315 | a C316

POPC SN_2 chain:

a C21 | a C22 | a C23 | a C24 | a C25 | a C26 | a C27 | a C28 | a C29 | a C210 | a C211 | a C212 | a C213 | a C214 | a C215 | a C216 | a C217 | a C218

POPC headgroup

r POPC & a O22 | a O21 | a C2 | a HS | a HX | a HY | a C3 | a O31 | a O32 | a C1 | a HA | a HB | a O11 | a P | a O12 | a O13 | a O14 | a C11 | a H11A | a H11B | a C12 | a H12A | a H12B | a N | a

C13 | a H13A | a H13B | a H13C | a C14 | a H14A | a H14B | a H14C | a C15 | a H15A | a H15B | a H15C

PLPC SN_1 chain:

a C31 | a C32 | a C33 | a C34 | a C35 | a C36 | a C37 | a C38 | a C39 | a C310 | a C311 | a C312 | a C313 | a C314 | a C315 | a C316

PLPC SN_2 chain:

a C21 | a C22 | a C23 | a C24 | a C25 | a C26 | a C27 | a C28 | a C29 | a C210 | a C211 | a C212 | a C213 | a C214 | a C215 | a C216 | a C217 | a C218

PLPC Headgroup:

r PLPC & a O22 | a O21 | a C2 | a C3 | a O31 | a O32 | a C1 | a O11 | a P | a O12 | a O13 | a O14 | a C11 | a C12 | a N | a C13 | a C14 | a C15 | a HX | a HY | a HS | a HA | a HB | a H11A | a H11B | a H12A | a H12B | a H13A | a H13B | a H13C | a H14A | a H14B | a H14C | a H15A | a H15B | a H15C

Cholesterol:

r CHL1 & a H3 | a O3 | a C1 | a H1A | a H1B | a C2 | a H2A | a H2B | a C3 | a H3 | a C4 | a H4A | a H4B | a C5 | a C6 | a H6 | a C7 | a H7A | a H7B | a C8 | a H8 | a C9 | a H9 | a C10 | a C19 | a H19A | a H19B | a H19C | a C11 | a H11A | a H11B | a C12 | a H12A | a H12B | a C13 | a C14 | a H14 | a C15 | a H15A | a H15B | a C16 | a H16A | a H16B | a C17 | a H17 | a C18 | a H18A | a H18B | a H18C | a C20 | a C21 | a H21A | a H21B | a H21C | a C22 | a H22A | a H22B | a C23 | a H23A | a H23B | a C24 | a H24A | a H24B | a C25 | a H25 | a C26 | a H26A | a H26B | a H26C | a C27 | a H27A | a H27B | a H27C

Figure 4.7: Index groups for all lipid acyl chains and headgroups.

4.2.8 Data analysis

In this study, the order parameters, membrane density and the secondary helical structure of Sp-B_{1-25} at the measured APLs and surface pressures were investigated. The deuterium order parameters are a measure of the motional disorder of the hydrocarbon chains measured in S_{CD}, which is a parameter that shows the orientational mobility of the C-H bonds on the lipid tails, SN_1 and SN_2, where a higher S_{CD} value implies more order within the lipid chains, and a lower S_{CD} value implies more motional disorder. The deuterium order parameters were calculated with the normal to the bilayer along the z axis using the gmx order command, where order (measured in S_{CD}) on the y axis was plotted against atom number on the x axis. The order parameters of the DPPC and POPG acyl chains were generated for each measured APL (approximately at 47, 55, and 72 $Å^2$), at 150 ns and 300 ns.

The gmx density command was used to create density plots for the headgroup, acyl chains, water (TIP3) and peptide, from the headgroup index files of DPPC and POPG as created previously. All groups were plotted on a partial density graph, with density (kg m^{-3}) on the y axis plotted against z-coordinates (nm), which shows the distance and position of each monolayer component from the box center. The partial density graphs of Sp-B_{1-25} in DPPC were analysed at each APL, at 150 ns and 300 ns. The RMSD and least squares fit of the peptide's trajectory with respect to the backbone reference structure was calculated with the gmx rms command, for the final 150 ns of the production simulations. This was done using the TPR files containing topology and coordinate information, and the compressed trajectory XTC files at each 10 ns step from 150 ns to 300 ns.

Parameters to analyse the alpha helix were computed using the analysis toll helix supplied with GROMACS, with the output XTC and TPR files at each 10 ns step from 150 ns to 300 ns. XVG files were generated from the following parameters: the average helical angle per residue (twist), rise per residue, helix radius, average H bond length, helicity, helix length, RMS deviation from ideal helix and the average RMSD per residue. Similarly, the hbond analysis tool by GROMACS was used to calculate the number of hydrogen bonds and the number of pairs within 0.35 nm, using the output files at each 10 ns step between 150 ns and 300 ns. The two specified groups were for Sp-B_{1-25} (432 elements) and DPPC (34450 elements).

4.3 Results

4.3.1 Surface pressure analysis

The surface pressure was calculated by subtracting the average value of the number of surfaces of the lipids multiplied by the surface tension in mN m^{-1}, from the experimental surface tension value of water at 37˚C (~310 K) of 70.09 mN m^{-1}, which was interpolated from the values obtained from Vargaftik *et al.* (Vargaftik, Volkov and Voljak, 1983). In all systems, there is an inverse relationship between surface pressure and APL (Figure 4.8). In the DPPC monolayer, the surface pressure decreases from 61.2 to 22.4 mN m^{-1} as the APL increases from 47.3 to 72 $Å^2$. In the DPPC POPG (80:20 mol%) monolayer, the surface pressure decreases from 56.5 to 22.5 mN m^{-1}, as the APL increases from 47.5 to 71.9 $Å^2$. In the DPPC PXPG (80:20 mol%) oxidised monolayer, the surface pressure decreases from 47 to 29.3 mN m^{-1} as the APL increases from 47.4 to 72 $Å^2$. In the unoxidised mixed monolayer, the surface pressure decreases from 60.1 to 17.3 mN m^{-1} as the APL increases from 47.3 to 73 $Å^2$. Finally, in the oxidised mixed monolayer, the surface pressure decreases from 61.9 to 18.9 mN m^{-1} while the APL increases from 47.1 to 73.3 $Å^2$. When the pressure-area isotherms of Sp-$B_{1\text{-}25}$ in a pure DPPC monolayer was compared to pressure-area isotherms of a DPPC monolayer from MD simulations by Baoukina *et al.*, the data was in agreement with the experiments done at 310 K (Figure 4.9).

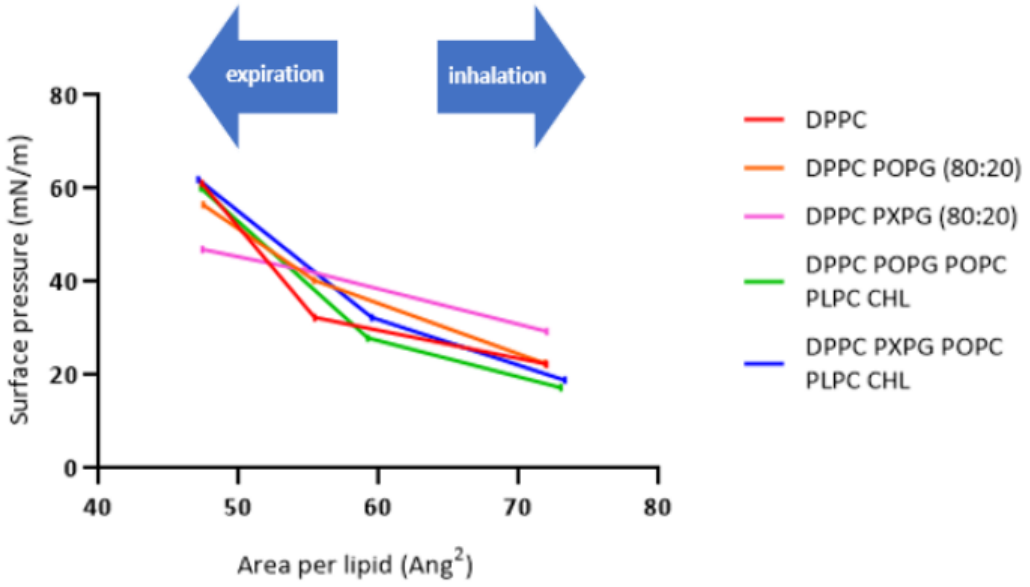


Figure 4.8: Surface pressure-area isotherms of Sp-$B_{1\text{-}25}$ in various lipid monolayers. DPPC (red), DPPC and POPG (80:20 mol%) (orange), DPPC and PXPG (80:20 mol%) (pink), mixed lipid monolayer with POPG (green) and mixed monolayer with PXPG (blue).

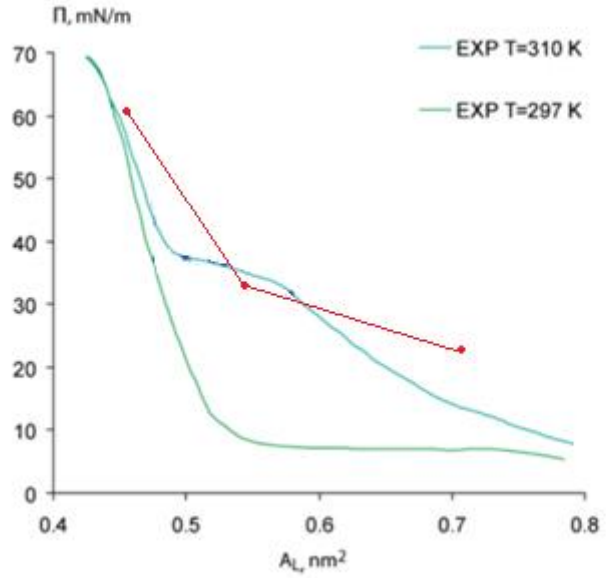


Figure 4.9: Surface pressure-area isotherms obtained from MD simulations of DPPC monolayers. Sp-$B_{1\text{-}25}$ in DPPC (red) and pure DPPC monolayers at 297 K (green) and 310 K (blue).

The surface pressure of both DPPC and DPPC/POPG (80:20 mol%) monolayers decrease with increasing APL (Figure 4.8). During inhalation, the alveoli expands, and the alveolar radius increases. The surfactant film expands, resulting in a larger APL and surface tension, whereas the surface pressure decreases and vice versa.

The saturated nature of DPPC allows pulmonary surfactant to withstand high surface pressures due to its ability to form highly packed liquid condensed phase at lung temperatures. Hence, at the lowest measured APL of 47.4 $Å^2$, the DPPC monolayer reaches a higher surface pressure of 61.2 mN m^{-1} compared to the DPPC/POPG monolayer with a surface pressure of 56.5 mN m^{-1}. In pulmonary surfactant, unsaturated lipids like POPG help to increase the fluidity and re-spreading of surfactant. The mixed DPPC/POPG monolayer has increased fluidity, hence surface pressure of the DPPC/POPG monolayer is 40.4 mN m^{-1}, while the DPPC monolayer is 32.3 mN m^{-1} at 55.4 $Å^2$, due to the lower compressibility of the mixed monolayer system. The highest measured APL of 71.9 $Å^2$ simulates the environment of pulmonary surfactant when the lungs are fully expanded. Both simulated monolayers showed a surface pressure value of 22.5 mN m^{-1}.

4.3.2 Further expansion of the DPPC monolayer

Expanding the DPPC monolayer beyond 72 Å^2 resulted in shrinkage of the monolayer to 69.5 Å^2 when equilibrated with a negative pressure of -25 bar in the xy direction. Increasing the negative pressure to -35 bar and equilibrating for 10 ns expanded the monolayer to 100.2 Å^2, and to stabilise it, the monolayer was equilibrated for another 2 ns at -25 bar, reaching an APL of 93.6 Å^2. Over-expanding the monolayer using high negative pressures resulted in instability, where the gradual expansion from 87.4 Å^2 (Figure 4.10a) to 100.2 Å^2 (Figure 4.10d) caused significant parts of the monolayer to disintegrate and rupture, and equilibrating the monolayer at a lower negative pressure after did not help with improvement of the monolayer stability.

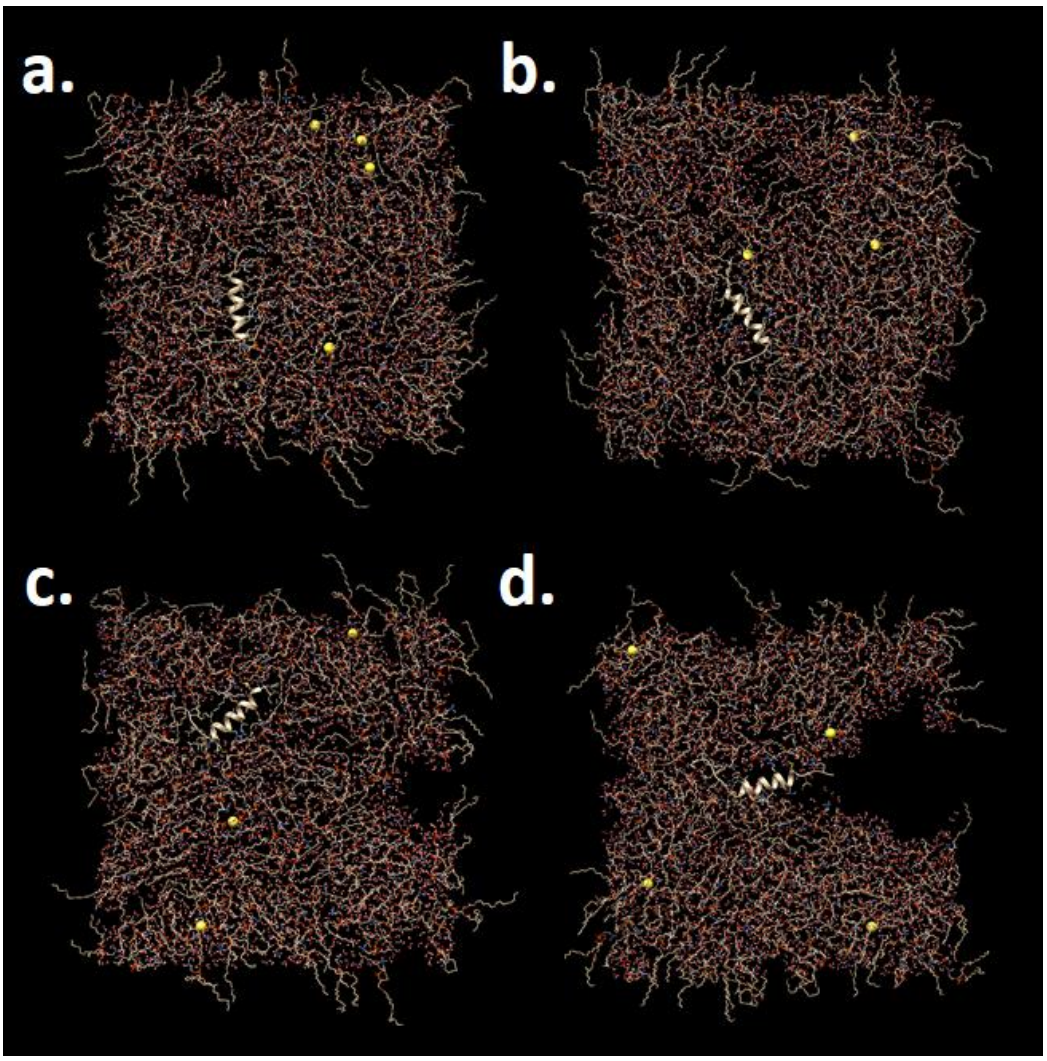


Figure 4.10: Sp-B$_{1-25}$ in DPPC expanded to an APL of **(a)** 87.4 Å^2, **(b)** 90.7 Å^2, **(c)** 96.7 Å^2 and **(d)** 100.2 Å^2.

Equilibration step (ns)	Box length (nm)	Box area (nm²)	Box area (Å²)	Area of lipid (upper leaflet) (Å²) (137 lipids in box)
ref_p in xy direction = -35.0 bar:				
2	10.94104	119.7063	11970.63	87.3769
4	11.14798	124.2774	12427.74	90.7135
6	11.33510	128.4844	12848.44	93.7843
8	11.50875	132.4513	13245.13	96.6798
10	11.71434	137.2258	13722.58	100.1648
ref_p = -25.0 bar:				
2	11.32451	128.2445	12824.45	93.6091

Table 4.14: Equilibration and expansion of the DPPC monolayer beyond 72 Å². The APL was calculated by multiplying the box lengths (Box X * Box Y) in Å², divided by the number of lipids in the upper leaflet (137 lipids).

4.3.3 Deuterium order parameters

Order parameter data was obtained to investigate the behaviour of the SN_1 and SN_2 lipid tails in the simulated monolayers, by measuring ordered or disordered the lipid acyl chains are. The motional disorder of the carbons on the acyl chains are measured in S_{CD}, where a higher value implies higher order in the chains and a lower value implies more disorder. The order parameters, S_{CD}, are calculated using Equation 4.7, where θ is the angle between the C-H bond vector and the normal of the monolayer, and the angular brackets refer to molecular and temporal ensemble averages.

$$S_{CD} = \frac{\langle 3\cos\theta \rangle}{2} \qquad (4.7)$$

Figures 4.11 to 4.15 show the order parameters for all simulated monolayers. With increasing APL, both the SN_1 and SN_2 chains of DPPC in the DPPC-only monolayer become more disordered, with Carbon 1 on the SN_1 chain being more ordered than Carbon 1 on the SN_2 chain, except at the highest APL (Figure 4.11).

In the DPPC:POPG and DPPC:PXPG systems, the DPPC, POPG and PXPG SN_1 and SN_2 chains are observed to behave in the same way, where both lipid tails become more disordered at higher APLs in both oxidised and unoxidised monolayers. The PXPG lipid tails appear to be more disordered than its POPG counterpart at the measured APLs. Carbons 8 to 11 on the SN_2 chain of POPG in both monolayers appear more disordered compared to the rest of the carbons on the same chain, due to the presence of the carbon double bond between Carbon 8 and 9. Comparing Figures 4.11 and 4.12, at the small and medium APL, it is observed that adding POPG or PXPG increases the order of the SN_1 chain of DPPC. Additionally, the SN_1 chains of DPPC are more ordered at the same APL compared to the SN_1 chains of POPG and PXPG, in the DPPC:POPG and DPPC:PXPG mixtures.

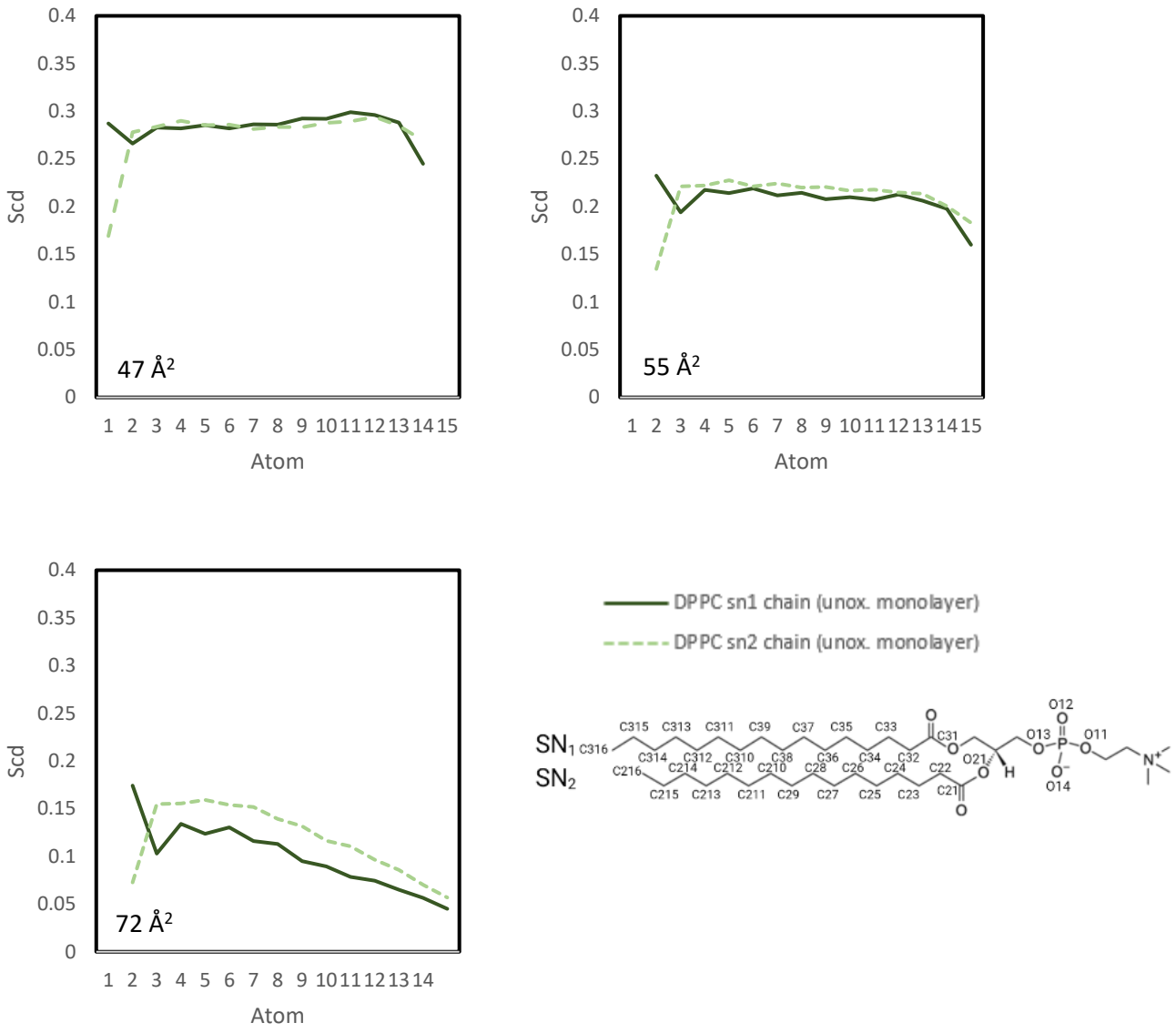
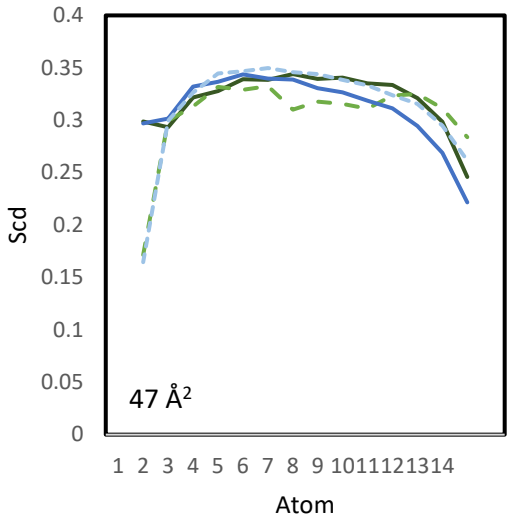
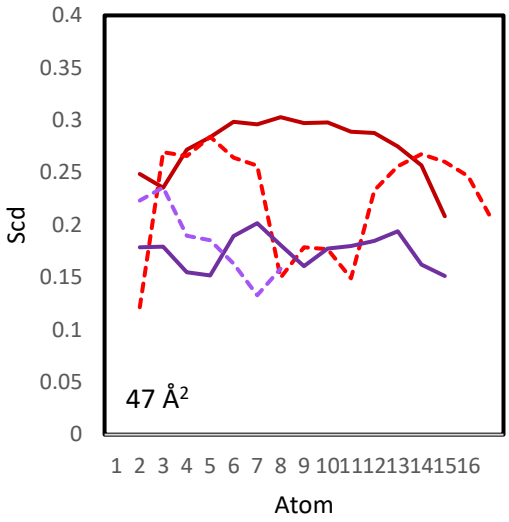
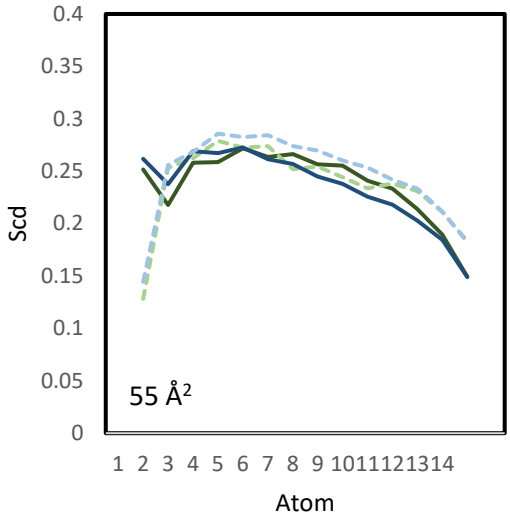
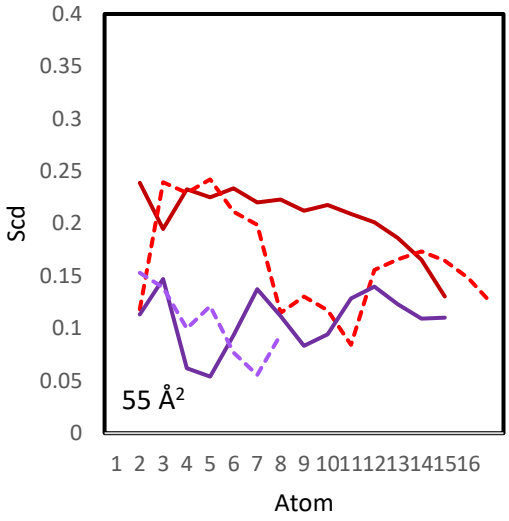


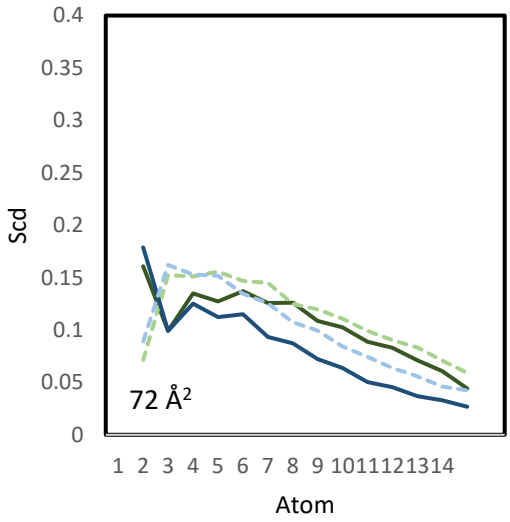
Figure 4.11: Order parameter profiles of the DPPC SN_1 (dark green) and SN_2 (light green) chains at the measured APLs.

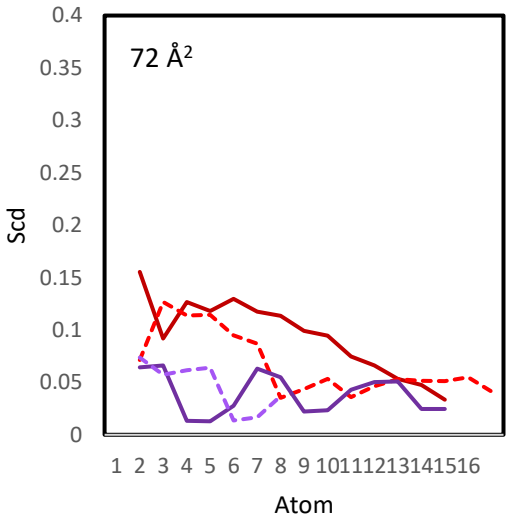

0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
47 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Atom


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
47 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Atom


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
55 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Atom


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
55 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Atom


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
72 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14
Atom


0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Scd
72 Å²
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Atom

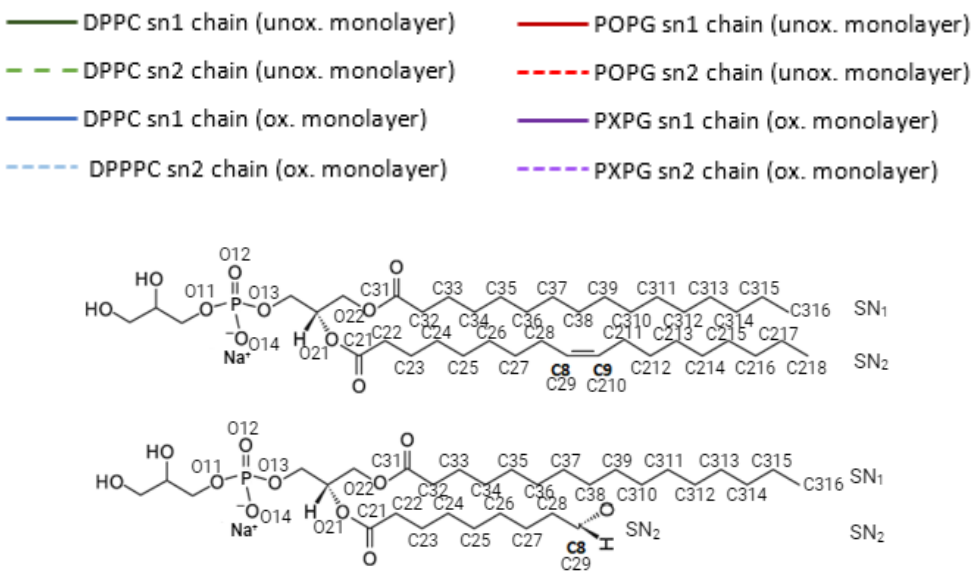
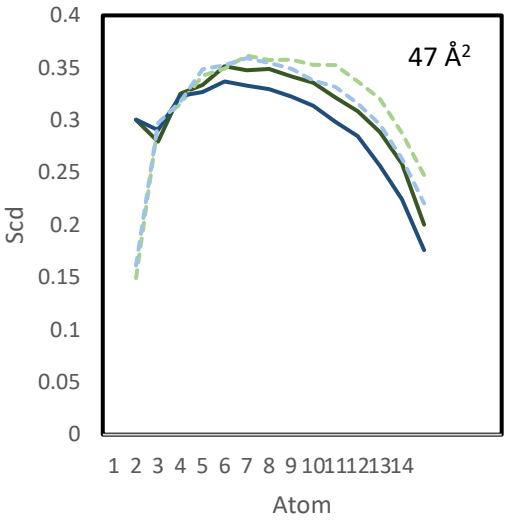
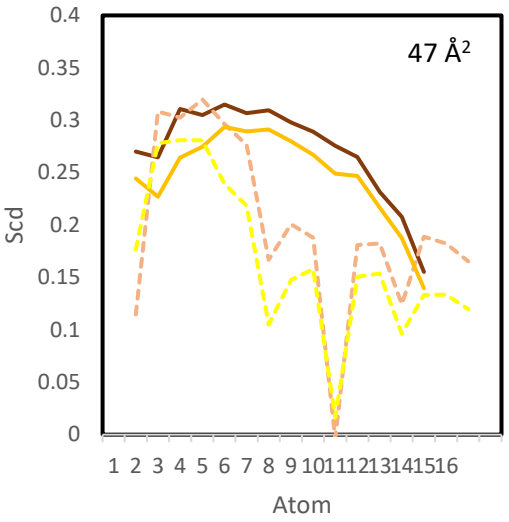
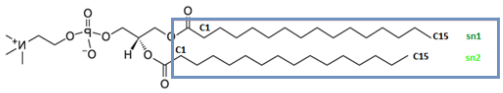


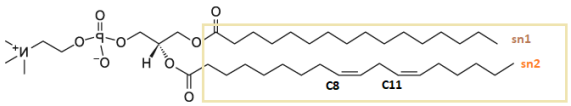
Figure 4.12: Order parameter profiles of the lipid acyl chains in DPPC:POPG (80:20 mol%) and DPPC:PXPG (80:20 mol%) monolayers. **Left column:** DPPC SN_1 (dark green) and SN_2 (light green, dashed) chains in the DPPC:POPG monolayer. DPPC SN_1 (dark blue) and SN_2 (light blue, dashed) chains in the DPPC:PXPG monolayer. **Right column:** Order parameter profiles of the POPG SN_1 (dark red) and SN_2 (light red, dashed) chains, and the PXPG SN_1 (dark purple) and SN_2 (light purple, dashed) chains.

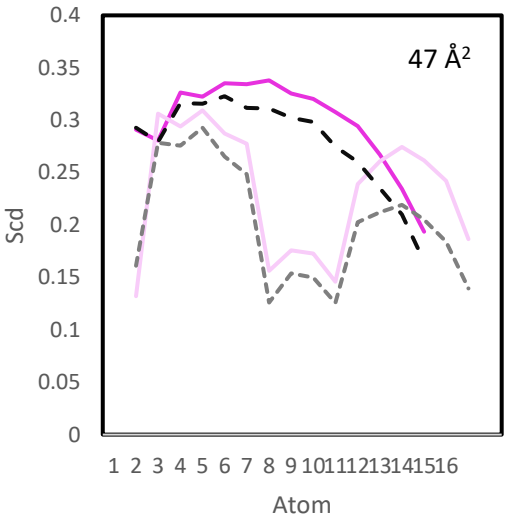
In the unoxidised and oxidised mixed monolayers, it was observed that there was a high level of disorder for the carbon atoms involved in, or close to the carbon double bond on the SN_2 chain of unsaturated lipids (PLPC, POPC, POPG and PXPG) (Figure 4.13-4.15). The SN_1 chains of DPPC at 47 and 72 $Å^2$ appear to have similar levels of order in both the unoxidised and oxidised mixed monolayers, and the DPPC:POPG and DPPC:PXPG mixtures (Figure 4.13-4.15). At 55 $Å^2$ the DPPC SN_1 chains in the mixed monolayers appear to be more disordered compared to the SN_1 chains in the DPPC:POPG and DPPC:PXPG mixtures (Figure 4.14). Interestingly, at the lowest APL, the POPG and PXPG SN_1 chains were more ordered in the mixed monolayers, compared to the DPPC:POPG and DPPC:PXPG mixtures (Figure 4.12, Figure 4.13). At 55 and 72 $Å^2$, the PXPG SN_1 chain is more ordered in the mixed monolayer compared to the DPPC:PXPG mixture (Figure 4.12, 4.14, 4.15). At all APLs, the SN_2 chain of the PXPG lipid is far more disordered in the mixed monolayer compared to the DPPC:PXPG mixtures (Figure 4.12-4.15).

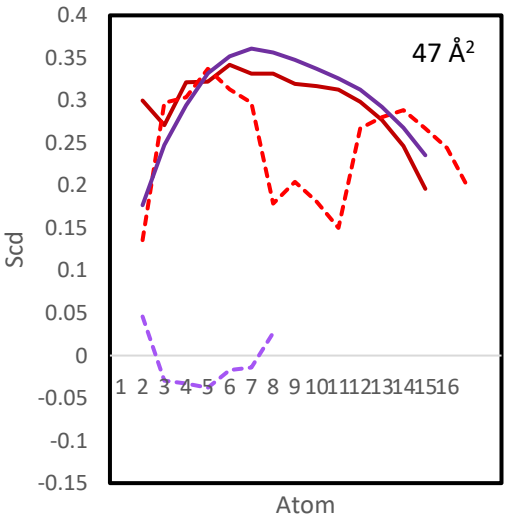


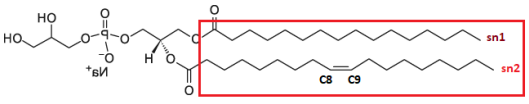


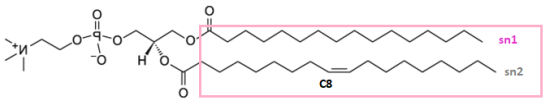


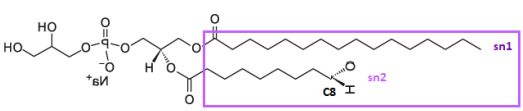












DPPC sn1 chain (unox. monolayer)
DPPC sn2 chain (unox. monolayer)
DPPC sn1 chain (ox. monolayer)
DPPC sn2 chain (ox. monolayer)
PLPC sn1 chain (unox. monolayer)
PLPC sn2 chain (unox. monolayer)
PLPC sn1 chain (ox. monolayer)
PLPC sn2 chain (ox. monolayer)

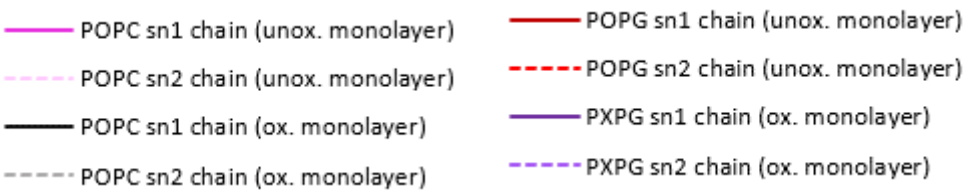


Figure 4.13: Order parameter profiles of the lipid acyl chains in the unoxidised and oxidised mixed monolayers, at 47 Å^2. **Top left:** DPPC SN_1 (dark green) and SN_2 (light green, dashed) chains in the unoxidised monolayer, DPPC SN_1 (dark blue) and SN_2 (light blue, dashed) chains in the oxidised monolayer. **Top right:** PLPC SN_1 (dark brown) and PLPC SN_2 (light brown, dashed) chains in the unoxidised monolayer, PLPC SN_1 (orange) and PLPC SN_2 (yellow, dashed) chains in the oxidised monolayer. **Bottom left:** POPC SN_1 (pink) and POPC SN_2 (light pink, dashed) in the unoxidised monolayer, POPC SN_1 (dark grey) and POPC SN_2 (light grey, dashed) in the oxidised monolayer. **Bottom right:** POPG SN_1 (red) and POPG SN_2 (light red, dashed) in the unoxidised monolayer, PXPG SN_1 (purple) and PXPG SN_2 (light purple) in the oxidised monolayer.

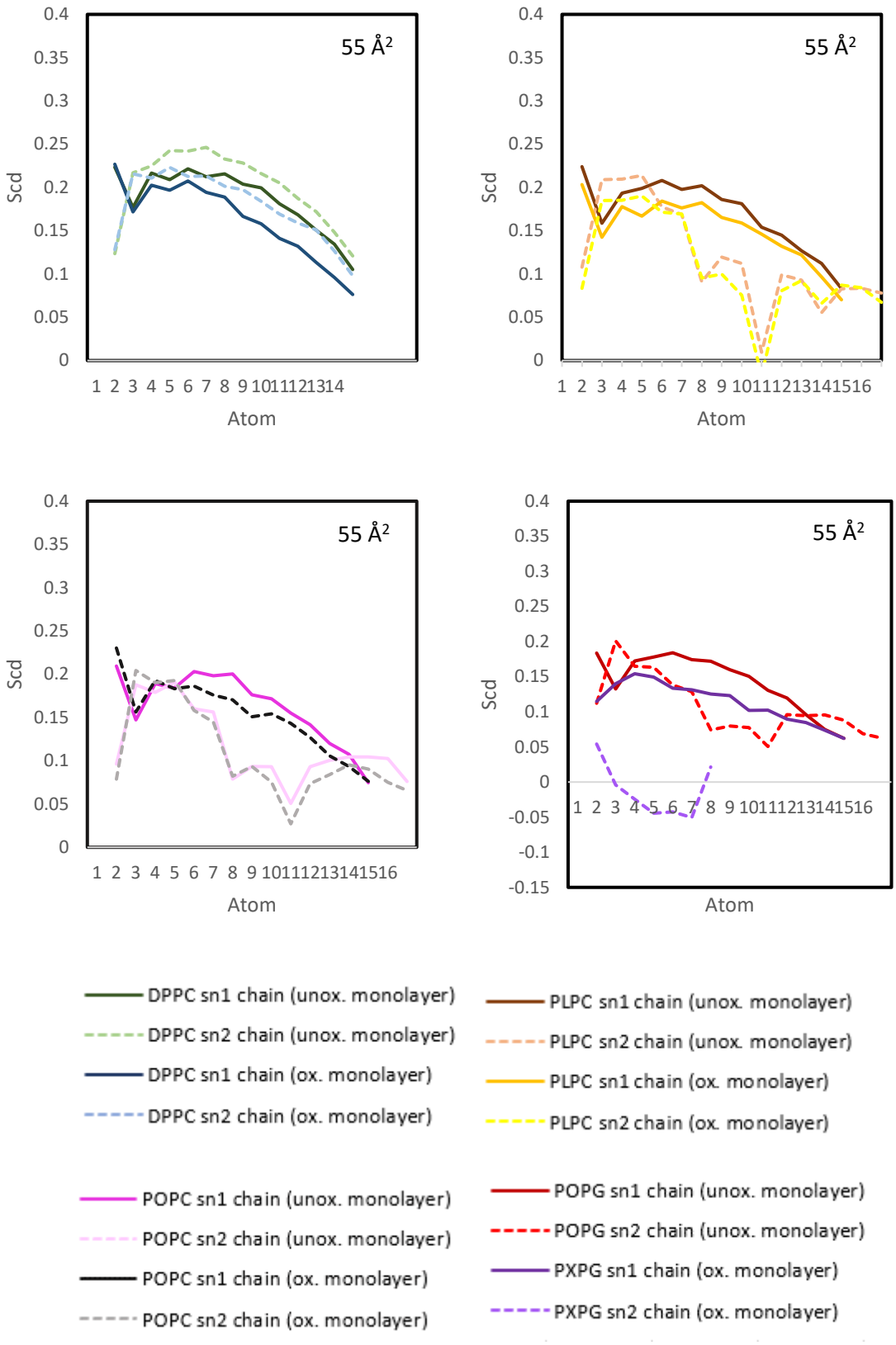


Figure 4.14: Order parameter profiles of the lipid acyl chains in the unoxidised and oxidised mixed monolayers, at 55 Å².

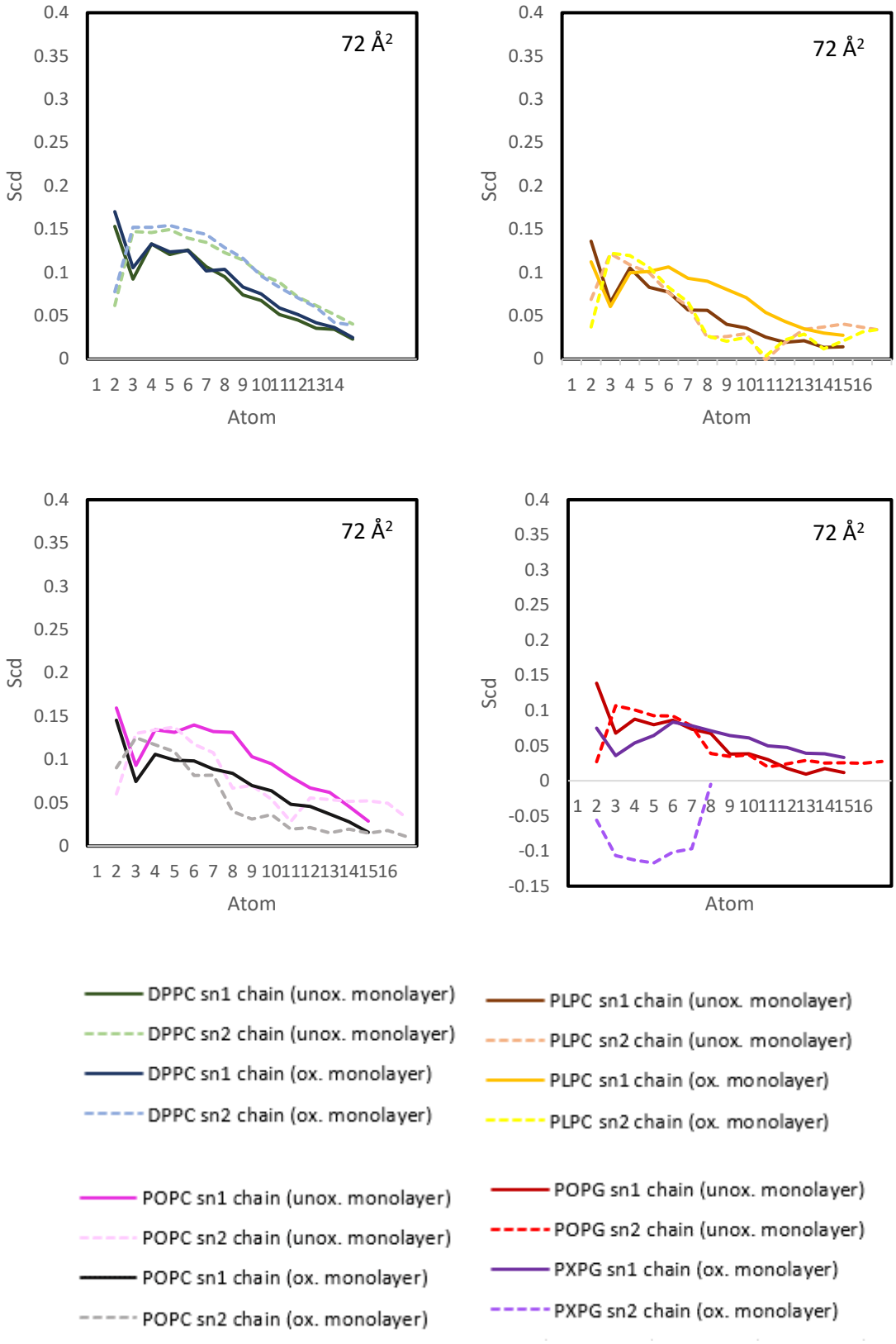


Figure 4.15: Order parameter profiles of the lipid acyl chains in the unoxidised and oxidised mixed monolayers, at 72 Å².

4.3.4 Partial density

The density profiles of the five monolayers at the three measured APLs are presented in Figures 4.16 to 4.20, illustrating the positioning of each group in relation to the simulation box. Only the leaflet containing Sp-$B_{1\text{-}25}$ is shown. The headgroup, indexed previously, was defined as containing the choline, phosphate, and glycerol groups. The lipid acyl chains were defined as the CH_2, CH_3 and CHO groups on the SN_1 and SN_2 chains. In all systems, the lipid tails face the air/vacuum interface while the headgroups face the water, as expected of their hydrophobic and hydrophilic nature. The overall density of the lipids is seen to decrease with increasing APL, due to more space between molecules.

In the DPPC-only monolayer, the DPPC headgroups can be seen to become less hydrated as the APL is decreased. The density of Sp-$B_{1\text{-}25}$ is lower at higher APL, which suggests that it adopts a different conformation at different APLs. Sp-$B_{1\text{-}25}$ can be seen to extend across a larger z-range as the APL decreases, suggesting a change in orientation. As the system expands, Sp-$B_{1\text{-}25}$ is seen to shift from being inserted between the lipid headgroup and tail regions, to being completely embedded in the solvated head region at the highest APL (Figure 4.16).

In the DPPC:POPG (80:20 mol%) monolayer, the DPPC headgroups and acyl chains generally appear to lie on top of each other, suggesting that the lipid tails are in the same place where they are not squeezed into the subphase. They also appear denser compared to the PXPG headgroups and acyl chains in the DPPC:PXPG (80:20 mol%) monolayer. In both systems, the peptide gradually inserts itself into the water region as the system expands to a higher APL. Most of the lipid headgroups in the unoxidised monolayer remained in the water region (Figure 4.17), and in the oxidised monolayer, the PXPG SN_2 acyl chains remain inserted between the air and water interface with the other lipid tails facing the air interface (Figure 4.18).

In both mixed monolayers containing cholesterol, Sp-$B_{1\text{-}25}$ exhibits similar behaviours, where it positions itself between the air-water interface at lower APLs and associates closely with the headgroup and water region at higher APLs. The presence of oxidative damage in the mixed monolayer did not largely affect the positioning of the peptide due to the presence of other unsaturated lipids (Figure 4.19, 4.20).

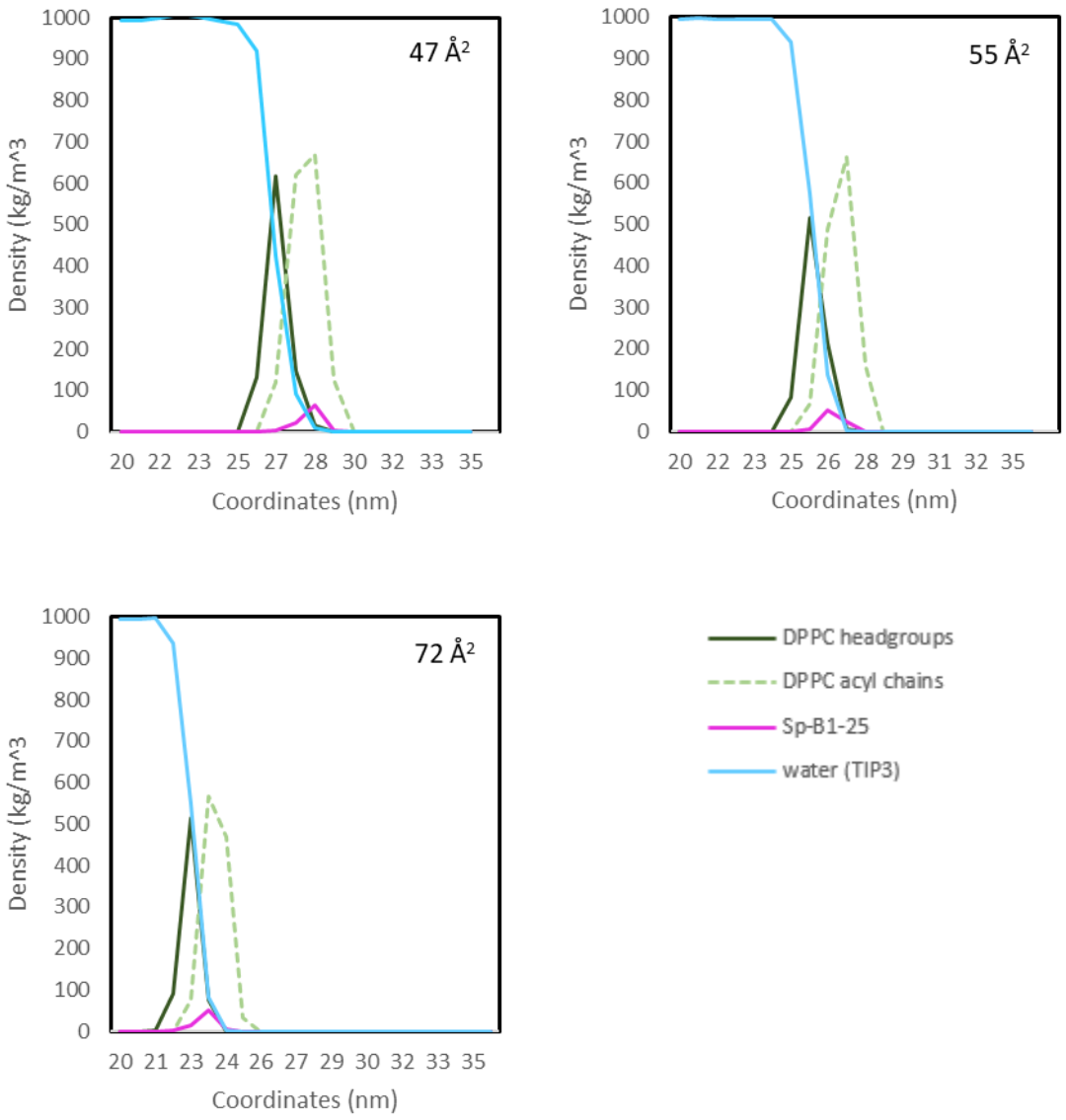


Figure 4.16: Density profiles of Sp-$B_{1\text{-}25}$ (pink), DPPC headgroups (dark green), DPPC acyl chains (light green) and water (blue) in DPPC-only monolayers, measured at 47 Å², 55 Å² and 72 Å².

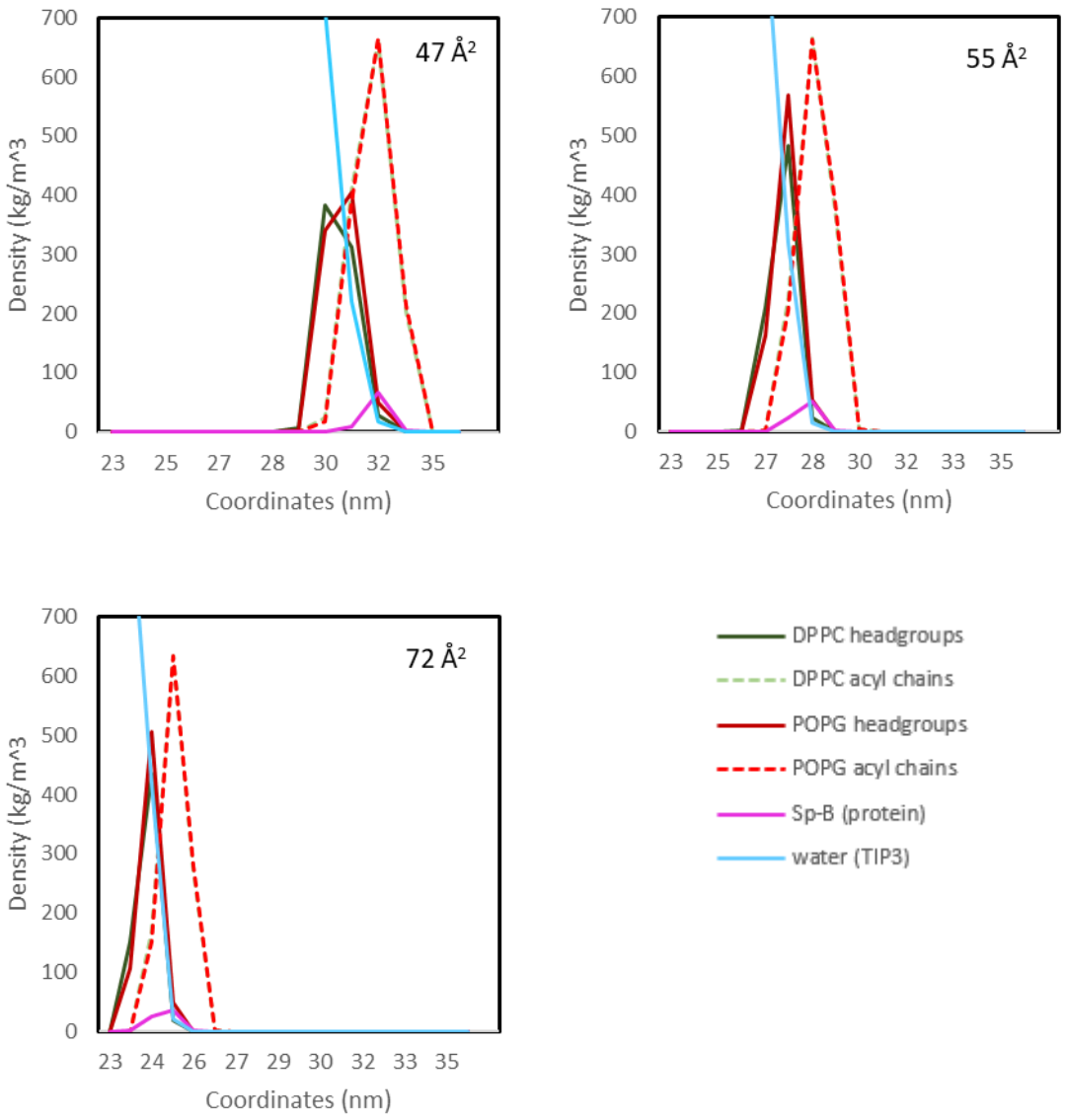


Figure 4.17: Density profiles of Sp-$B_{1\text{-}25}$ (pink), DPPC headgroups (dark green), DPPC acyl chains (light green), POPG headgroups (dark red), POPG acyl chains (red) and water (blue) in DPPC:POPG (80:20 mol%) monolayers, measured at 47 Å², 55 Å² and 72 Å².

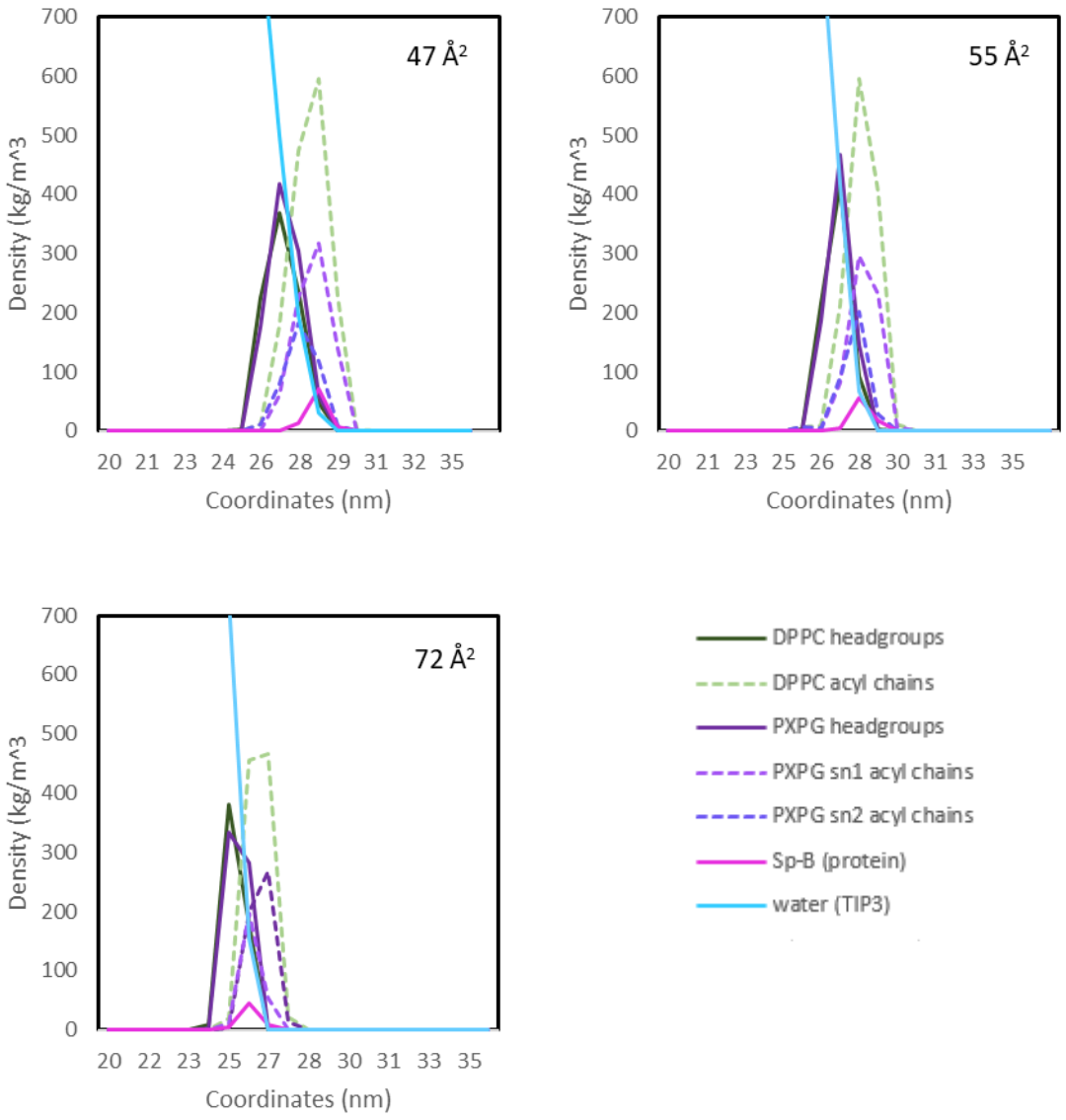


Figure 4.18: Density profiles of Sp-$B_{1\text{-}25}$ (pink), DPPC headgroups (dark green), DPPC acyl chains (light green), PXPG headgroups (dark purple), PXPG SN_1 (light purple) and SN_2 (purple) acyl chains and water (blue) in DPPC:PXPG (80:20 mol%) monolayers, measured at 47 Å², 55 Å² and 72 Å².

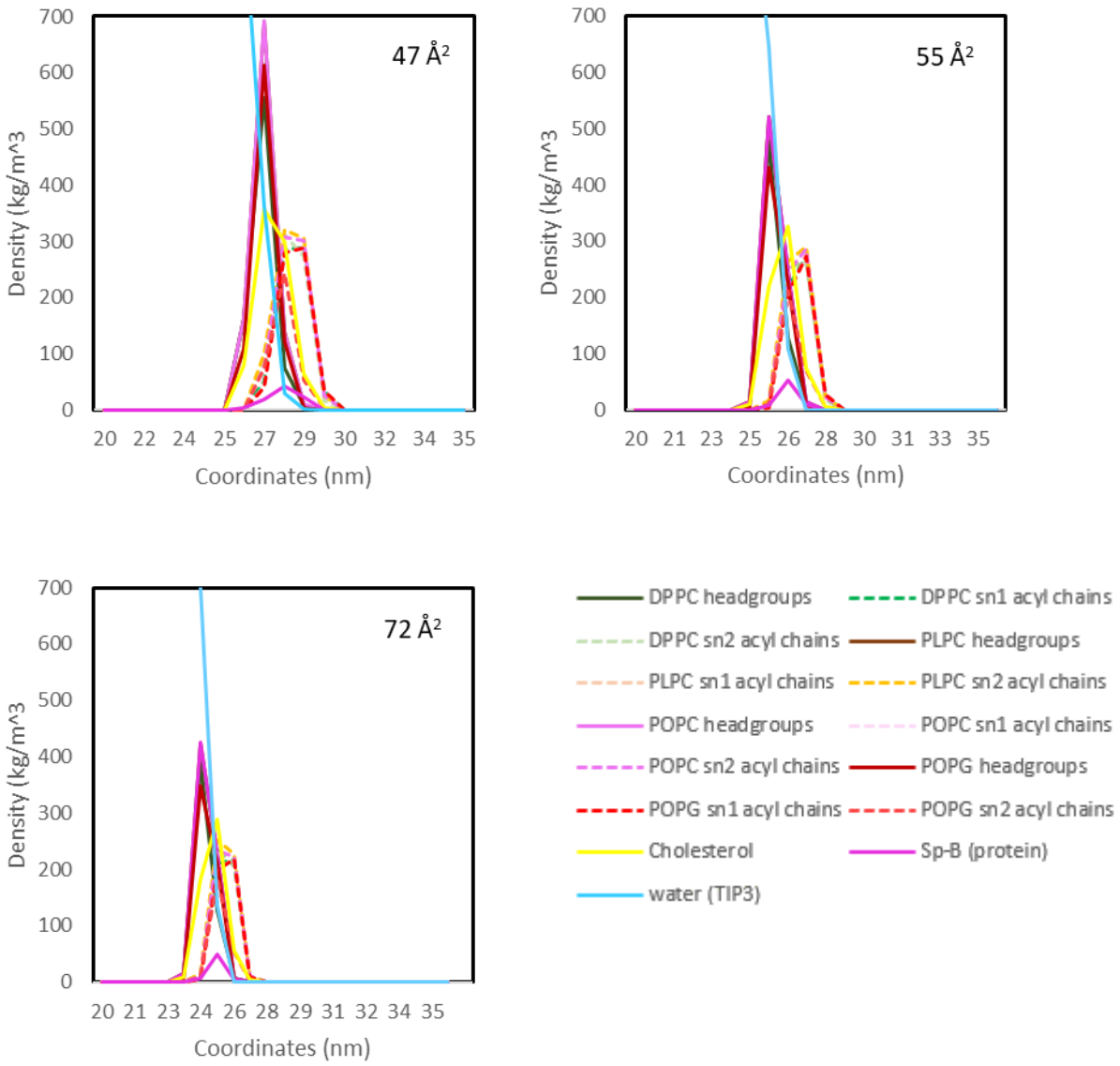


Figure 4.19: Density profiles of Sp-$B_{1\text{-}25}$ (pink), DPPC, PLPC, POPC, and POPG headgroups, SN_1 and SN_2 acyl chains, cholesterol (yellow) and water (blue) in unoxidised mixed monolayers, measured at 47 Å², 55 Å² and 72 Å².

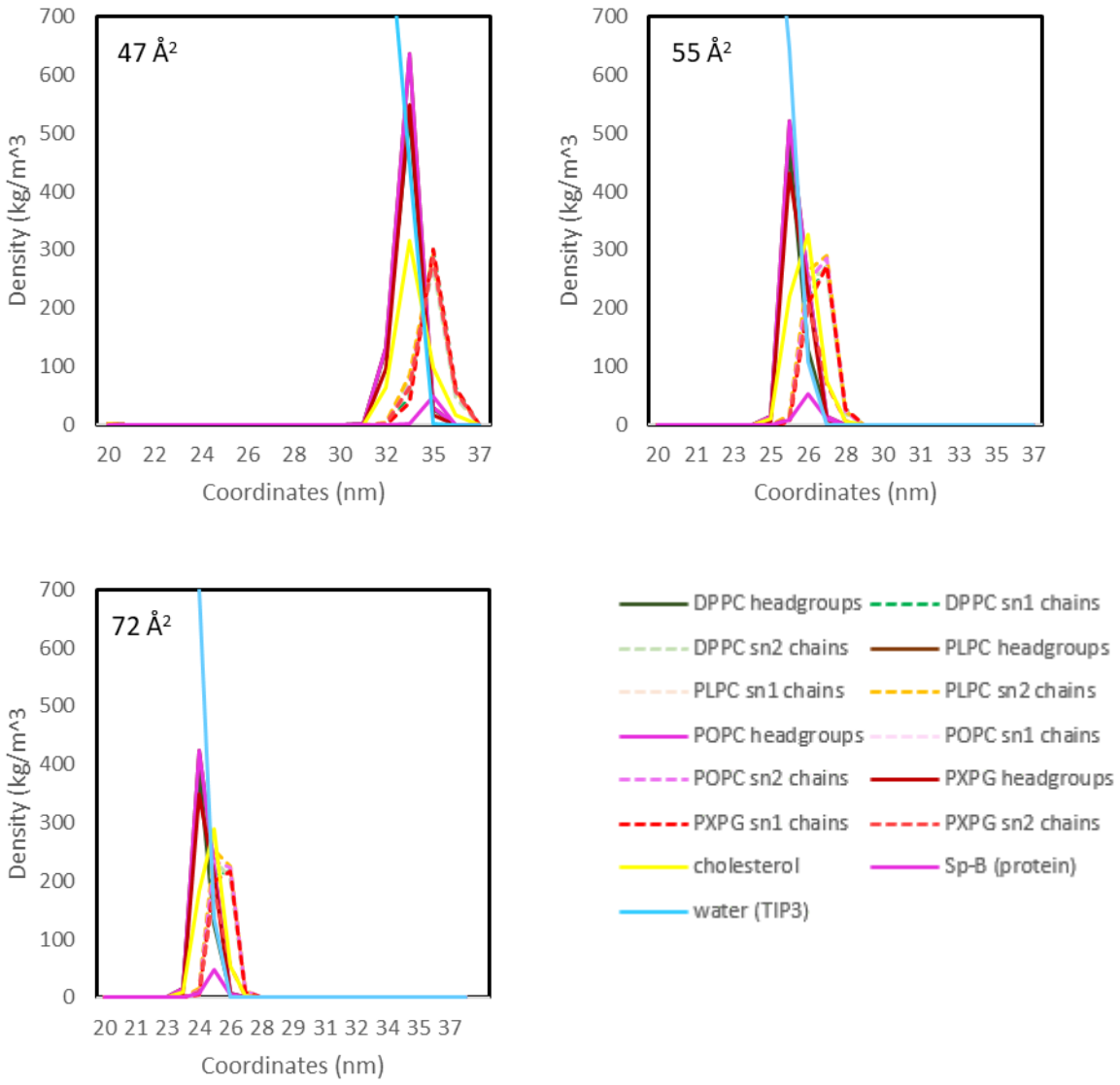


Figure 4.20: Density profiles of Sp-$B_{1\text{-}25}$ (pink), DPPC, PLPC, POPC, and PXPG headgroups, SN_1 and SN_2 acyl chains, cholesterol (yellow) and water (blue) in oxidised mixed monolayers, measured at 47 Å², 55 Å² and 72 Å².

4.3.5 Secondary structure analysis of Sp-B_{1-25}

To examine how the structure of Sp-B_{1-25} changes in an environment with different surfactant lipid compositions and APLs, several helix properties were calculated using the tools provided with GROMACS. Tables 4.15 to 4.18 show the helix parameters of Sp-B_{1-25} computed using the gmx helix command, where the mean values were taken from the last 150 ns simulation of each system.

The average helical angle per residue provides insight into the type of helix the peptide has. Fundamentally, alpha helices possess 3.6 residues per turn, with an average helical angle per residue of 100˚ around the helical axis. A 3_{10} helix would have 3 residues per turn, with an angle of 120˚ per residue (Vieira-Pires and Morais-Cabral, 2010), and a π helix having 4.1 residues per turn with an angle of 87˚ per residue. Table 4.15 shows the data agreeing with Sp-B_{1-25} having a largely alpha helical structure.

A smaller RMSD from the ideal helix would indicate more similarity between the Cα atoms to that of the ideal helix. The Cα atom is simply the first carbon atom attached to a functional group, an example being the central point and backbone of every residue connecting the amino, carboxyl group and sidechains. Data in Table 4.16 agrees with Sp-B_{1-25} having a low RMSD from the ideal alpha helix.

The helix radius refers to the RMS deviation in two dimensions for all Cα atoms, where in this equation it is computed as the square root of the sum of the RMSD of the two dimensions, x and y, divided by the number of backbone atoms N as shown in Equation 4.8. In an ideal helix, this results in a radius of approximately 0.23 nm. In Table 4.17, it is seen that Sp-B_{1-25} has an ideal helix radius.

$$\sqrt{\left(\frac{sum_i\left(x^{2(i)}+y^{2(i)}\right)}{N}\right)} \qquad (4.8)$$

The rise per residue is the difference in z-coordinate between Cα atoms, with the value being 0.15 nm for an ideal helix. Table 4.18 shows that the values obtained from Sp-B_{1-25} were again indicative of an ideal helix.

Monolayer	47 Å²	55 Å²	72 Å²
DPPC	99.809	98.584	99.373
DPPC POPG	100.46	98.864	100.55
DPPC PXPG	110.84	99.911	99.515
Mixed unox. monolayer	97.974	99.346	99.129
Mixed ox. monolayer	101.47	99.067	99.987

Table 4.15: Average helical angle per residue of Sp-B_{1-25}. A value of 100° is expected for an alpha helix.

Monolayer	47 Å²	55 Å²	72 Å²
DPPC	0.0427	0.0456	0.0459
DPPC POPG	0.0349	0.0427	0.0447
DPPC PXPG	0.0399	0.0382	0.0384
Mixed unox. monolayer	0.0577	0.0452	0.0498
Mixed ox. monolayer	0.0386	0.0531	0.0443

Table 4.16: RMS deviation from the ideal helix of Sp-B_{1-25}.

Monolayer	47 Å²	55 Å²	72 Å²
DPPC	0.2314	0.2327	0.2299
DPPC POPG	0.2303	0.2333	0.2289
DPPC PXPG	0.228	0.2311	0.2309
Mixed unox. monolayer	0.234	0.2318	0.2282
Mixed ox. monolayer	0.2302	0.2281	0.2298

Table 4.17: Alpha helical radius of Sp-B_{1-25}. A value of 0.23 nm is expected for an alpha helix.

Monolayer	47 Å²	55 Å²	72 Å²
DPPC	0.1518	0.1541	0.1552
DPPC POPG	0.1524	0.1521	0.1549
DPPC PXPG	0.1535	0.1521	0.153
Mixed unox. monolayer	0.1535	0.1533	0.1571
Mixed ox. monoalyer	0.1488	0.1582	0.1549

Table 4.18: Rise per residue in Sp-B_{1-25}. A value of 0.15 nm is expected for an alpha helix.

4.3.6 Helicity

Looking at alpha helical properties alone only helps to give an indication as to whether Sp-B$_{1\text{-}25}$ remains in an alpha helical state. To supplement the analysis of the secondary structure of Sp-B$_{1\text{-}25}$, the helicity of Sp-B$_{1\text{-}25}$ is analysed. This is the percentage of time that each individual residue remains in the helix during the simulation, giving an indication of the amount of helical content the peptide has.

Figures 4.21 to 4.23 illustrate the average residue-wise helicities of each system at each measured APL, obtained from the last 150 ns of simulation for every amino acid residue found in the sequence of Sp-B$_{1\text{-}25}$. Table 4.19 illustrates the total percentage of residues that remained helical at least 80% of the time during the simulations, for all monolayers and APLs. In all monolayers, it was observed that the residues **LCRALIKR** always remained in the helix at least 80% of the time. In the DPPC-only monolayers, the percentage of residues that remain in the helix more than 80% of the time increased from 40-48% as the APL increases (Figure 4.21). In the DPPC:POPG (80:20 mol%) monolayer, 44% of residues remained helical at all measured APLs. In the oxidised DPPC:PXPG (80:20 mol%) monolayer, the percentage of helical residues decreased from 44 to 32% as APL increases (Figure 4.22). In the unoxidised mixed monolayers with POPG, 36-44% of the residues remained helical, and 40-48% of residues remained helical in the monolayers containing oxidised PXPG (Figure 4.23).

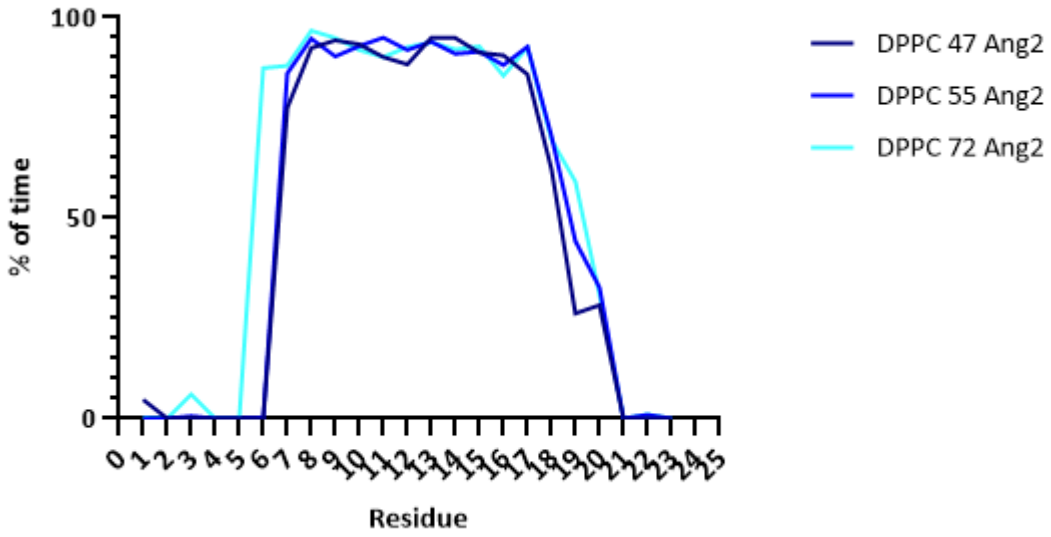


Figure 4.21: Average residue helicity of Sp-B$_{1\text{-}25}$ in the DPPC-only monolayers at three measured APLs.

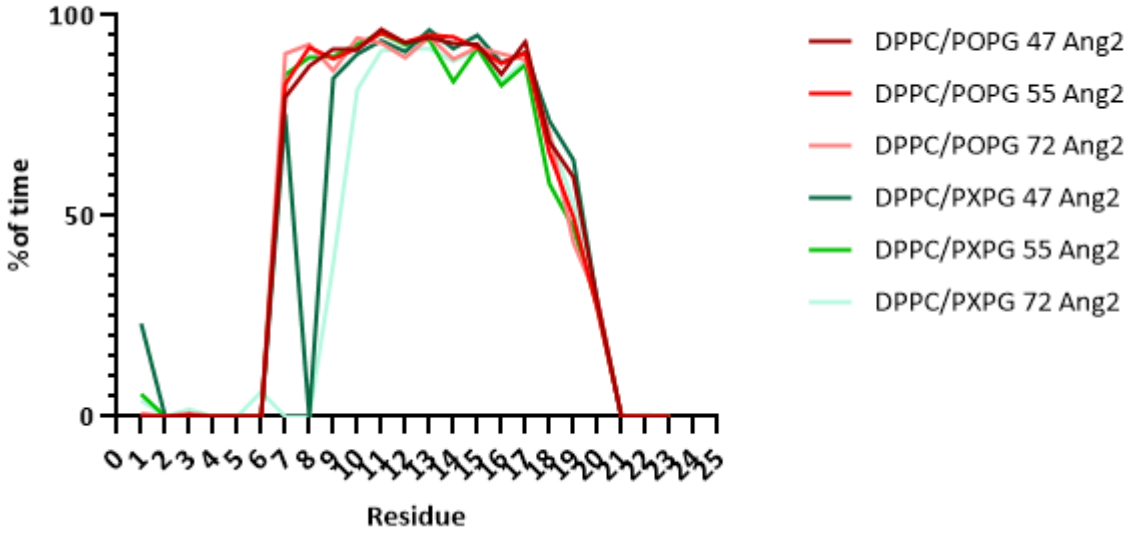


Figure 4.22: Average residue helicity of Sp-$B_{1\text{-}25}$ in the DPPC:POPG (80:20 mol%) and DPPC:PXPG (80:20 mol%) monolayers at three measured APLs.

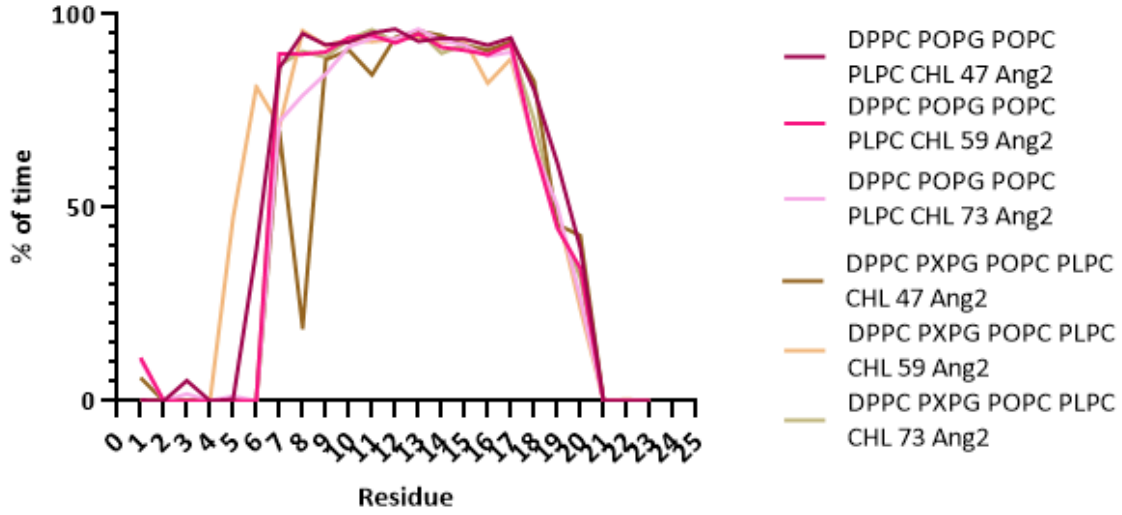


Figure 4.23: Average residue helicity of Sp-$B_{1\text{-}25}$ in the unoxidised and oxidised mixed monolayers at three measured APLs.

Monolayer	Residues	% of residues remaining helical >80% of the time
100% DPPC 47 Å^2	CWLCRALIKR	40%
100% DPPC 55 Å^2	YCWLCRALIKR	44%
100% DPPC 72 Å^2	PYCWLCRALIKR	48%
DPPC POPG (unoxidised) 47 Å^2	YCWLCRALIKR	44%
DPPC POPG (unoxidised) 55 Å^2	YCWLCRALIKR	44%
DPPC POPG (unoxidised) 72 Å^2	YCWLCRALIKR	44%
DPPC PXPG (oxidised) 47 Å^2	YCWLCRALIKR	44%
DPPC PXPG (oxidised) 55 Å^2	WLCRALIKR	36%
DPPC PXPG (oxidised) 72 Å^2	LCRALIKR	32%
Mixed (unoxidised) 47 Å^2	YCWLCRALIKR	44%
Mixed (unoxidised) 55 Å^2	YCWLCRALIKR	44%
Mixed (unoxidised) 73 Å^2	WLCRALIKR	36%
Mixed (oxidised) 47 Å^2	WLCRALIKRI	40%
Mixed (oxidised) 55 Å^2	P,CWLCRALIKRI	48%
Mixed (oxidised) 73 Å^2	YCWLCRALIKR	44%

Table 4.19: Total percentage of Sp-B$_{1-25}$ residues remaining in the alpha helix during the final 150 ns simulation.

4.3.7 Hydrogen bond analysis

To understand the behaviour of Sp-B$_{1-25}$ in the monolayers at different areas per lipids and how it interacts with the lipid headgroups and acyl chains, it is useful to study the hydrogen bonds formed between the residues on Sp-B$_{1-25}$ and the lipids. Figure 4.25 shows the lipids involved in hydrogen bond formation with Sp-B$_{1-25}$ at each measured APL for the DPPC monolayer (Figure 4.25, a-c) and DPPC/POPG (80:20 mol%) monolayer (Figure 4.25, d-f). Additionally, Table 4.20 shows the total number of hydrogen bonds formed between the lipids and peptides, and specifically which residues on Sp-B$_{1-25}$ are directly involved in bond formation with either DPPC or POPG lipids.

Generally, more hydrogen bonds form between the lipid and peptide in the mixed monolayer compared to the DPPC monolayer. In this monolayer, the total number of hydrogen bonds formed between Sp-B$_{1-25}$ and DPPC lipids increases when the monolayer expands from 47.4 Å^2 (6 bonds) to 71.9 Å^2 (12 bonds). In the mixed monolayer, the total number of hydrogen bonds formed is highest at 55.4 Å^2 (18 bonds), and lowest at 71.9 Å^2 (9 bonds). The positively charged residues on Sp-B$_{1-25}$ strongly interact with lipid headgroups and more hydrogen bonds seem to form with POPG lipids except for the

simulation at the largest APL. Figure 4.24 shows a snapshot of 2 POPG lipids forming intermolecular hydrogen bonds with Sp-B$_{1\text{-}25}$, with the residues Arginine 17 and Lysine 24 and Glycine 25, where arginine and lysine are positively charged residues.

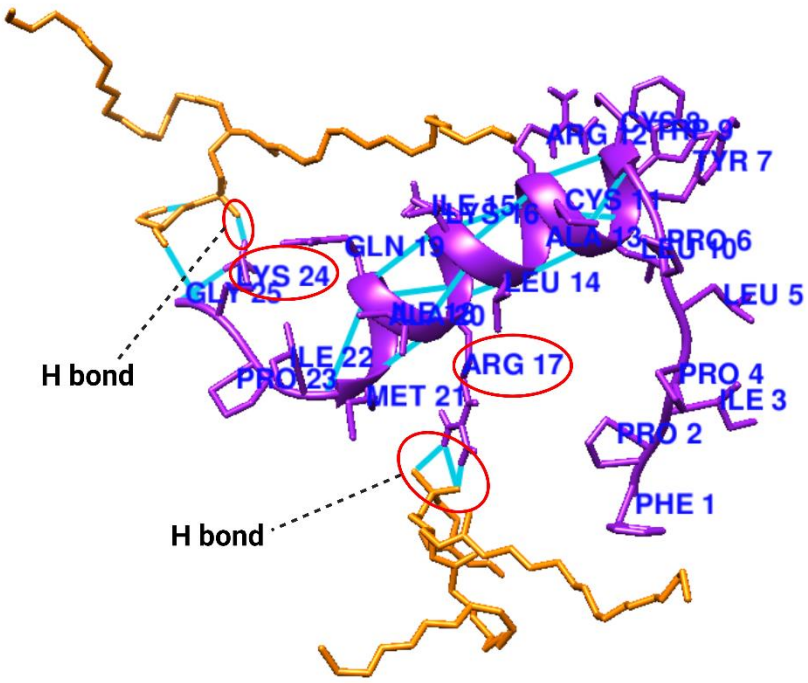


Figure 4.24: Zoomed in image of 2 POPG lipids (orange) forming hydrogen bond interactions (blue) with Sp-B$_{1\text{-}25}$ (purple). Positively charged residues that form hydrogen bonds with the peptide are circled in red.

Sp-B$_{1\text{-}25}$ consists of no negatively charged residues and 4 positively charged residues; lysine 24, lysine 16, arginine 12 and arginine 17. The positively charged residues strongly interact with the negatively charged phosphates on the lipid headgroups. In the DPPC monolayer, at 47.4 Å^2 33% of hydrogen bonds were formed with positively charged residues on the peptide, 62.5% at 55.4 Å^2, and 50% at 71.9 Å^2. In comparison, the peptide in the mixed monolayer form more hydrogen bonds with the lipid headgroups, 66.7% of hydrogen bonds were formed with positively charged residues at 47.4 Å^2, 77.8% at 55.4 Å^2, and 66.7% at 71.9 Å^2, suggesting that Sp-B$_{1\text{-}25}$ associates more closely with the headgroup region of the lipids in the presence of saturated and unsaturated lipids.

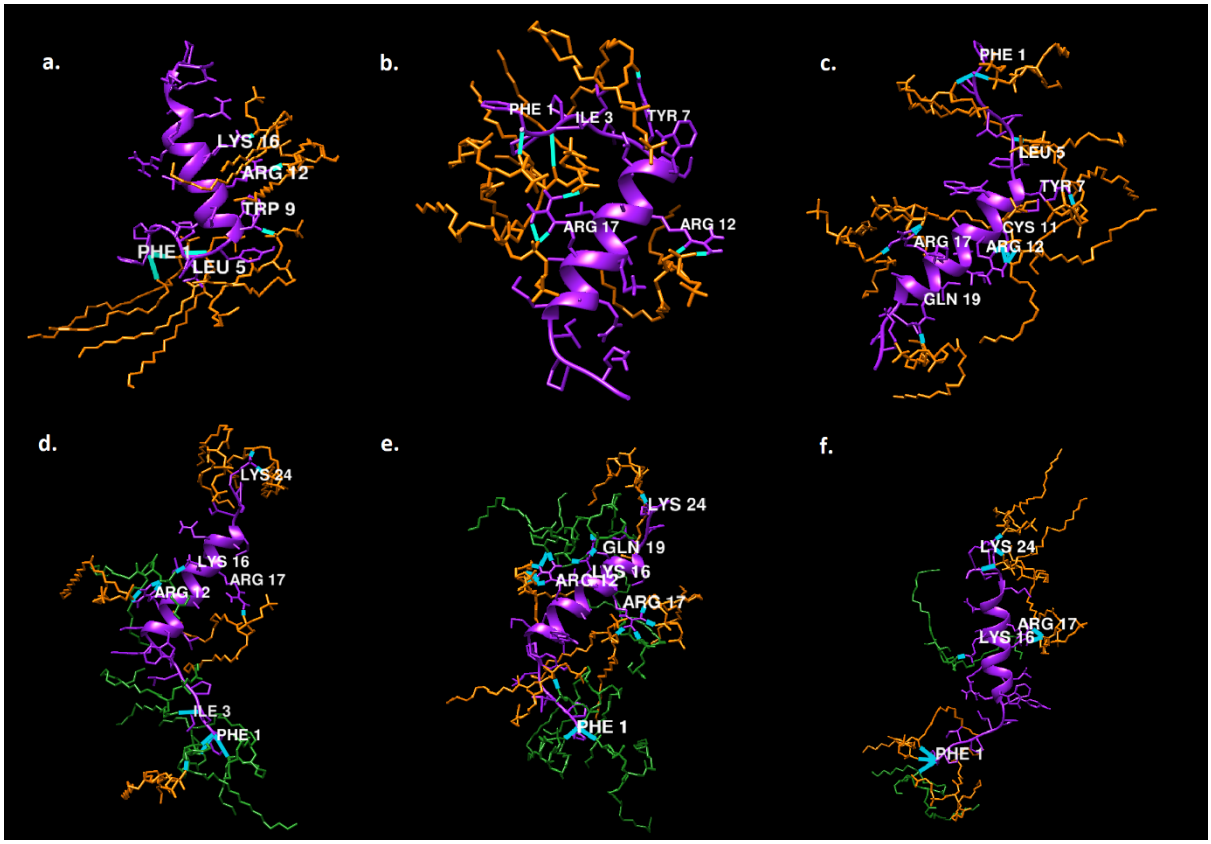


Figure 4.25: Hydrogen bonds between Sp-B_{1-25} and DPPC and/or POPG residues. Sp-B_{1-25} (purple), DPPC (orange), POPG (green), hydrogen bonds (blue). **a)** Sp-B_{1-25} in DPPC, 47.4 Å^2, **b)** Sp-B_{1-25} in DPPC, 55 Å^2, **c)** Sp-B_{1-25} in DPPC, 72 Å^2, **d)** Sp-B_{1-25} in DPPC/POPG, 47 Å^2, **e)** Sp-B_{1-25} in DPPC/POPG, 55 Å^2, **f)** Sp-B_{1-25} in DPPC/POPG, 72 Å^2. Structures visualised in Chimera were taken at the end of a 300 ns simulation.

Monolayer	**H bonds formed between Sp-B_{1-25} and lipid**	**Total H bonds**
DPPC 47 Å^2	PHE 1 with DPPC 200 (O14, O22) LEU 5 with DPPC 173 (O14) TRP 9 with DPPC 175 (O14) ARG 12 with DPPC 170 (O32) LYS 16 with DPPC 180 (O14)	6 Total of 5 lipids involved
DPPC 55 Å^2	PHE 1 with DPPC 222 (O14) ILE 3 with DPPC 215 (O22) TYR 7 with DPPC 202 (O32) ARG 12 with DPPC 170 (O11, O13) ARG 17 with DPPC 167 (O13, O13) ARG 17 with DPPC 215 (O11)	8 Total of 6 lipids involved
DPPC 72 Å^2	PHE 1 with DPPC 167 (O14)	12

	PHE 1 with DPPC 201 (O13) LEU 5 with DPPC 237 (O32) TYR 7 with DPPC 217 (O22) CYS 11 with DPPC 223 (O22) ARG 12 with DPPC 223 (O13, O13, O11) ARG 17 with DPPC 165 (O32, O32) ARG 17 with DPPC 252 (O32) GLN 19 with DPPC 236 (O22)	Total of 8 lipids involved
80:20 mol% DPPC/POPG 47 Å^2	PHE 1 with POPG 272 (O13, O22) PHE 1 with POPG 286 (OC2) ILE 3 with POPG 267 (O32) ARG 12 with DPPC 142 (O13, O13) ARG 12 with POPG 265 (O11 O11) LYS16 with POPG 265 (O13) ARG 17 with DPPC 138 (O13) LYS 24 with DPPC 141 (O13, O32)	12 Total of 8 lipids involved (5 POPG 3 DPPC)
80:20 mol% DPPC/POPG 55 Å^2	PHE 1 with POPG 268 (O13) PHE 1 with POPG 273 (O13) PHE 1 with POPG 282 (OC2) ARG 12 with DPPC 150 (O11, O13, O13, O21, O32) ARG 12 with POPG 275 (2 O32) LYS 16 with POPG 270 (O14) LYS 16 with POPG 275 (O14) ARG 17 with DPPC 136 (O11) ARG 17 with DPPC 167 (O14, O32) ARG 17 with POPG 269 (O14) GLN 19 with POPG 270 (O11) LYS 24 with DPPC 231 (O14)	18 Total of 10 lipids involved (4 DPPC 6 POPG)
80:20 mol% DPPC/POPG 72 Å^2	PHE 1 with DPPC 136 (O14, O22) PHE 1 with POPG 268 (O14) LYS16 with POPG 280 (O14) ARG17 with DPPC 235 (O22, O22) LYS 24 with DPPC 168 (O13, O14) LYS 24 with DPPC 200 (O14)	9 Total of 6 lipids involved (4 DPPC 2 POPG)

Table 4.20: Total number of hydrogen bonds formed between Sp-$B_{1\text{-}25}$ and the lipids.

4.4 Discussion

Simulations of the monolayers at the measured APLs were conducted to investigate the MD of surfactant lipids in an environment mimicking the inspiration and expiration of air in the lungs. The mixed monolayers are more physiologically relevant compared to the monolayers containing one or two lipid types. In all monolayers, it was seen that as the APL increases, the surface pressure of the monolayer decreases, which is in agreement with previous surface pressure-area isotherms obtained from Olżyńska *et al.* (Olżyńska et al., 2016).

The compressibility (k), shown in Equation 4.9, of a monolayer relates to the slope of the pressure-area isotherm, where A refers to the molecular area of the monolayer, π as the surface pressure, $(\frac{\partial A}{\partial \pi})$ representing the partial derivative of molecular area with respect to surface pressure and the subscript T showing that the partial derivative is taken at constant temperature. Compressibility is defined in terms of a decrease in area (APL) with an increase in surface pressure (ie. compression). Hence, a negative compressibility value shows that the monolayer is less compressible, requiring a higher pressure to be applied to achieve a change in APL.

$$k = -\frac{1}{A}\left(\frac{\partial A}{\partial \pi}\right)_T \qquad (4.9)$$

Surface pressure-area isotherms can discern whether the monolayers existed in a liquid condensed or expanded phase at the measured APLs. For instance, in the liquid condensed phase, an isotherm would exhibit a much steeper decrease in surface pressure as the APL increases. This indicates a more ordered and compact arrangement of lipids in the monolayer. In relation to compressibility, a steeper slope in the isotherm corresponds to a more negative compressibility, and a gradual slope corresponds to a less negative compressibility. This means that monolayers with a high compressibility are more loosely packed, allowing surface pressure to change easily in response to the APL. Low compressibility monolayers require larger changes in APL to significantly alter the surface pressure, suggesting that the lipid molecules follow a densely packed, ordered arrangement.

In the DPPC pressure-area isotherm, there was a steeper increase in surface pressure as the APL decreased from 55 to 47 Å^2, which may indicate the monolayer was in the liquid condensed phase as the lipids became more ordered and compact, and from 55 to 72 Å^2,

there was a more gradual increase in surface pressure which implies the lipids became more disordered and loosely packed. There was a much more gradual increase in surface pressure as the APL decreased from 72 to 47 $Å^2$, in the DPPC:PXPG compared to the DPPC:POPG monolayer. It appeared that the DPPC:PXPG monolayer had the highest compressibility out of all monolayers, and compressibility decreased when the PXPG lipids were replaced with POPG in the same mixture. This could be explained by the absence of the unsaturated carbon double bond in the PXPG lipids which affected the phase of the lipid monolayers. The presence of both saturated and unsaturated lipids in the monolayer resulted in better surfactant function, as the monolayer can withstand and reach a larger range of surface pressures as the APL changes (Figure 4.8). Meanwhile, the mixed monolayer with unoxidised lipids appeared to have the lowest compressibility, as significantly larger surface pressures must be applied to compress the monolayer down to 47 $Å^2$. In the case of large APLs beyond 80-100 $Å^2$, the simulated monolayers were seen to rupture, which led to negative surface pressures (Figure 4.10).

The order parameters give a quantitative measurement of how ordered the carbon atoms are on the phospholipid tails. Order parameter data (Figure 4.11-4.15) showed how structural changes in the tails of unsaturated lipids were determined by compression and expansion of the monolayer. In all systems, it was shown that the SN_1 and SN_2 lipid tails of DPPC became more disordered as the APL increases. The SN_2 tails on the unsaturated lipid, POPG, in the DPPC:POPG monolayer also became increasingly disordered with increasing APL, with the carbon atoms in the vicinity of the double bond being systematically more disordered compared to the other atoms. In the DPPC:PXPG monolayer, the absence of the carbon double bond on the SN_2 tail due to oxidative damage resulted in the PXPG tails being more disordered than the POPG tails simulated at the same APL. Similarly, this effect is observed on the SN_2 tails of the unsaturated lipids in the mixed monolayer. However, the order of DPPC lipid tails in the oxidised monolayer did not seem to be affected by the absence of the unsaturated bond on the SN_2 chain of PXPG.

At an APL of 47 $Å^2$, the carbon atoms on the SN_2 chain of PXPG are seen to be far more disordered than the other lipid tails in the mixed monolayer, where Carbons 3 to 7 of the SN_2 chain have negative S_{CD} values. Using Equation 4.7, a negative S_{CD} value implies that the angle between the carbon deuterium bond vector and the monolayer normal

(reference axis) is more than 90°. For instance, an S_{CD} value of 0.2 gives an angle of 82.3°, while a value of -0.05 would give an angle of 91.9°. This reveals information on the orientation of the lipid tails with respect to the normal. As shown in Figure 4.26, the angle of the SN_2 chain of PXPG with respect to the monolayer normal is far larger than the SN_2 chain of POPG, which reflects a higher level of disorder.

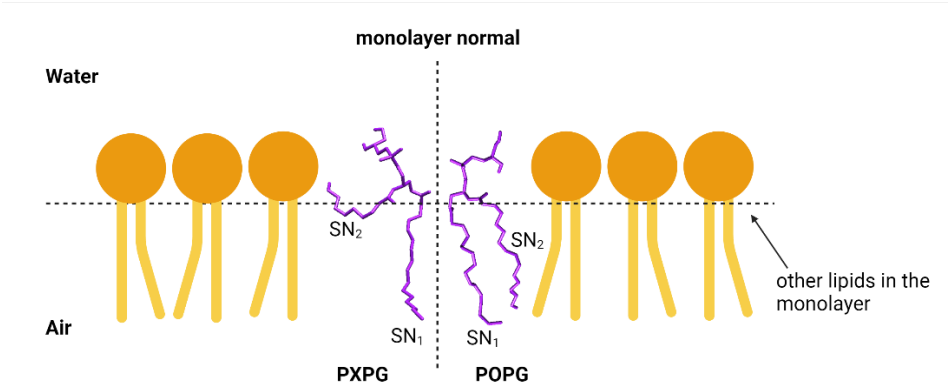


Figure 4.26: Diagram illustrating an example of how POPG and PXPG lipid tails may be positioned in the simulated monolayer. Created in Chimera and BioRender.

The density profiles of the monolayers were visualised to understand the positionings of the lipid headgroups, tails and peptide in the simulated systems. As the monolayer expands to a larger APL, the simulated system becomes less dense as seen from the decrease in size of the density plots, which agrees with the surface pressure-area isotherm data.

It was revealed that all monolayers were less dense and more spread out as the APL increases. At the lowest APL (most compressed state) in the monolayers, Sp-$B_{1\text{-}25}$ was seen to be positioned equally between the headgroups and acyl chain regions, and gradually the peptide becomes completely embedded in the headgroup region, towards the water region as the monolayer expands. Replacing 20% of DPPC in the monolayer with oxidatively damaged PXPG resulted in Sp-$B_{1\text{-}25}$ fully embedding itself in the headgroup region instead of associating with the lipid tails at the measured APLs. This is expected of amphipathic proteins like Sp-$B_{1\text{-}25}$; as seen in Figure 4.27, where one side of the helix and both ends of the helix contain mostly hydrophobic residues, with hydrophilic and amphipathic residues residing on another side of the helix.

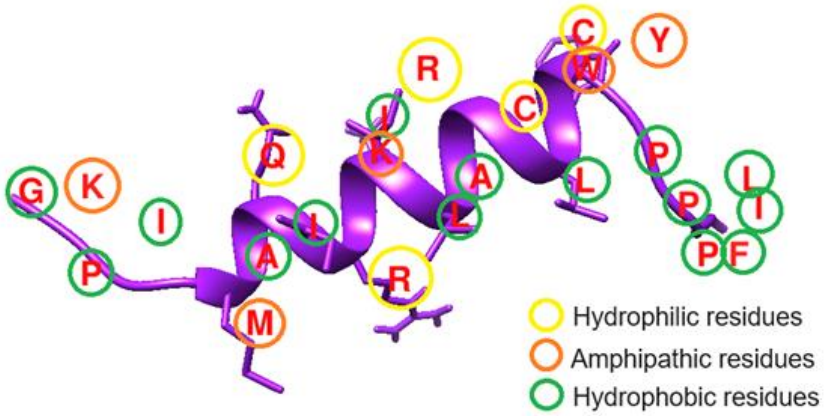


Figure 4.27: Hydrophilic, hydrophobic, and amphipathic residues found within the structure of Sp-B_{1-25}. Created in Chimera.

As seen in Figure 4.28, Sp-B_{1-25} is seen to situate itself between the acyl chains and headgroup region at higher APL (72 Å^2), while it positions closer towards the lipid acyl chains at lower APL (47 Å^2). This is in agreement with the partial density profile from Figure 4.17, and the predicted orientation of Sp-B_{1-25} in a lipid monolayer as shown in Figure 4.29. As the surface pressure decreases and the APL increases, the hydrophilic side of the Sp-B_{1-25} helix would be orientated towards the hydrophilic lipid headgroups, with the hydrophobic side of the helix orientated towards the lipid tails (Figure 4.29).

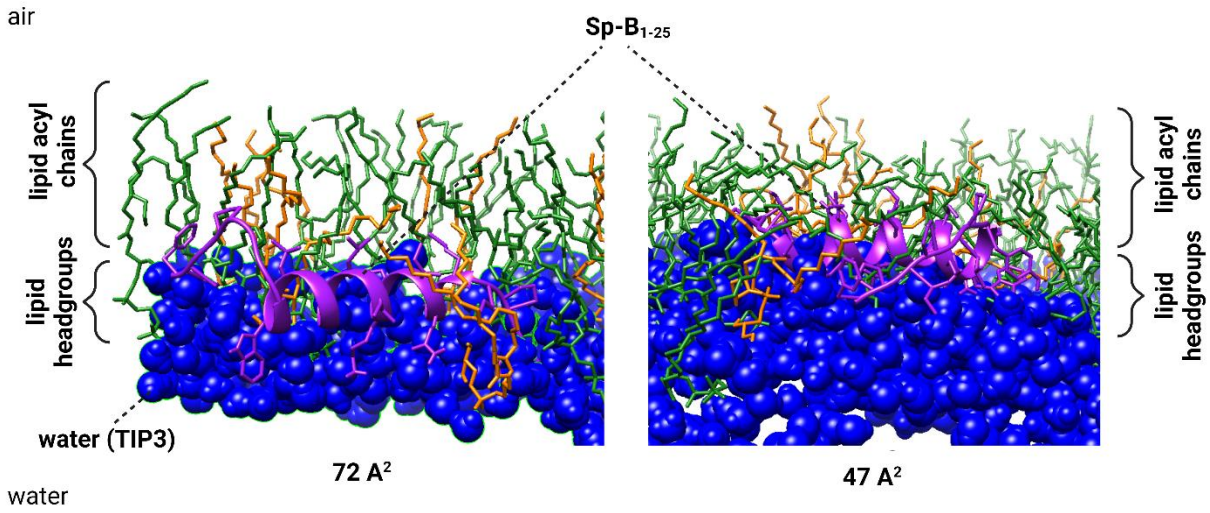


Figure 4.28: Snapshot of an MD simulation of Sp-B_{1-25} in DPPC:POPG monolayer at 72 and 47 Å^2. Sp-B_{1-25} is shown in purple, DPPC lipids are shown in green, POPG lipids in orange and water (TIP3) molecules shown in blue. Made using Chimera and BioRender.

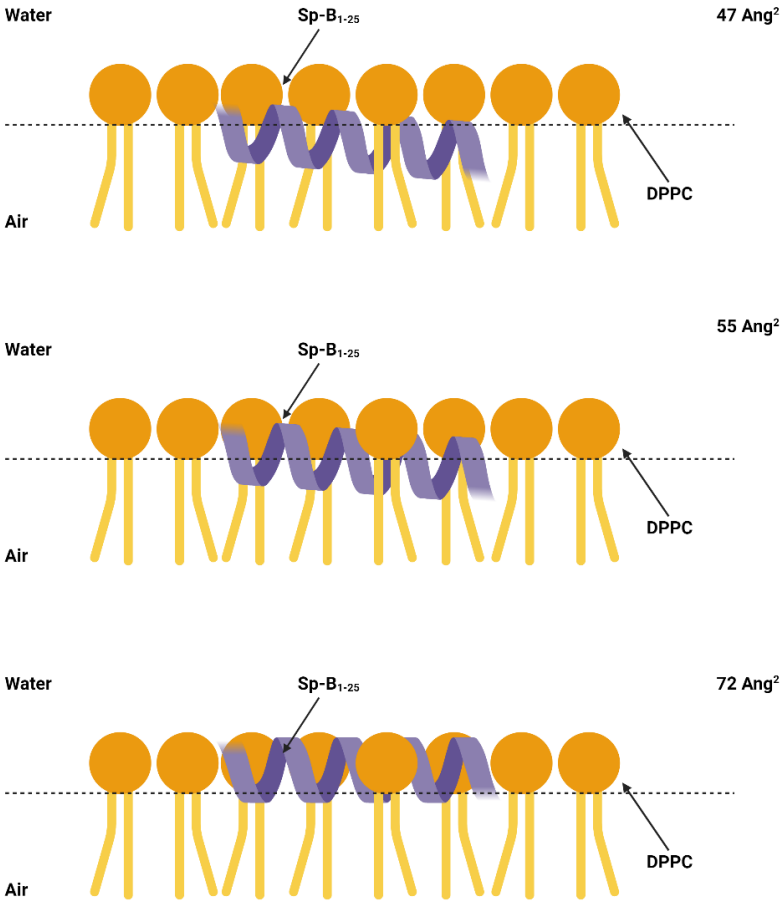


Figure 4.29: Predicted orientations of the Sp-B_{1-25} helix at various APLs, with respect to the lipid tails, headgroups and air-water interface. Created in BioRender.

A study by Kandasamy *et al.* performed MD simulations on Sp-B_{1-25} in DPPC monolayers, where it was discovered that the peptide helix orientates itself parallel to the air-water interface. It was proposed that the orientation of the peptide was largely determined by how favourable it was to form hydrogen bonds between Sp-B_{1-25} and charged residues on the DPPC headgroups (Kandasamy and Larson, 2005). An alpha helical structure is created from the formation of hydrogen bonds between the carbonyl oxygen (C=O) bond on a residue, with an amide hydrogen (N-H) bond on a residue that is four amino acids ahead in the protein sequence. Interactions with lipids in the surrounding environment will influence the structure of an alpha helix; for example, lipid tails have the propensity to form interactions with hydrophobic residues of the peptide sequence. The structure of Sp-B_{1-25} (1DFW) obtained from the PDB showed that residues 7 to 22 (Tyrosine 7 to Isoleucine 22) are contained within the helical structure of the peptide. Residues 1 to 6

(Phenylalanine 1 to Proline 6) and 23 to 25 (Proline 23 to Glycine 25) located at the N and C-terminals of the peptide were not in the helix (Figure 4.30).

The PDB structure of Sp-B_{1-25} generated by AlphaFold2 predictions showed that residues 5 to 21 (Leu5 to Met21) was part of an alpha helical structure, while the PDB structure of experimentally determined Sp-B_{1-25} revealed that residues 7 to 22 (Tyr7 to Ile22) was alpha helical, as seen in Figure 4.30. In all simulations, the C-terminal residues (Isoleucine 18, Glutamine 19, Alanine 20, Methionine 21, and Isoleucine 22) did not remain in the helix, possibly due to hydrophobic and hydrophilic interactions between the peptide and lipid. It was seen that in the DPPC-only monolayer, Proline 6 was part of the helix at 72 $Å^2$. In the unoxidised DPPC, DPPC:POPG and mixed monolayers, residues 7 to 17 remained in the helix at all APLs. In the oxidised DPPC:PXPG monolayer, residue 8 (Cysteine 8) was not in the helix at 47 and 72 $Å^2$ (Figure 4.22). In the unoxidised mixed monolayer, residue 8 was not in the helix at 47 $Å^2$ (Figure 4.23). Combined with data from Table 4.19, it confirms the significance of saturated and unsaturated lipid tails in maintaining the helical structure of Sp-B_{1-25}.

Oxidative damage to all unsaturated lipid tails results in the unwinding of the alpha helix at both N and C-terminal ends. Lipid oxidation of the unsaturated lipid tails causes the formation of polar aldehyde groups, dramatically altering the properties and structure of the lipids. Compared to Figure 4.30, which shows the residues that remain in the helical structure of native Sp-B_{1-25}, it was observed that Cysteine 8, a hydrophilic residue, possibly interacted with the aldehyde group on the oxidised PXPG lipid tail in the DPPC:PXPG monolayer, causing it to unwind from the alpha helix as pictured in Figure 4.30. This may affect the positioning of undamaged lipids; for instance, higher disorder in the SN_2 tails of oxidatively damaged unsaturated lipids could lead to residues in the helix to adjust their conformation to accommodate lipid tail positioning in the monolayers.

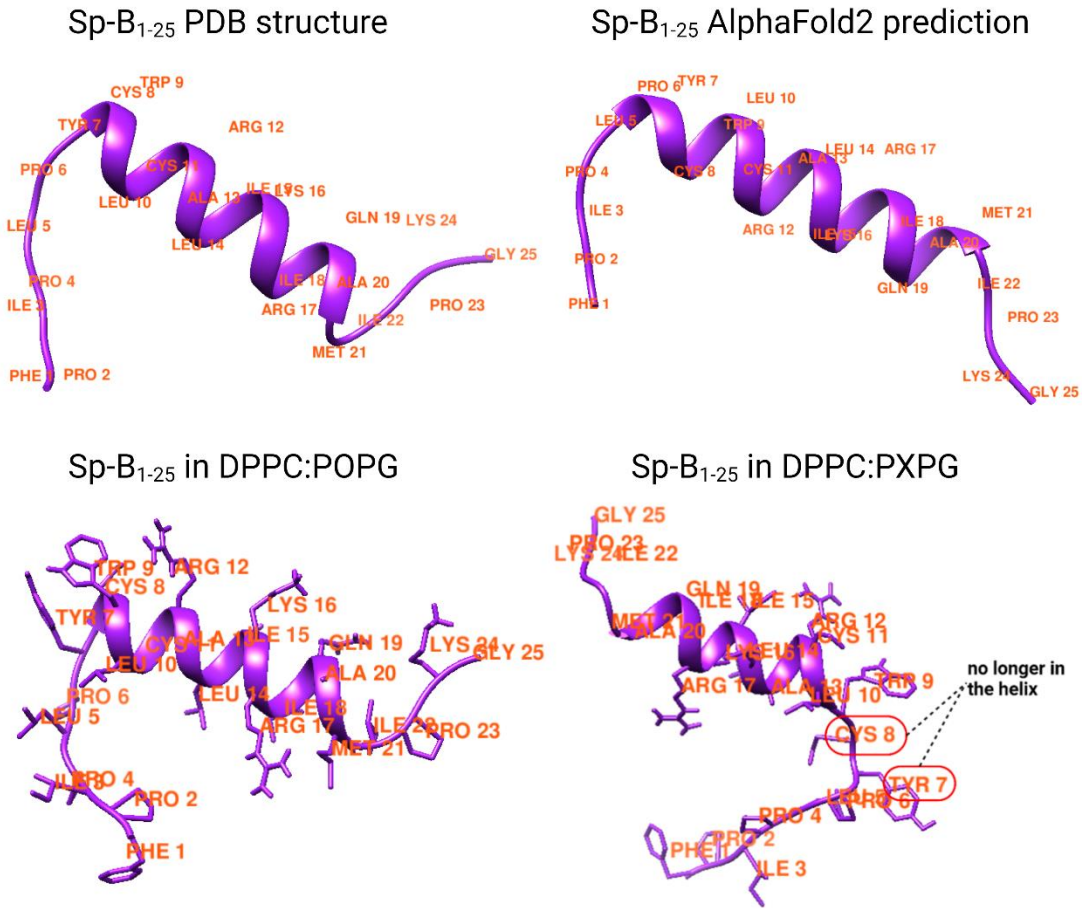


Figure 4.30: Experimentally determined, predicted, and simulated structures of Sp-B$_{1\text{-}25}$. **Top left:** Structure of experimentally determined Sp-B$_{1\text{-}25}$ obtained from the PDB. **Top right:** Structure of Sp-B$_{1\text{-}25}$ predicted using AlphaFold2. **Bottom left:** Structure of Sp-B$_{1\text{-}25}$ obtained from a snapshot of a simulation in the DPPC:POPG monolayer. **Bottom right:** Structure of Sp-B$_{1\text{-}25}$ obtained from a snapshot of a simulation in the DPPC:PXPG monolayer. Residues that are no longer a part of the alpha helix are circled in red (Cys8, Tyr7). Created in Chimera and BioRender.

4.5 Conclusions

This chapter aimed to simulate a truncated version of Sp-B, Sp-B_{1-25}, in monolayers containing varying compositions of surfactant lipids at different areas per lipid, in order to study surfactant protein-lipid interactions during expansion and compression of the lungs. It also investigated whether oxidative damage would affect the positioning of the lipid tails, headgroups and peptide.

Overall, the order parameter data of the monolayers agrees with data presented from previous MD simulations done on surfactant lipids such as DPPC and POPC (Huynh et al., 2014; Olżyńska et al., 2016). The membrane density data suggests that Sp-B_{1-25} positions itself parallel to the monolayer normal, where it situates between the lipid tails and headgroups at lower APLs, and more towards the headgroups at at higher APLs.

Understanding that the helicity of Sp-B_{1-25} changes within different lipid environments and APLs suggest that the structure of full-length Sp-B is highly influenced by the surrounding lipid environment. Oxidative damage to surfactant lipids would therefore impair the structure and function of Sp-B, since it is known that surfactant proteins Sp-B and Sp-C have a predominantly alpha helical structure which contributes to their biophysical function and the regulation of surfactant function in preventing alveolar collapse. The absence of cholesterol may have also caused the lipids to interact with the C-terminal residues on Sp-B_{1-25}, disrupting the structure of the alpha helix. Further work could include simulating the monolayer with cholesterol and other surfactant lipids, to understand cholesterol's role in maintaining monolayer fluidity and how it influences the positioning of other lipids.

It was discovered that complete oxidative damage to unsaturated lipid tails in a monolayer would result in decreased biophysical function, as the surfactant is unable to reach a wide range of surface pressures compared to the undamaged monolayers. However, this effect is greatly reduced if there are undamaged unsaturated lipids present.

Chapter 5 – Investigation of Protegrin (PG-1) in Lipid Monolayers at the Air-Water Interface with and without Oxidised Lipids using Neutron Reflection

5.1 Introduction

At the centre of an atom, the atomic nucleus contains protons, neutrons, and electrons. In an atom, positively charged protons determine the atomic number and chemical properties, while the uncharged neutrons determine the isotopic number or nuclear species of the atom. A neutron (n^0) is a subatomic particle with a mass of 1.674×10^{-27} kg (Tiesinga et al., 2019). Intrinsically, neutrons have a quantum mechanical property at the subatomic level known as spin, which influences their magnetic behaviour. Magnetic moment arises as a result of neutrons spinning which generate a magnetic field. Although neutron spin is independent and unaffected by the particle's orbital motion, it is directly proportional to the neutron's magnetic moment.

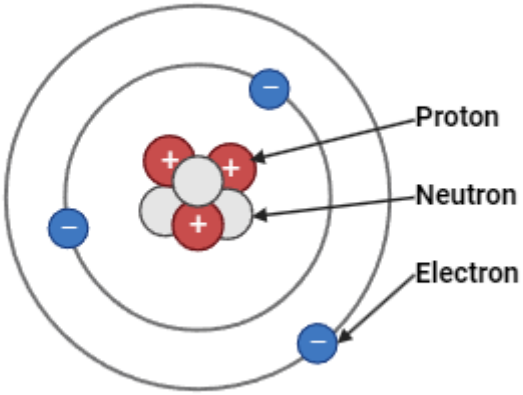


Figure 5.1: Composition of an atomic nucleus. Protons have a positive charge; electrons have a negative charge and neutrons have a neutral charge. Created in BioRender.

Fundamentally, the properties of neutrons play a crucial role in many areas of structural biology research. Free neutrons can be considered as particles or waves, which have an associated wavelength. For instance, when a neutron encounters the atomic nuclei of an atom, this results in neutron scattering due to the nuclear forces or interactions of magnetic moments between the two particles, which can be particularly useful in the study of membrane proteins. An important consequence of the fact that neutrons interact

with nuclei is reflected in their ability to distinguish isotopes via scattering cross sections, for example in isotopic substitution experiments involving hydrogen (^{1}H) and deuterium (^{2}H) (Young and Koppel, 1964). Since neutrons weakly interact with the atomic nucleus, they can easily penetrate many types of materials without destroying the samples, making them essential in the study of biological interfaces beyond the surface layers.

5.2 Principles of neutron reflectivity

When neutrons encounter an interface between two media of different neutron refractive indexes, reflection and/or refraction may occur. The technique of NR depends on measuring the intensity of the reflected neutron beam as a function of angle, incidence and/or wavelength. The extreme sensitivity of neutron beams to the structural and magnetic profiles of interfaces allows for the scattering length density (SLD) profiles of said interfaces to be obtained via a fitting process.

5.2.1 Neutron sources

Every NR experiment begins at a neutron source, which contains and produces an intense beam of neutrons to target the sample. Accelerator-based neutron sources typically rely on particle accelerators to accelerate a beam of charged particles to a high energy. Protons are generated in an ion source, and subsequently injected into the particle accelerator and further accelerated in a synchrotron via the usage of magnetic and electric fields in a circular path. This results in the formation of very high-energy particles, which can reach up to 800 MeV as seen in accelerators found in the ISIS Neutron and Muon Source. To optimally measure the transfer of energy and momentum between the neutrons and the sample, moderators containing high-hydrogen density substances are used to slow down these high-energy neutrons, which in turn provide better resolution in NR experiments (Anderson *et al.*, 2016).

Spallation neutron sources contain an accelerator which accelerate protons onto a heavy metal target, such as lead, mercury, or tungsten. This results in free neutrons “spalling” off upon impact with the nuclei of the heavy metal. Continuous spallation sources operate similarly to accelerator-based sources, where a steady stream of neutrons would be delivered to the target, making them suitable for certain spectroscopy and diffraction

experiments. One advantage of spallation neutron sources would be its ability to produce short and intense neutron pulses. Short-pulsed spallation neutron sources have a high-intensity neutron flux to provide short but high-resolution results for time-of-flight (TOF) experiments. Long-pulsed neutron sources fire lower-intensity but longer pulses of neutrons towards the target, to obtain a wider range of TOF results at a lower resolution (Bauer, 2001)

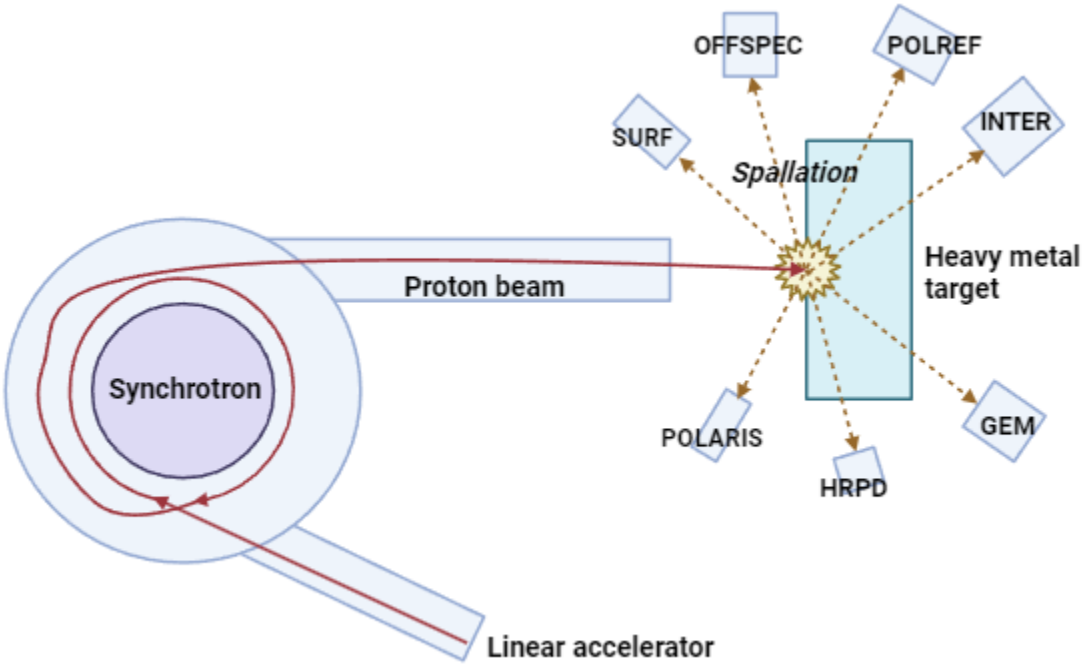


Figure 5.2: An example of a synchrotron-based spallation neutron source. The proton beam is accelerated in the synchrotron and directed at a heavy metal target, causing spallation of the protons to produce free neutrons, which are moderated before directed to various neutron diffractometers and reflectometers. Created using BioRender based on (Molina de la Peña, Calvo and F. Álvarez-Estrada, 2019) and instruments found at the ISIS Neutron and Muon Source.

5.2.2 Coherent and incoherent neutron scattering

Upon interacting with a sample, neutrons typically scatter via coherent or incoherent scattering. Generally, coherent scattering occurs when neutrons are scattered by a group of atoms in relation to their neighbouring atoms, maintaining the phase relationship between incident and scattered waves. In contrast, incoherently scattered neutrons carry information on single, individual atoms with no phase relationship between incident and

scattered waves, implying that scattering is uniform in all directions and independent of other events (Arbe et al., 2020).

The scattering length depends on the cross-section of individual atoms, which measures the strength of a particular interaction between a neutron and a target atom. As represented in the equation below, the total probability of interaction between the neutron and target atom is summed by the cross-section of scattered and incident neutrons that preserve the phase relationship ($\sigma_{coherent}$) and the cross-section of scattered and incident neutrons that disrupt the phase relationship ($\sigma_{incoherent}$).

$$\sigma_{total} = \sigma_{coherent} + \sigma_{incoherent} \qquad (5.1)$$

5.2.3 Scattering length density

The neutron SLD measures the extent of which a molecule can scatter neutrons, calculated from the ratio of the scattering length per molecule divided by the number of scatterers per unit volume. In the Equation below, ρ represents the SLD, where bi is the scattering length of atom type *i*, and *Ni* is the number of atoms of type *i* in the molecule. The summation is over all atom types in the molecule and *V* is the volume occupied by a single molecule.

$$\rho = \frac{\Sigma(bi \times Ni)}{V} \quad (5.2)$$

The SLD is essential in probing the composition and dynamics of thin lipid films and interfaces. An initial modelled SLD profile created using a Fourier transform produces a reflectivity curve, which is directly compared to the actual reflectivity data curve. Through an iterative fitting process, the initial curve and SLD profiles are modified until it matches the generated reflectivity curve based on a statistical goodness-of-fit and agreement. Typically, the SLD can be fitted using a series of layers, namely the structural layer thickness, roughness of the interface and the density (Arnold, 2013).

In this thesis, the reflected intensity was measured as a function of *q*, which depends on the wavelength and angle of incidence as written in Equation 5.3. This was subsequently compared to modelling profiles for different lipid interfaces, while varying the SLD and thicknesses of the layers, working out the best fit to the measured curves. The total SLD of the individual lipids were calculated using Equations 5.4 and 5.5. The SLD of the lipid

tails or heads were calculated by dividing the sum of all scattering lengths of the atoms in the tails or heads (b), divided by the volume of the component.

$$q_z = \frac{4\pi sin\theta}{\lambda} \quad (5.3)$$

$$SLD_{total} = \frac{\Sigma\,(biNi)_{total}}{Volume_{total}} \quad (5.4)$$

$$SLD_{total}\ \frac{\Sigma\,(biNi)_{total}}{Volume_{total}} = \frac{\Sigma\,(biNi)_{heads}+\Sigma\,(biNi)_{tails}}{Volume_{heads}+Volume_{tails}} \quad (5.5)$$

As ~80:20 mol% lipid mixtures were prepared, by mole, the scattering length of the average lipid in, for example, the dPOPG:POnPC (80:20 mol%) lipid film was calculated using Equation 5.6, where $x1$ is the mole fraction of lipid 1 (0.80) and $x2$ is the mole fraction of of lipid 2 (0.20).

$$SLD_{mixture} = \frac{\Sigma\,(biNi)_{total}}{Volume_{total}} = \frac{x1\Sigma\,(biNi)_{lipid\ 1}+x2\Sigma\,(biNi)_{lipid\ 2}}{x1Volume_{lipid1}+x2Volume_{lipid2}} \quad (5.6)$$

5.2.4 The significance of the critical angle of D_2O in neutron reflectivity

Raw data from NR is collected and plotted as a reflectivity curve, showing the neutrons reflected from the monolayer surface as a function of the momentum transfer perpendicular to the surface. The reflectivity or reflected intensity, was measured as a function of q, where λ is the neutron wavelength and Θ is the angle of incidence, as shown in Equation 5.3.

The critical angle is a specific angle of incidence arising from the refractive index of neutrons. The neutron refractive index relates to the SLD in Equation 5.7, where n is the refractive index, ρ is the SLD and λ is the neutron wavelength.

$$n = 1 - \frac{\lambda^2}{2\pi}\rho \quad (5.7)$$

For many materials, the SLD is positive, thus the refractive index is less than one. In this case, when the reflectivity at the interface reaches the critical angle given by $\cos\ \theta = n_2/n_1$, total external reflection occurs, where a "critical edge" is observed in the reflectivity profile below which the reflectivity is one. The position and shape of the critical angle reveals information on the roughness and thickness of the lipid films. Unlike most materials, including D_2O, the SLD of H_2O is negative. By adjusting the ratio of H_2O

and D_2O in the solvent through contrast variation, the SLD of the solvent can be matched to that of the sample, rendering it "invisible" to the neutron beam.

Neutrons reflected off different layers at an interface can interfere, leading to the appearance of fringes. The distance between the fringe periodicity in the momentum transfer space ($\boldsymbol{\Delta q}$) between consecutive fringes is determined by thickness of the lipid film (D) in equation 5.8, where a shorter distance correlates to a thicker lipid film (Figure 5.3). Furthermore, the uniformity of the fringe spacing, and intensity also suggest a uniform thickness and interface roughness in the lipid film. A loss of coherence in the reflected neutrons due to the presence of rough interfaces can cause dampening of the fringes due to path differences between those neutrons (Squires, 2012).

$$D = \frac{2\pi}{\boldsymbol{\Delta q}} \quad (5.8)$$

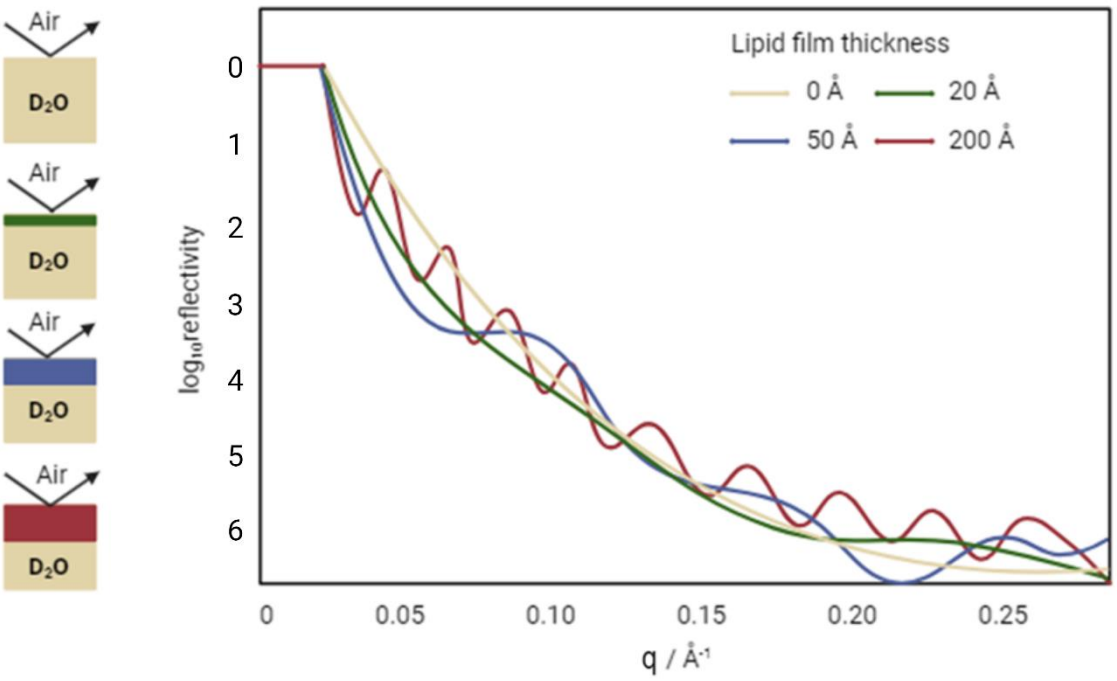


Figure 5.3: NR curve of pure D_2O at the air-liquid interface at various lipid film thicknesses. The critical edge is at $\log_{10}\text{reflectivity}(1) = 0$. Created in BioRender based on (Cousin and Fadda, 2020).

5.2.5 Applications of contrast variation in solvents

The presence of different isotopes can significantly impact the SLD value, as seen in hydrogen-deuterium substitution experiments. This utilises their difference in scattering lengths, where hydrogen possesses a negative scattering length and deuterium has a positive scattering length. As a result, this allows for contrast variation, where adjusting the composition of hydrogen and deuterium present in a buffer can emphasise or hide certain components within a sample.

The SLD of H_2O and D_2O have been recorded as -0.55×10^{-6} Å^{-2} and 6.35×10^{-6} Å^{-2} respectively (Heenan, 2010). By calculating the volume ratio, VR, combining 92% H_2O with 8% D_2O results in a SLD of 0 Å^{-2}, which makes null-reflecting water (NRW) (Equation 5.9). Contrast matching the amount of deuteration in a surfactant sample for instance, allows us to resolve the scattering of the surfactant layer without the subphase containing only H_2O (Figure 5.4).

$$(SLD\ H_2O\ \times\ VR\ H_2O) + (SLD\ H_2O\ \times\ VR\ H_2O) = SLD\ NRW \quad (5.9)$$

$$\left(-0.55\ \times\ 10^{-6}\ \times\ \frac{92}{100}\right) + \left(6.35\ \times\ 10^{-6}\ \times\ \frac{8}{100}\right) =\ 0\ \text{Å}^{-2}$$

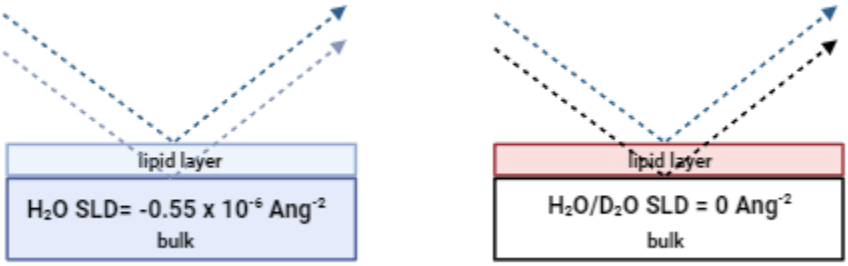


Figure 5.4: Applications of contrast variation in the bulk subphase. Created in BioRender. **Left:** Lipid layer is unresolved from the bulk. **Right:** A volume ratio of 92:8 H_2O to D_2O resolves the SLD of the lipid layer from the bulk subphase.

5.2.6 Setting up the neutron reflectivity instrument

In this thesis, the high-intensity chemical interface neutron reflectometer (INTER) was used to study the behaviour of lipid films with and without the peptide protegrin (PG-1), in D_2O and NRW-based buffers. Figure 5.5 illustrates how the beamline is set up for INTER, with the horizontal black line being the incident beam and the red arrow

representing the reflected beam. A beam of free neutrons generated from the source passes through a series of low-efficiency detectors (monitors), which scale and normalise the signal. The beam also passes through two slits, which collimate the beam to set the angular divergence and resolution, and to optimise the illumination area on the sample. The neutron beam then reaches the sample horizon, at an incident angle θ_i, represented by the dotted black line with respect to the incident beam. A high-efficiency detector, located at an angle of $2\theta_{det}$, records the final scattering angle, θ_f, between the reflected beam and sample horizon (Arnold, 2013).

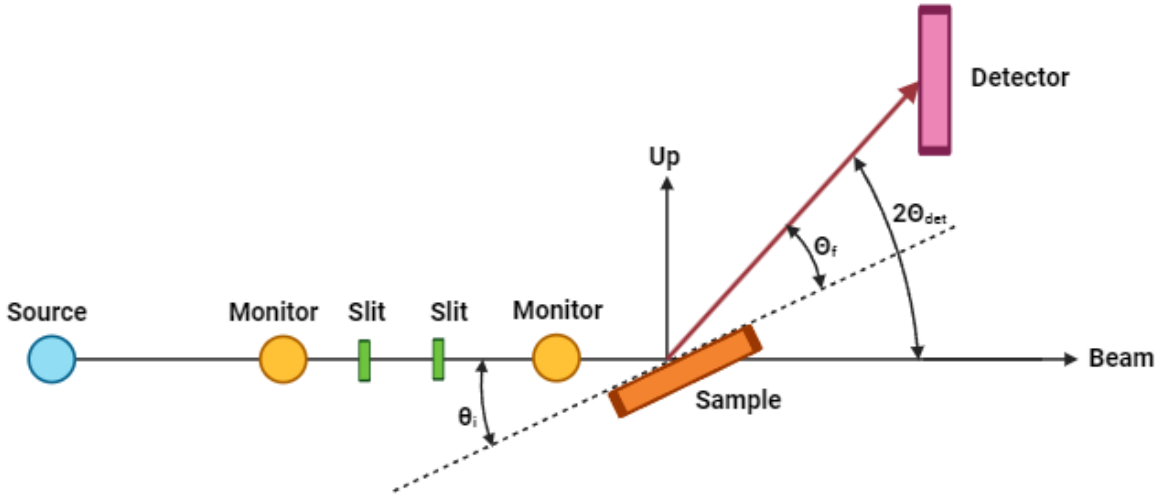


Figure 5.5: Schematic diagram of an INTER reflectometer typically encountered at ISIS. Source of neutrons (blue) is directed in a beam through the monitors (yellow) and slits (green), reaching the sample horizon (orange) where the reflected beam (red) then reaches the detector (pink). Created in BioRender.

5.3 The Langmuir trough

The surface tension and NR of the lipids were simultaneously measured on a Langmuir trough. This was done by forming a monolayer of lipids on an aqueous subphase, containing either D_2O or NRW. The lipid monolayer is contained between the barriers and are orientated with the hydrophilic headgroups facing the subphase and hydrophobic tails facing the air. PG-1 was injected into all four corners of the trough, in the subphase right below the barriers.

As seen in Figure 5.6, the trough barriers can be adjusted to go in either direction to compress or expand the monolayer. The surface pressure (Π), is calculated from the difference between the surface tension, (γ_0) and with (γ) in the lipid monolayer present

(Equation 5.10). In other words, γ_0 also refers to the surface tension of the subphase alone. The Wilhelmy plate is situated hanging from one of the barriers, partially dipped into the subphase, allowing the measurement of changes in surface pressure exerted by the lipid monolayer against the subphase. The plate is also attached to a sensor, which measures the force exerted and converts it into surface tension, measured in $mN\ m^{-1}$.

$$\Pi = \gamma_0 - \gamma \quad (5.10)$$

Compressing the barriers in the Langmuir trough causes a reduction of surface area of the lipid monolayer and an increase in density of lipids per unit area. This means that there is an increase in surface pressure, and a decrease in surface tension at the monolayer interface. Surface pressure area isotherms are plotted against the area per molecule as the barriers expand or compress. Such changes in the surface pressure provide information on any changes in interactions, orientations and properties of the lipids and PG-1, in relation to reflectivity.

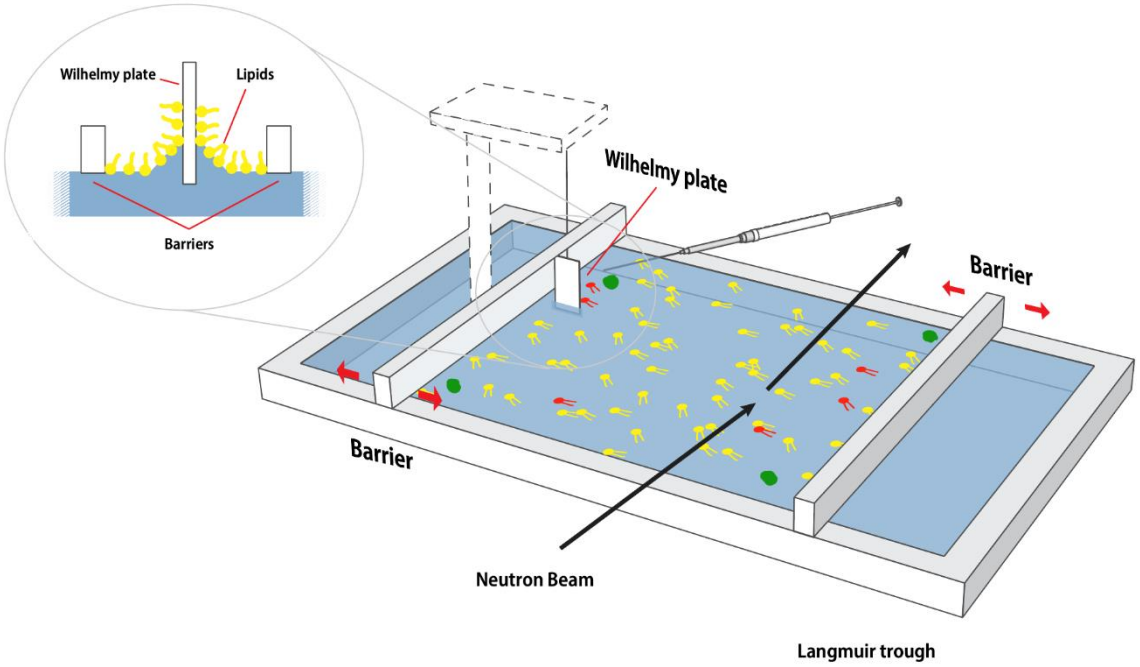


Figure 5.6: Diagram of a Langmuir trough. A lipid monolayer (yellow, red) is formed on the aqueous subphase (blue) with PG-1 (green) added under the four corners of the trough. The Wilhelmy plate is dipped into the subphase and attached to a sensor that detects changes in forces exerted. The reflectivity can be simultaneously measured as the incident beam is directed at the monolayer between the barriers.

5.4 Interaction of anti-microbial peptides with lipids

Anti-microbial peptides (AMPs) are small proteins that are found in the innate immune system of mammals (Ishitsuka et al., 2006). They were first discovered in porcine leukocytes and primarily serve to protect the host against microbial cells such as bacteria, fungi, and viruses. Bacterial cell membranes contain large amounts of anionic PG lipids, whereas mammalian cell membranes contain more PC lipids. This difference allows AMPs to preferentially interact with bacterial cell membranes. Furthermore, their amphiphilic nature aids insertion into the microbial cell membranes, causing an increase in membrane permeability, eventually leading to cell lysis and death.

In a study by Gidalevitz *et al.*, it was found that PG-1 inserts into anionic DPPG, POPG and lipid A films (Gidalevitz et al., 2003). However, it inserts significantly less into monolayers containing zwitterionic DPPC, POPC and di-palmitoyl-phosphatidylethanolamine (DPPE) lipids. The level of PG-1 insertion was monitored in Langmuir trough experiments with model lipid membranes by measuring changes in surface area, and it was shown that PG-1 inserts significantly more into membranes containing negatively charged PG lipids compared to zwitterionic PC lipids. It was concluded that PG-1 completely disturbed the structure of DPPG, significantly disordered the structure of DPPC, and marginally affected the structure of DPPE. PG-1 also led to the complete destabilisation of Lipid A films, which is a crucial component of the outer membrane of Gram-negative bacteria.

In this chapter, PG-1 was used to investigate the effects of AMPs on models of bacterial lipid monolayers. PG-1 is an 18-amino acid residue peptide containing a high proportion of arginine and cysteine residues. It forms a β sheet loop structure, stabilised by disulfide bonds between the cysteine residues (Figure 5.7).

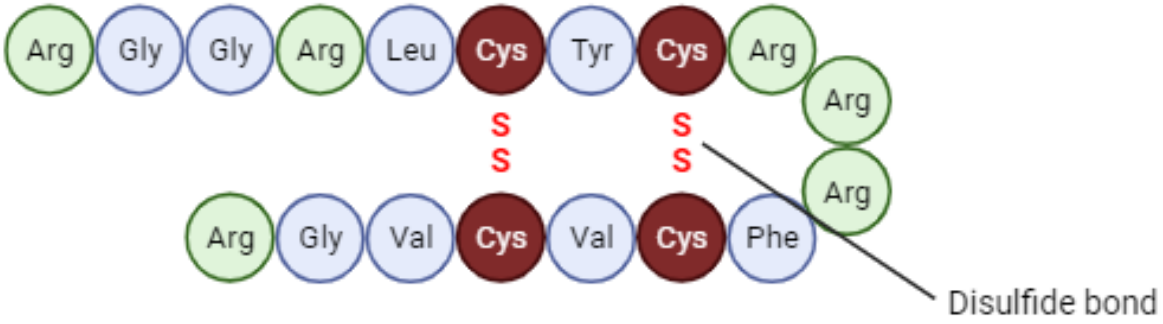


Figure 5.7: Structure of PG-1. Cationic residues are highlighted in green. Created in BioRender based on (Yasin et al., 1996).

Changes in the reflectivity profile can provide insight into how PG-1 disrupts the modelled lipid monolayers, and how deeply they penetrate the modelled cell membrane. The lipids used in this experiment were lipids typically found in bacterial cell membranes, such as deuterated DPPG (d_{62}DPPG), POPC and POPG. In more realistic models, lipids in both microbial and mammalian cells are also subjected to continuous oxidative attack, due to external environmental factors. Hence, oxidised POPC (POnPC) was used to explore the effects of oxidised lipids within the modelled cell monolayers, and how oxidation changes the extent of PG-1 interacting with the lipid monolayers.

5.5 Methods

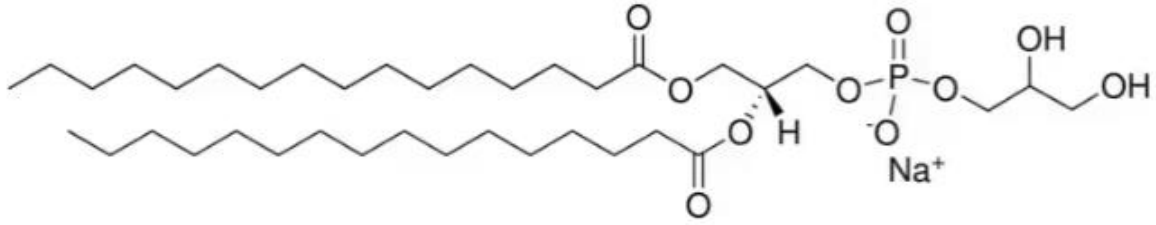
5.5.1 Materials

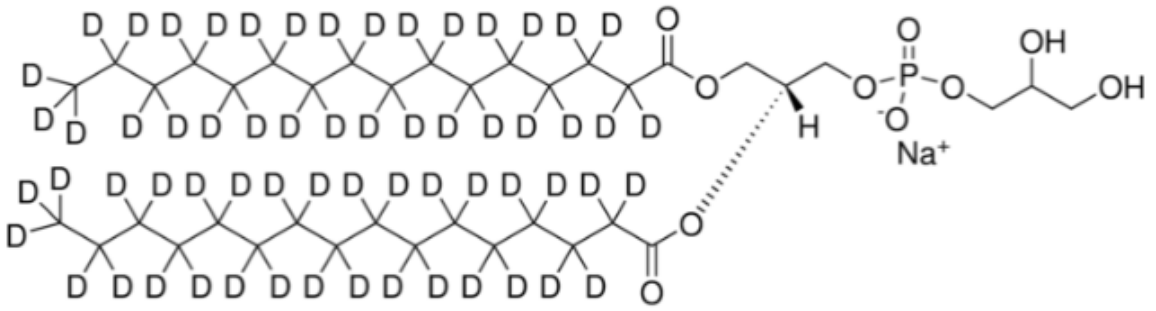
PG-1 was synthesised by Peptide Synthetics (Peptide Protein Research Ltd), using automated synthesis with a purity of >95%, determined by high-performance reversed-phase high performance liquid chromatography (RP-HPLC). The chemicals used in this thesis are recorded in Table 5.1, and Figure 5.8 shows the chemical structure of all the lipids used.

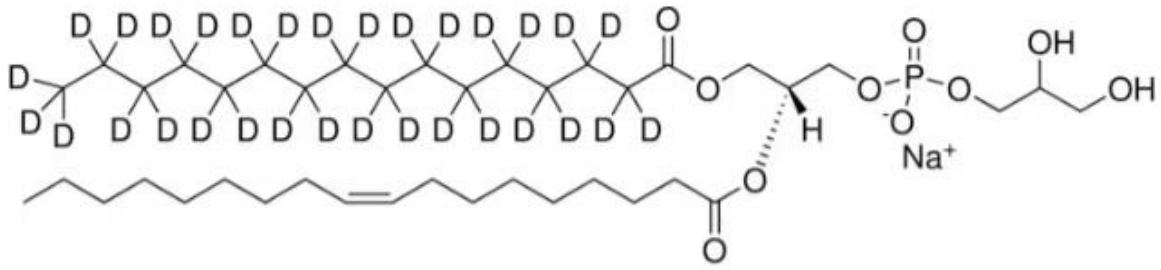
The lipids used were purchased from Avanti Polar Lipids. The carbons on both acyl chain tails of d_{62}DPPG are bonded to deuterium rather than hydrogen, with a total of 62 deuteriums in its structure. d_{31}POPG contain 31 deuteriums, all of which are located only on the saturated acyl chain, and 33 hydrogens on the unsaturated lipid tail.

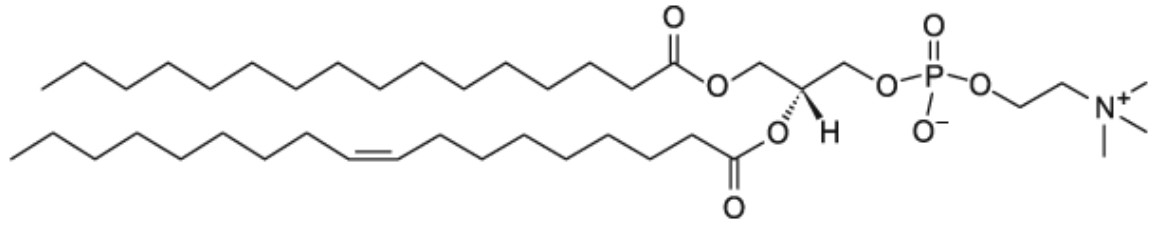
Compound	Manufacturer
PG-1 (≥ 95%)	Peptide Synthetics
dDPPG (> 99%, CAS 200880-41-7)	Avanti Polar Lipids
d_{62}DPPG (> 99%, CAS 326495-46-9)	Avanti Polar Lipids
d_{31}POPG (> 99%, CAS 327178-87-0)	Avanti Polar Lipids
POPC (> 99%, CAS 26853-31-6)	Avanti Polar Lipids
POnPC (> 99%, CAS 135726-46-4)	Avanti Polar Lipids
Chloroform (99-99.4%, CAS 67-66-3)	Honeywell
Methanol (99.9%, CAS 67-56-1)	ThermoScientific Chemicals
Oxoid™ Phosphate Buffered Saline Tablets	ThermoFisher

Table 5.1: Compounds used in NR experiments.









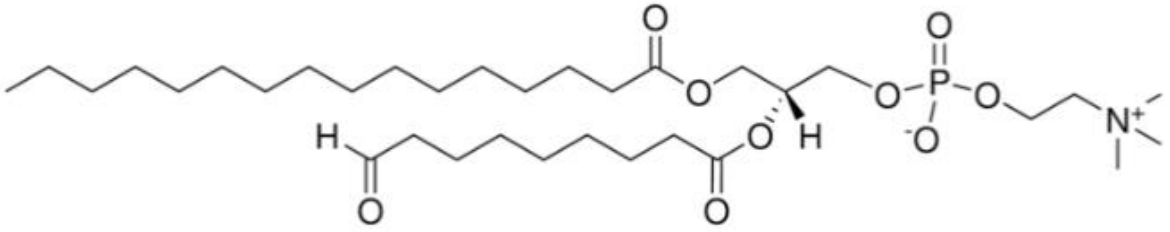


Figure 5.8: Structure of lipids used in NR experiments. **Top to bottom:** 1,2-dipalmitoyl-sn-glycero-3-phospho-(1′-rac-glycerol) (sodium salt) **(DPPG)**, 1,2-dipalmitoyl-d62-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt) **(d_{62}DPPG)**, 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-[phospho-rac-(1-glycerol)] (sodium salt) **(d_{31}POPG)**, 1-hexadecanoyl-2--(9Z-octadecenoyl)-sn-glycero-3-phosphocholine **(POPC)**, 1-palmitoyl-2-(9′-oxo-nonanoyl)-sn-glycero-3-phosphocholine **(POnPC)**. Obtained from Merck Pharmaceuticals.

5.5.2 Measuring the surface pressure of the monolayers

In all experiments, the subphases were prepared using Oxoid™ Phosphate Buffered Saline, PBS, Tablets (Dulbecco A) (ThermoFisher), with 1 tablet dissolved per 100 mL of H_2O/D_2O solution to give pH 7.3. This solution contained 8 g L^{-1} sodium chloride (NaCl), 0.2 g L^{-1} potassium chloride (KCl), 1.15 g L^{-1} disodium hydrogen phosphate (Na_2HPO4) and 0.2 g L^{-1} monopotassium phosphate (KH_2PO_4) (Dulbecco and Vogt, 1954). A pure D_2O-based buffer subphase was prepared, as well as a NRW-based buffer containing approximately 91.9% H_2O and 8.1% D_2O by mass. For example, 1 L of NRW-based buffer was prepared by dissolving 10 PBS tablets in 89.7g of D_2O, and making the volume up to 1 L.

NR experiments were conducted at the ISIS Neutron and Muon Source in the STFC Rutherford Appleton Laboratory, using the INTER neutron reflectometer beamline. The neutron beam was run at a low and high angle of incidence of 0.8 and 2.3˚. The experiments were made in a Teflon Langmuir trough (Nima Technology), with calibrated barrier positions, where 465 mL of aqueous subphase was added each time with the temperature maintained at 25˚C. Chromatography-grade filter paper was used as the Wilhelmy plate, which attaches to a sensitive microbalance that measured the difference in surface tension between the interface of the subphase and the lipid monolayer ie. the surface pressure.

Prior to measuring the reflectivity of the lipid monolayers, the cleanliness of the subphase surface is checked. The Langmuir trough was filled with D_2O-based subphase, and the barriers were closed to compress the trough area to a minimum. If the surface pressure fluctuated beyond 1 mN m^{-1}, the surface was aspirated to remove impurities.

Vials containing 1 mg of peptide PG-1 was dissolved in 1 mL of either D_2O or NRW-based buffer to make a peptide solution of 1 mg mL^{-1}. Lipid solutions of DPPG ($C_{38}H_{74}O_{10}PNa$)

and d_{62}DPPG ($C_{38}H_{12}O_{10}PNaD_{62}$) was made up to 1 mg mL^{-1} in a 2:1 v/v chloroform:methanol solution. d_{31}POPG ($C_{40}H_{45}O_{10}PNaD_{31}$), POPC ($C_{42}H_{82}NO_8P$) and POnPC ($C_{33}H_{64}NO_9P$) were made up to 1 mg mL^{-1} in chloroform stabilised with ethanol. The ratio of mixed lipids prepared are shown in Table 5.2.

The reflectivity of a pure D_2O subphase was recorded in the trough as a control, and to find the scale factor. For each dataset, the reflectivity profiles were recorded from increasing pressures 10 or 15 mN m^{-1}, 20 mN m^{-1}, 35 mN m^{-1} and subsequently down to 20 mN m^{-1} then 10 or 15 mN m^{-1} to simulate the compression and expansion of the surfactant monolayer.

The barriers were moved to the maximum trough area, and the lipid monolayers were formed by using a Hamilton syringe to add droplets of the prepared lipid solutions to the surface of the subphase, until the surface pressure reached approximately 5 mN m^{-1}. The volume of lipid monolayer mixtures added to the trough are recorded in Table 5.2.

A mixture of 80:20 mol:mol dDPPG:POPC was prepared by mixing 400 µL of 1 mg mL^{-1} (1.24×10^{-3} M) dDPPG with 94 µL of 1 mg mL^{-1} (1.13×10^{-3} M) POPC. Similarly, an 80:20 mol:mol of dDPPG:POnPC was prepared by mixing 400 µL of 1 mg mL^{-1} (1.24×10^{-3} M) dDPPG with 80 µL of 1 mg mL^{-1} (1.54×10^{-3} M) POnPC. Next, an 80:20 mol:mol of d_{31}POPG:POPC was made by mixing 400 µL of 1 mg mL^{-1} ($1.25\ 10^{-3}$ M) d_{31}POPG with 94 µL of 1 mg mL^{-1} (1.13×10^{-3} M) POPC. Finally, a mixture of 80:20 mol:mol d_{31}POPG:POnPC was prepared by mixing 400 µL of 1 mg mL^{-1} ($1.25\ 10^{-3}$ M) d_{31}POPG with 94 µL of 1 mg mL^{-1} ($1.31\ 10^{-3}$ M) POnPC.

After the reflectivity of the lipid monolayers were measured, 20 µL of 1 mg mL^{-1} PG-1 solution was added to each of the four corners of the trough (80 µL in total) underneath the barriers in all runs, achieving a concentration of 2.3×10^{-4} mg mL^{-1}. The peptide was incubated for 30 minutes before the measurement. Each reflectivity profile took approximately 40 minutes per session, with repeated measurements taking 3.5 hours in total.

Monolayer	Buffer	Volume of buffer added (mL)	Volume of lipid added (µL)	Volume of PG-1 added (µL)
d_{62}DPPG	D_2O	347	160	80
d_{62}DPPG	NRW	347	160	80
d_{62}DPPG:POPC (80:20 mol%)	NRW	347	180	80
d_{62}DPPG:POnPC (80:20 mol%)	NRW	347	160	80
d_{31}POPG	NRW	347	110	80
d_{31}POPG:POPC (80:20 mol%)	NRW	347	90	80
d_{31}POPG:POPC (80:20 mol%)	D_2O	347	103	80
d_{31}POPG:POnPC (80:20 mol%)	NRW	347	120	80
d_{31}POPG:POnPC (80:20 mol%)	D_2O	347	106	80

Table 5.2: Volumes of lipids and PG-1 added to the trough in D_2O or NRW-based buffer.

5.5.3 Data fitting

The initial model based on a hypothesised structure of the monolayer is generated from the reflectivity data, separated into layers or slabs with its own thickness, SLD and interface roughness. In this thesis, the data was fitted using the programs *refnx* (Nelson and Prescott, 2019) and *motofit* (Nelson, 2006). These programs calculate the reflectivity profile from the initial model of each dataset based on the Abeles matrix method, which is a computational technique that calculates the reflectivity of neutrons through thin films. Refnx is a python-based program that creates multiple layered models containing its own thickness, SLD and roughness parameters, and infers the probability distribution of the model parameters using a Bayesian approach. Motofit is an integrated fitting program that models multilayered structures with their own defined parameters. These models are fitted with a genetic algorithm, which can handle contrast variation data when different subphases are used.

For the monolayers where the reflectivity of the monolayers was recorded in both NRW and D_2O, the data for the monolayers in NRW-based buffer was fitted alone to determine the SLD of the tails to be used in the second fitting with both NRW and D_2O. The background was adjusted to 1.76×10^{-5}, dq resolution set at 5.5%, roughness set to 1.5 Å and SLD backing set to 0. The heads and tails volume fraction solvent (VFS) parameters

were linked and fitted, and the thicknesses of the heads and tails were fitted. In the second fitting, the parameters for the heads and tails thicknesses, SLD of the tails, and VFS of the heads were linked between the NRW and D_2O datasets. The background for the D_2O dataset was set to 1.07×10^{-5}, and the SLD backing set to 6.18×10^{-5} $Å^{-2}$. The background, dq resolution, roughness and SLD backing values for the NRW dataset remained the same as the first fitting. The value for SLD of the tails and the VFS of the heads were obtained from the first fitting with NRW alone. In the second fitting, only the thicknesses of the lipid heads and tails were fitted.

If the reflectivity curves of a monolayer were only recorded in NRW, there was only one fitting. The background was adjusted to 1.76×10^{-5}, dq resolution set at 5.5%, roughness set to 1.5 Å and SLD backing set to 0. The heads and tails VFS parameters were linked and fitted to give the VFS, while the heads and tails thicknesses were fitted.

5.6 Results

To investigate how oxidative damage may affect the insertion of the antimicrobial peptide, PG-1, into bacterial cell membranes, several monolayers have been constructed and set up on the Langmuir trough. Bacterial cell membranes typically contain large amounts of anionic PG lipids, such as DPPG and POPG. It was known that unsaturated lipids such as POPG and POPC can be subject to oxidative attack. Hence, oxidatively damaged PC lipids were incorporated into some monolayers, as previously described in Table 5.2, to explore how continuous oxidative damage would affect the degree of peptide insertion into the bacterial lipid membrane. Oxidatively damaged PG lipids were not commercially available.

The scattering length values of atoms contained in the lipids were taken from the NIST Center for Neutron Research (Table 5.3), and the total scattering length sum of each lipid was calculated in Å. The volumes of the lipids were taken from various sources (Pan *et al.*, 2012; Nagle *et al.*, 2019) The SLD of the lipid tails and headgroups were calculated by dividing the total SLD of the tail or headgroup region, by the volume of the tail or headgroups respectively (Table 5.4). The SLD of an average lipid in the lipid mixtures were also calculated in Å, in an 80:20 mol% ratio. The upper limit of the SLD for these mixtures, used in neutron fitting, was calculated by dividing the SLD of an average lipid in that mixture, by the respective lipid volumes in the 80:20 mol% ratio.

Atom	Scattering lengths / 10^{-15} m
C	6.65
H	-3.74
D	6.67
O	5.80
P	5.13
Na	3.63
N	9.36

Table 5.3: The scattering length values for the respective atoms in the lipids (Sears, 1992).

	Volume (total) / Å	Volume (headgroup) / Å^3	SLD (tails) / Å^{-2}	SLD (headgroup) / Å^{-2}
d_{62}DPPG	1111.7	291	7.47×10^{-6}	2.58×10^{-6}
d_{31}POPG	1205.4	291	3.24×10^{-6}	2.46×10^{-6}
POPC	1256.5	331	-1.86×10^{-7}	1.82×10^{-6}
POnPC	1090	331	-1.75×10^{-7}	1.82×10^{-6}

Table 5.4: Volumes of lipids obtained from various sources interpolated to a temperature of 25°C.

Figures 5.9 and 5.10 shows the NR curves of a monolayer containing only d_{62}DPPG lipids on a D_2O-based buffer subphase, recorded at a starting surface pressure of 10-15 mN m^{-1}, which increased to 30 mN m^{-1} upon compression by the trough barriers and down to 10-15 mN m^{-1} again upon expansion of the barriers. Error bars were included for the plots of d_{62}DPPG with and without PG-1 at 10 mN m^{-1}, where the error bars were computed by dividing the error in the reflectivity by the value of the reflectivity, multiplied by $\frac{1}{ln10}$ (0.4343). Error bars were not included in other plots for the purpose of clarity. At all measured surface pressures, there was no significant change in reflectivity for the d_{62}DPPG monolayers with and without PG-1. Figures 5.11 and 5.12 show d_{62}DPPG monolayers with and without PG-1 on an NRW-based buffer subphase, where it was observed that at lower surface pressures, the reflectivity of the monolayer with PG-1 was lower compared to the monolayer without PG-1.

Table 5.5 shows the fitting data of the two monolayers with and without PG-1, where it was observed that the VFS of the heads in both d_{62}DPPG monolayers with and without PG-1 decreased with increasing surface pressure and increased with decreasing surface pressure. The percentage change between the monolayers with and without PG-1 were calculated for the total thickness and the VFS in Table 5.6. There was generally no increase in total thickness in the d_{62}DPPG monolayer even when PG-1 was added. The VFS of the heads decreased at 30 mN m^{-1} and 10 mN m^{-1} expansion but no conclusive differences were observed to the monolayer upon the addition of PG-1.

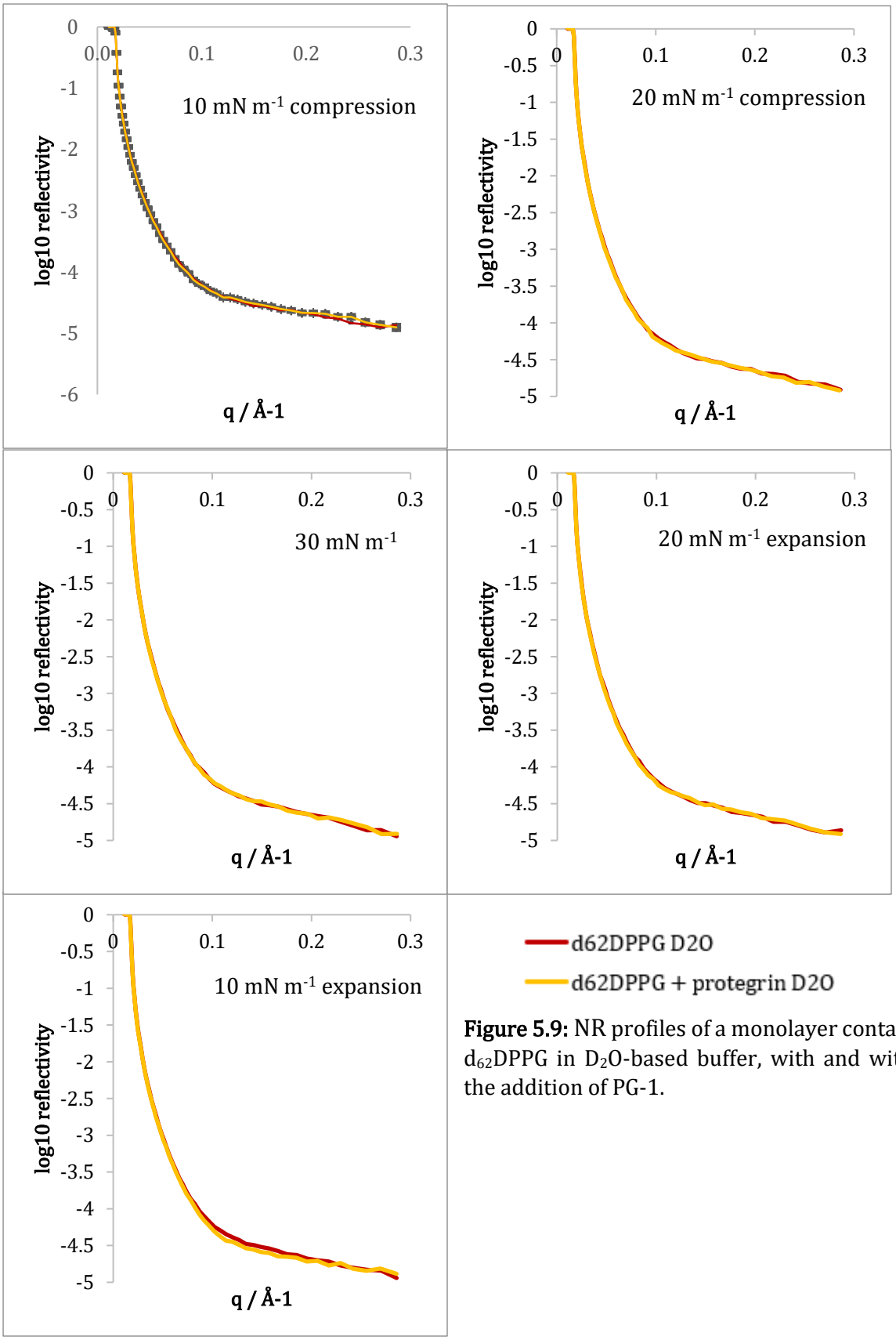


Figure 5.9: NR profiles of a monolayer containing d_{62}DPPG in D_2O-based buffer, with and without the addition of PG-1.

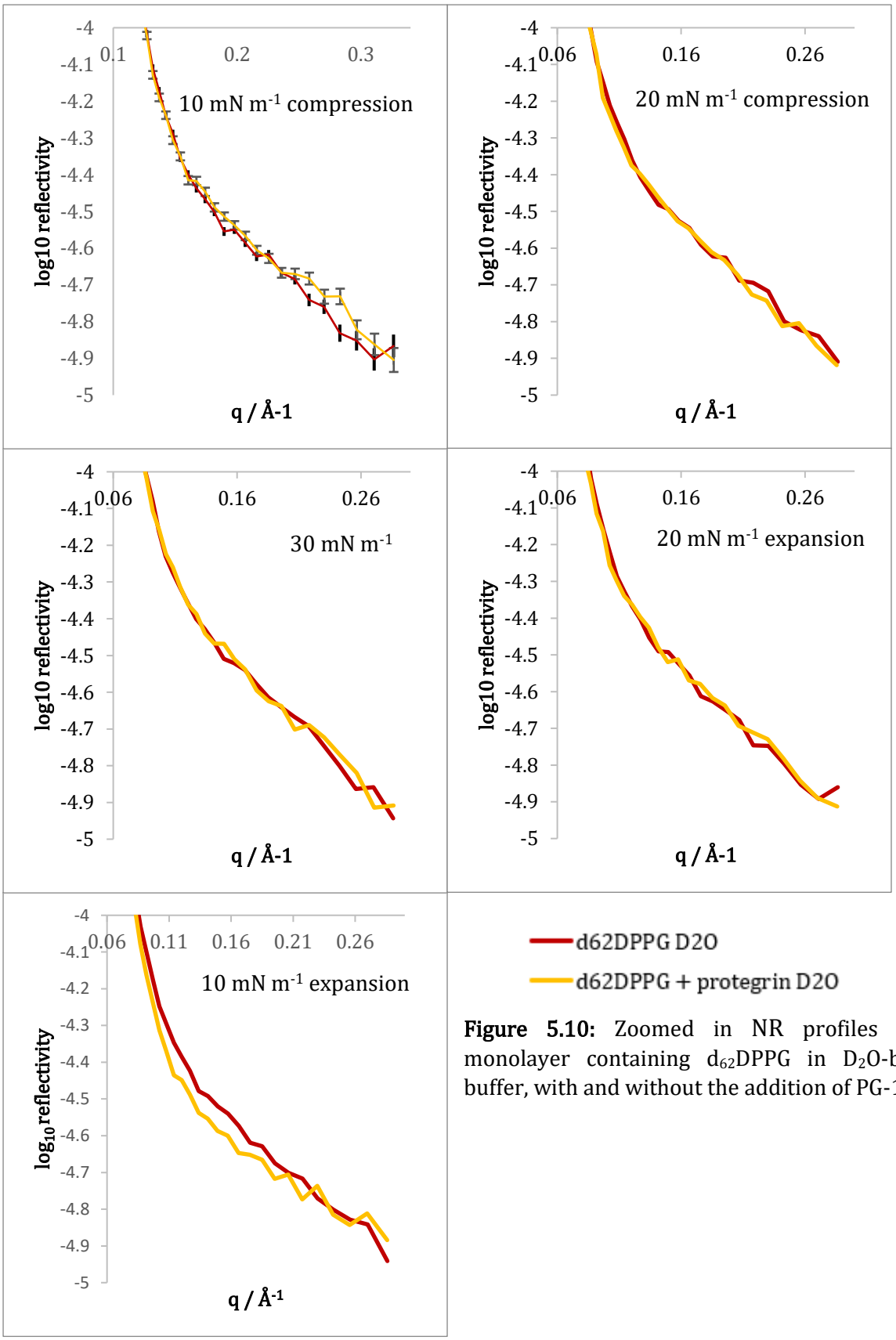


Figure 5.10: Zoomed in NR profiles of a monolayer containing d_{62}DPPG in D_2O-based buffer, with and without the addition of PG-1.

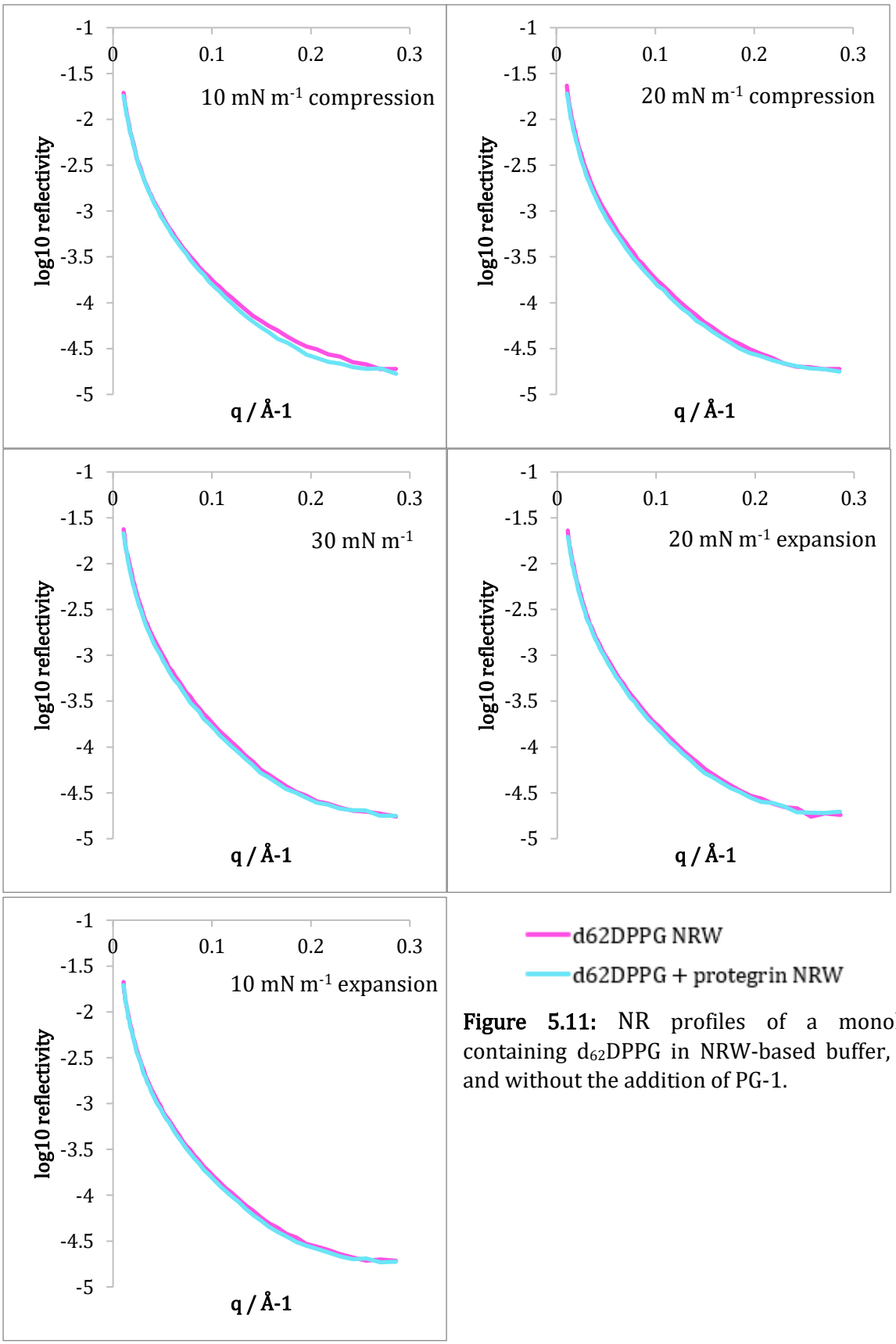


Figure 5.11: NR profiles of a monolayer containing d_{62}DPPG in NRW-based buffer, with and without the addition of PG-1.

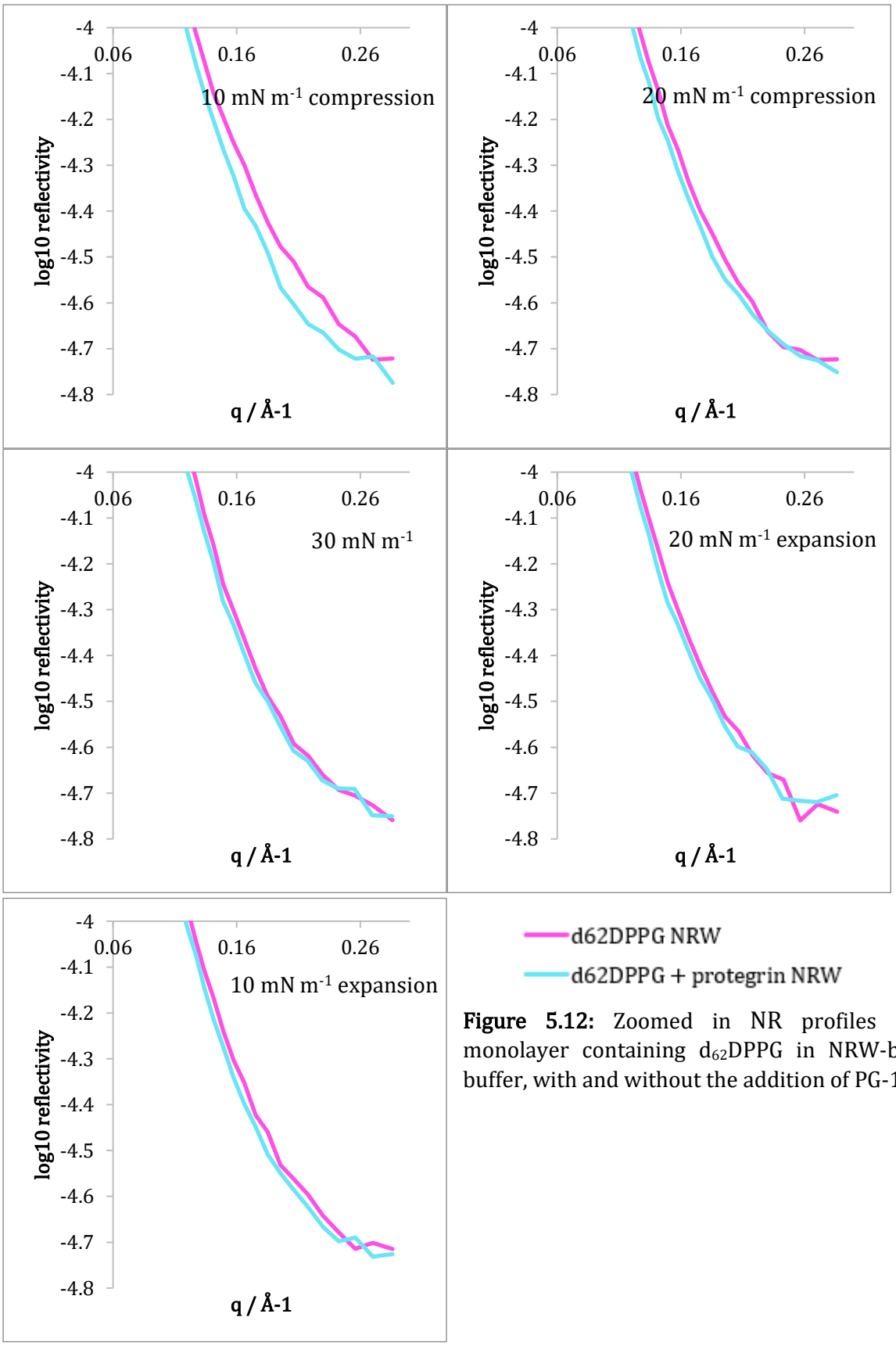


Figure 5.12: Zoomed in NR profiles of a monolayer containing d_{62}DPPG in NRW-based buffer, with and without the addition of PG-1.

d_{62}DPPG								
Surface pressure / mN^{-1}	**Thickness (tails) / Å**	**Sigma / Å**	**Thickness (heads) / Å**	**Sigma in / Å**	**Thickness (total) / Å**	**SLD (tails) / × 10^{-6} $Å^{-2}$**	**SLD (heads) / × 10^{-6} $Å^{-2}$**	**VFS (heads)**
10	17.57	0.03	3.80	0.06	21.37	6.35	2.58	0.15
20	17.16	0.03	5.24	0.06	22.40	6.72	2.58	0.1
30	16.21	0.02	6.42	0.05	22.63	7.01	2.58	0.06
20	16.70	0.03	5.48	0.06	22.19	6.72	2.58	0.1
10	17.18	0.03	4.47	0.06	21.65	6.35	2.58	0.15
With PG-1:								
10	17.12	0.03	4.81	0.06	21.93	6.24	2.58	0.16
20	16.27	0.03	5.57	0.06	21.84	6.58	2.58	0.12
30	14.60	0.02	6.83	0.05	21.43	7.17	2.58	0.04
20	15.66	0.03	6.40	0.06	22.06	6.72	2.58	0.1
10	14.95	0.03	6.47	0.06	21.42	6.64	2.58	0.11

Table 5.5: Fitting results of the d_{62}DPPG monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	**Percentage change in total layer thickness**	**Change VFS**
10	103	+0.01
20	98	+0.02
30	95	-0.02
20	99	0
10	99	+0.04

Table 5.6: Percentage changes in the total thickness (heads and tails) and VFS (heads) between the d_{62}DPPG monolayers with and without PG-1.

In the d_{62}DPPG:POPC (80:20 mol%) monolayer (Figure 5.13, 5.14), the reflectivity curve of the monolayer with PG-1 was observed to be lower than the monolayer without PG-1, upon compression and expansion to 20 mN m^{-1}. A small change in reflectivity was also seen during the 10 mN m^{-1} compression, but no significant differences were observed at 30 mN m^{-1}. Generally, the VFS of the lipid heads in the monolayer without PG-1 is lower than the monolayer containing PG-1 at all surface pressures (Table 5.7, 5.8). The total thickness of the lipid heads and tails were seen to be higher in the monolayers containing PG-1, where an increase in thickness was observed at all surface pressures as seen in Table 5.8.

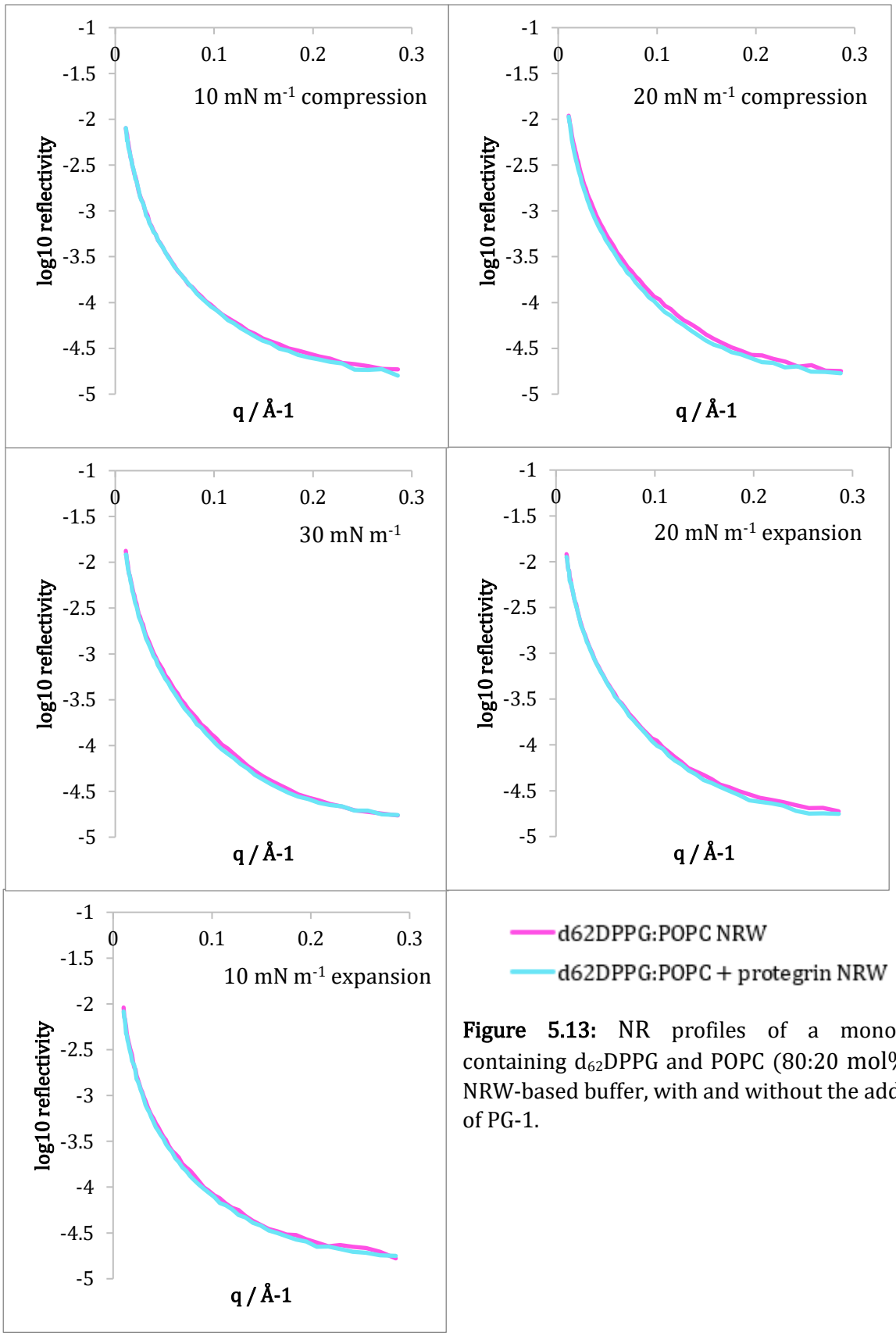


Figure 5.13: NR profiles of a monolayer containing d_{62}DPPG and POPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

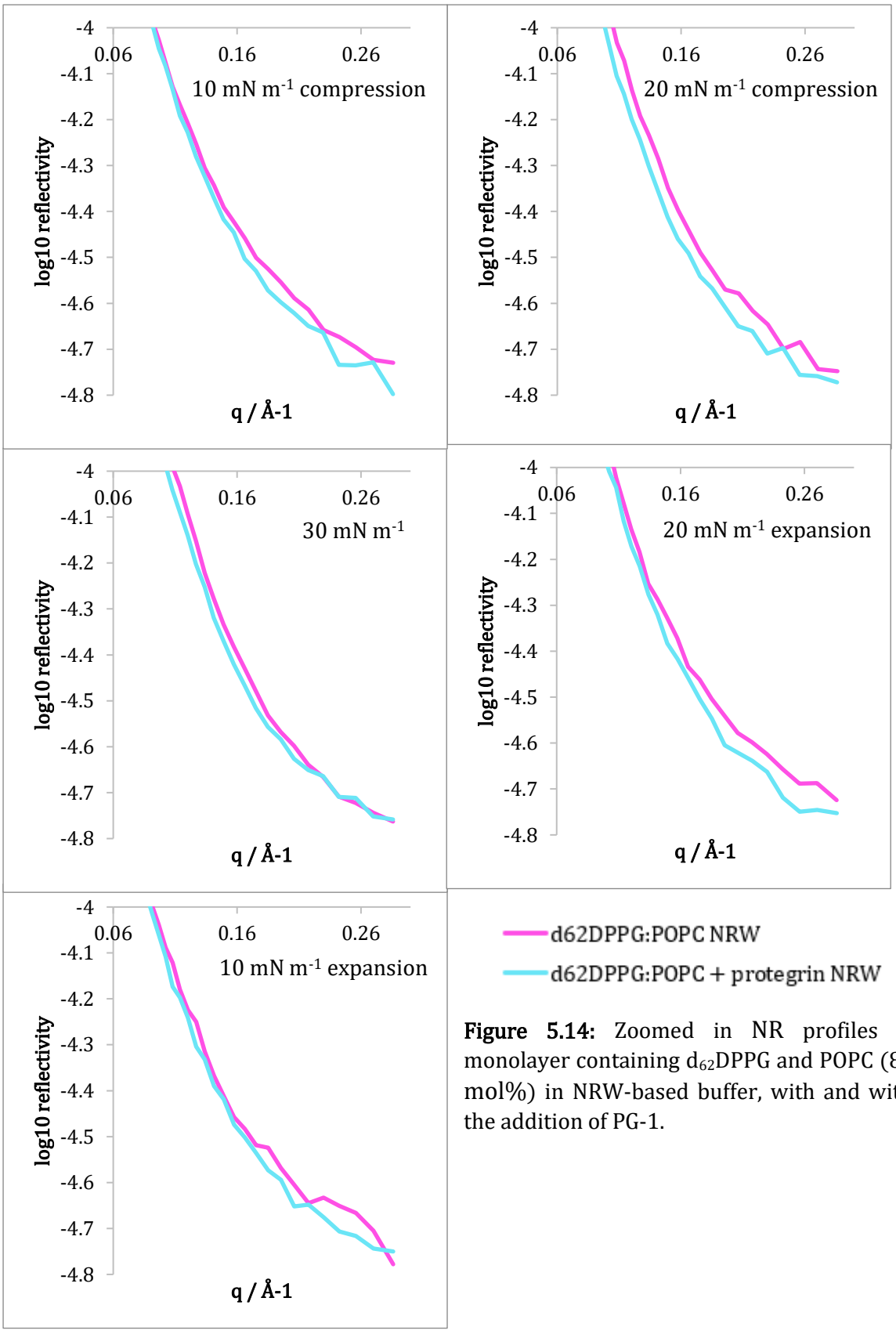


Figure 5.14: Zoomed in NR profiles of a monolayer containing d_{62}DPPG and POPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

d_{62}DPPG with POPC (80:20 mol%)								
Surface pressure / mN^{-1}	**Thickness (tails) / Å**	**Sigma / Å**	**Thickness (heads) / Å**	**Sigma in / Å**	**Thickness (total) / Å**	**SLD (tails) / × 10^{-6} $Å^{-2}$**	**SLD (heads) / × 10^{-6} $Å^{-2}$**	**VFS (heads)**
10	10.14	6.73	7.92	5.36	18.05	5.79	2.41	0.02
20	12.07	2.60	9.44	2.10	21.51	5.79	2.41	0
30	13.71	1.59	9.30	1.42	23.01	5.79	2.41	0.01
20	12.38	2.43	8.50	2.12	20.88	5.79	2.41	0
10	10.19	11.61	9.06	8.60	19.26	5.79	2.41	0.05
With PG-1:								
10	13.48	2.97706	6.66	3.10	20.14	5.79	2.41	0.19
20	14.38	1.90029	8.75	1.80	23.12	5.79	2.41	0.17
30	12.54	2.17367	11.40	1.61	23.94	5.79	2.41	0.03
20	12.40	2.45116	10.44	1.90	22.84	5.79	2.41	0.05
10	11.38	4.96644	8.85	4.00	20.23	5.79	2.41	0.15

Table 5.7: Fitting results of the d_{62}DPPG:POPC (80:20 mol%) monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	**Percentage change in total layer thickness**	**Change in VFS**
10	112	+0.17
20	108	+0.17
30	104	+0.02
20	109	+0.05
10	105	+0.10

Table 5.8: Percentage changes in the thickness and VFS (heads) between the d_{62}DPPG:POPC (80:20 mol%) monolayers with and without PG-1.

The reflectivity curves of the d_{62}DPPG:POnPC (80:20 mol%) monolayers with and without PG-1 in NRW-based buffer, did not show any significant changes in reflectivity at the measured surface pressures (Figure 5.15, 5.16). There was a positive percentage change in total thickness of the lipid heads and tails from the monolayer without PG-1 to the monolayer with PG-1, at all surface pressures (Table 5.10). In the monolayer without PG-1, the VFS was largest at the lowest measured surface pressure of 10 mN m^{-1}, which decreased to 0 as the surface pressure increased to 30 mN m^{-1}. The VFS of the lipid heads was generally observed to be larger in the monolayer containing PG-1, except for the VFS at the 20 mN m^{-1} expansion (Table 5.9).

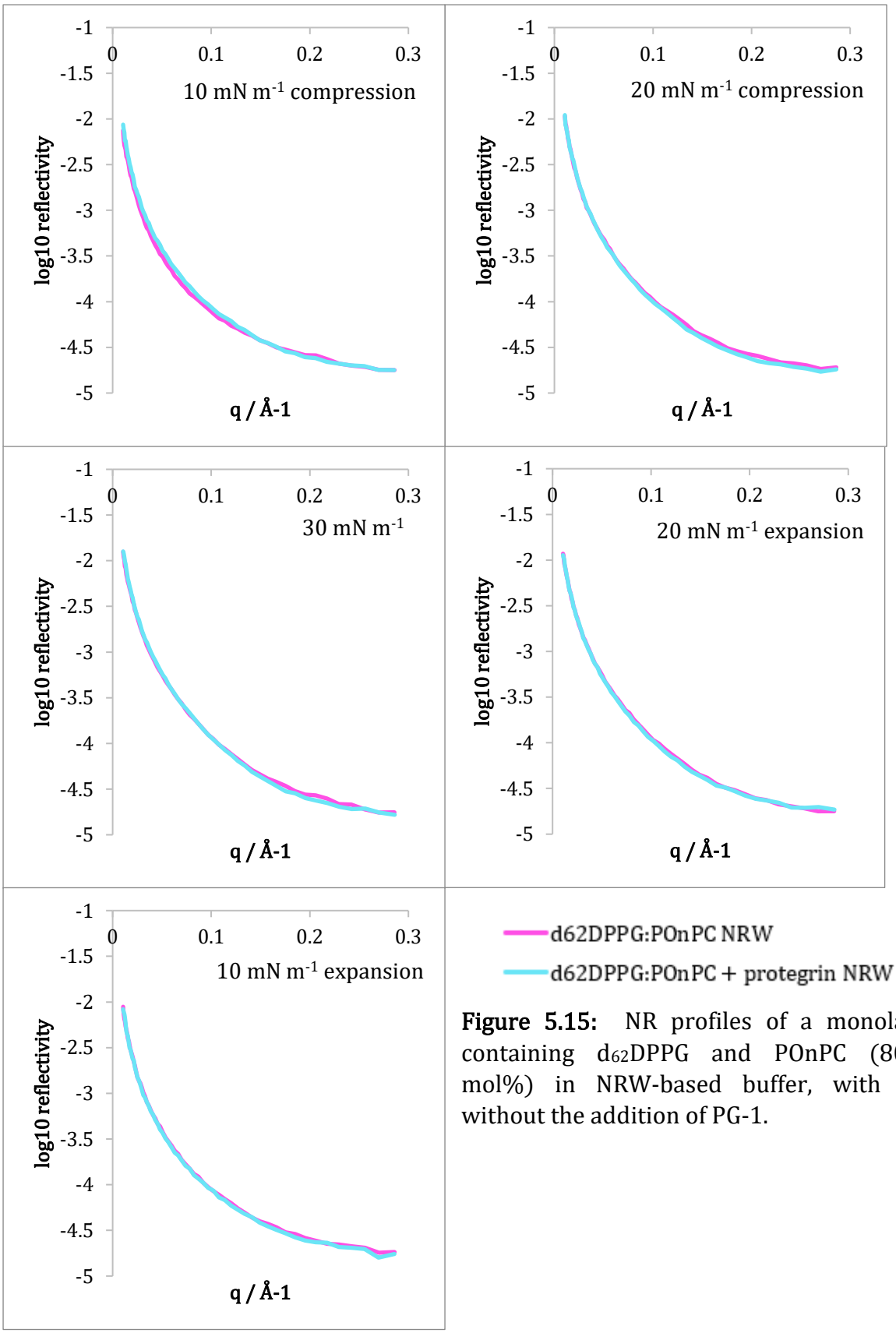


Figure 5.15: NR profiles of a monolayer containing d_{62}DPPG and POnPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

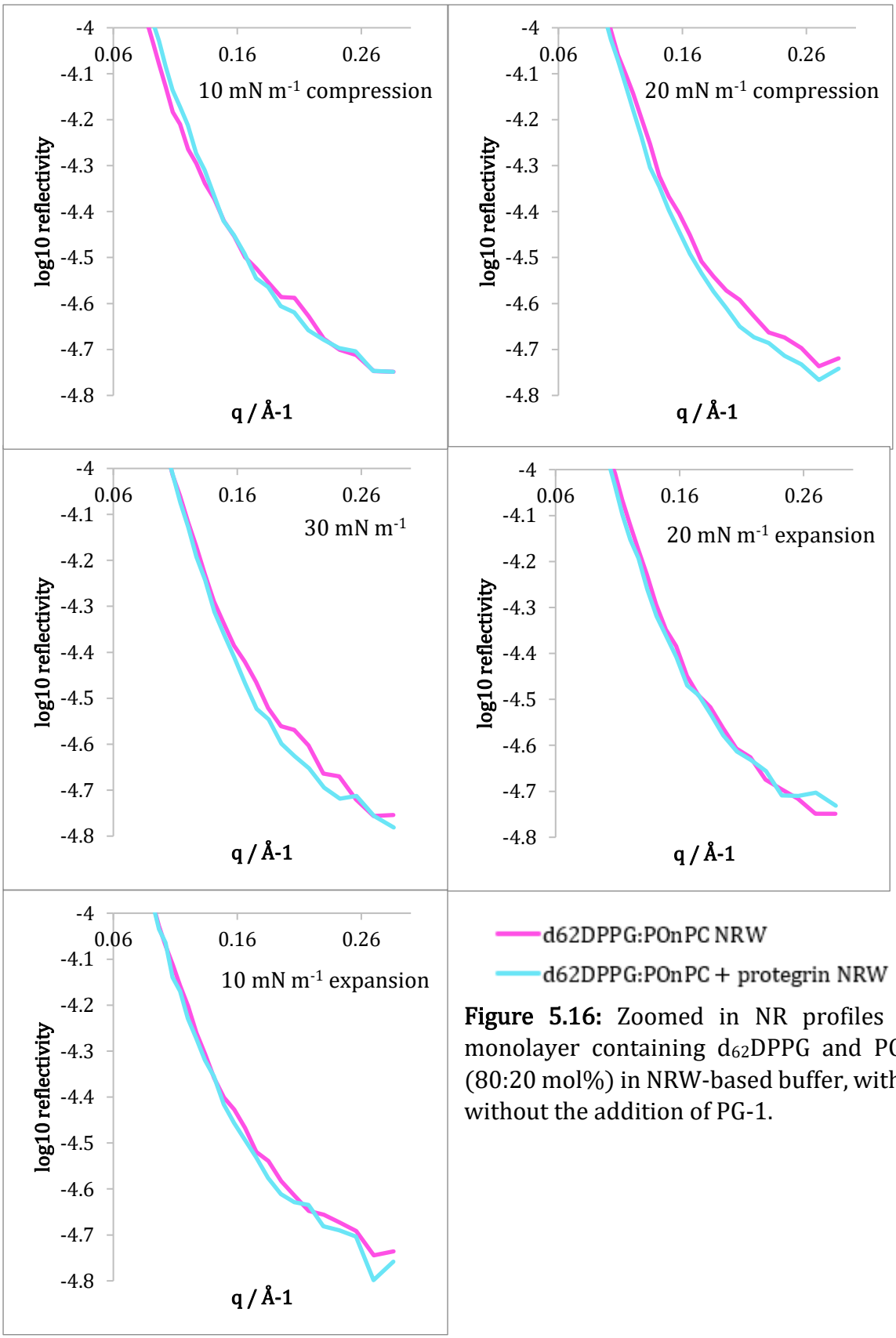


Figure 5.16: Zoomed in NR profiles of a monolayer containing d_{62}DPPG and POnPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

d_{62}DPPG with POnPC (80:20 mol%)								
Surface pressure / mN^{-1}	**Thickness (tails) / Å**	**Sigma / Å**	**Thickness (heads) / Å**	**Sigma in / Å**	**Thickness (total) / Å**	**SLD (tails) / × 10^{-6} $Å^{-2}$**	**SLD (heads) / × 10^{-6} $Å^{-2}$**	**VFS (heads)**
10	10.93	5.05	7.19	4.48	18.12	6.04	2.41	0.15
20	11.49	3.03	9.31	2.40	20.80	6.04	2.41	0.05
30	11.92	2.31	10.19	1.78	22.11	6.04	2.41	0
20	13.09	1.79	9.29	1.56	22.38	6.04	2.41	0.06
10	10.94	4.32	8.90	3.40	19.85	6.04	2.41	0.10
With PG-1:								
10	12.75	2.76	8.01	2.56	20.77	6.04	2.41	0.19
20	14.29	1.66	9.08	1.56	23.37	6.04	2.41	0.19
30	13.78	1.40	10.69	1.16	24.47	6.04	2.41	0.09
20	11.42	2.92	11.50	2.02	22.92	6.04	2.41	0.02
10	13.06	3.15	8.00	2.97	21.06	6.04	2.41	0.20

Table 5.9: Fitting results of the d_{62}DPPG:POnPC (80:20 mol%) monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	**Percentage change in total layer thickness**	**Change in VFS**
10	115	+0.04
20	112	+0.14
30	111	+0.09
20	102	-0.04
10	106	+0.10

Table 5.10: Percentage changes in the thickness and VFS (heads) between the d_{62}DPPG:POnPC (80:20 mol%) monolayers with and without PG-1.

The reflectivity profile of the d_{31}POPG monolayer in NRW-based buffer with PG-1 was lower than the monolayer without PG-1 when they were expanded to a surface pressure of 20 mN m^{-1}. No observable differences were seen at the other surface pressures with and without PG-1, and data for the d_{31}POPG monolayer containing PG-1 at the 20 mN m^{-1} expansion could not be obtained for comparison (Figure 5.17, 5.18). Generally, the thicknesses of the lipid tails were seen to be larger in the monolayer with PG-1 at the corresponding surface pressures. On the other hand, the thicknesses of the lipid heads were larger in the monolayer without PG-1, except for the lipid head thickness measured at the 20 mN m^{-1} expansion. The VFS of the heads appeared larger at lower surface pressures for the monolayer with PG-1 added, except at the highest surface pressure of 30 mN m^{-1}. In the monolayer containing PG-1, the VFS of the lipid heads was seen to decrease with increasing surface pressure (Table 5.11). There was a positive change in percentage thickness at the 15 mN m^{-1} compression and 20 mN m^{-1} expansion, and a negative change observed for the 20 mN m^{-1} compression and 30 mN m^{-1}. The percentage change in VFS was smallest at the highest surface pressure, and a larger change was observed between the monolayers with and without PG-1 at lower surface pressures of 10 mN m^{-1} (Table 5.12).

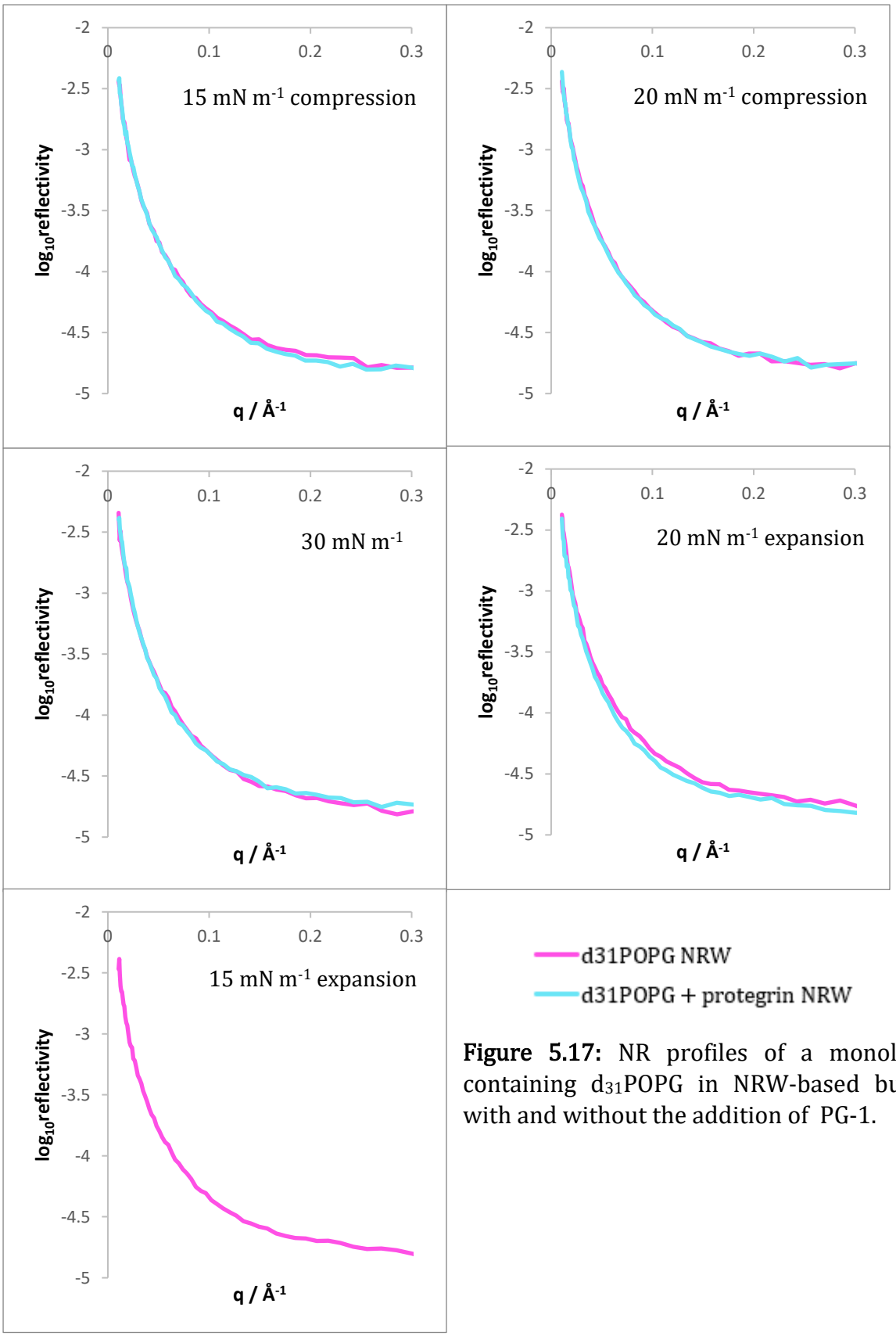


Figure 5.17: NR profiles of a monolayer containing d_{31}POPG in NRW-based buffer, with and without the addition of PG-1.

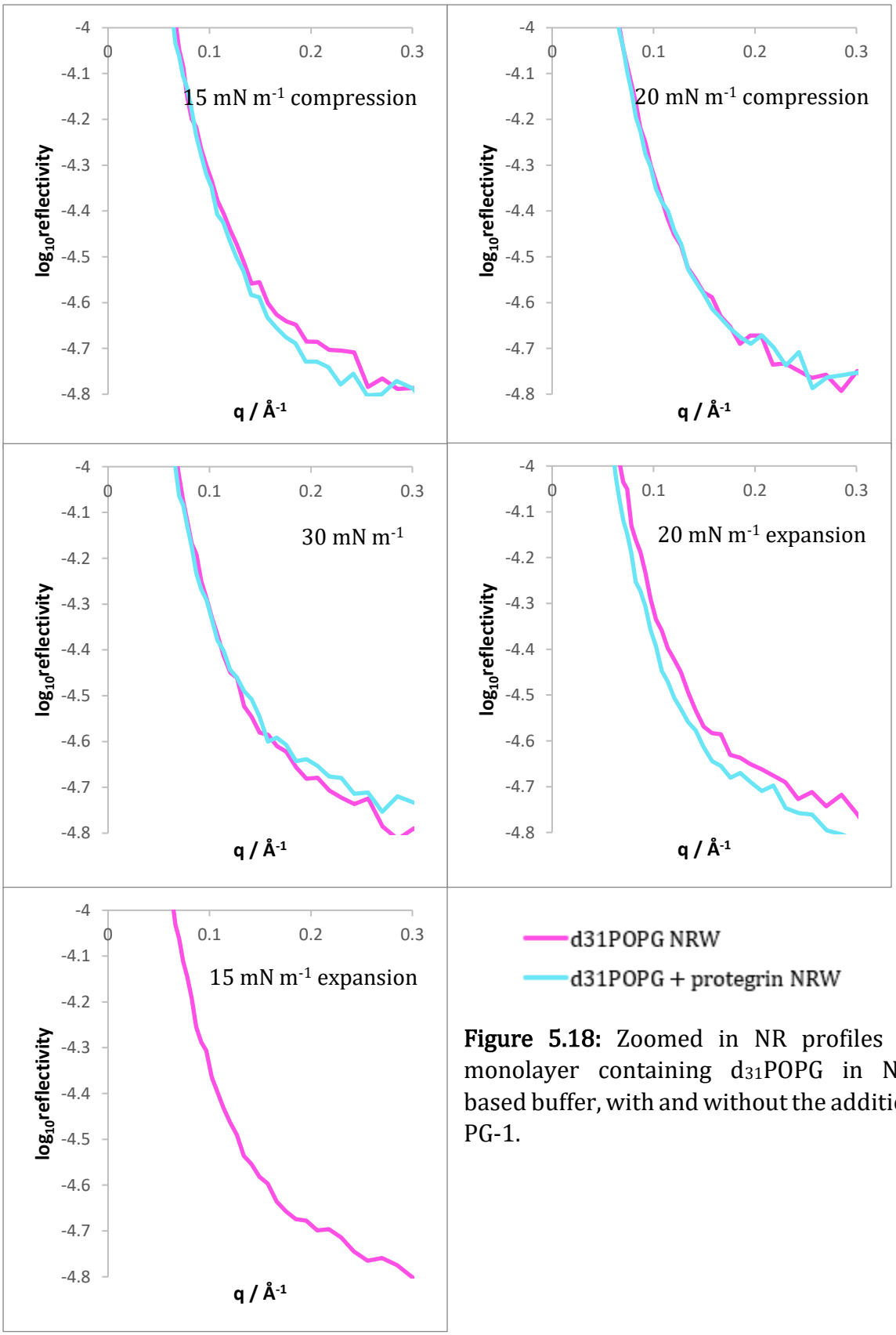


Figure 5.18: Zoomed in NR profiles of a monolayer containing d_{31}POPG in NRW-based buffer, with and without the addition of PG-1.

d_{31}POPG								
Surface pressure / mN^{-1}	**Thickness (tails) / Å**	**Sigma / Å**	**Thickness (heads) / Å**	**Sigma in / Å**	**Thickness (total) / Å**	**SLD (tails) / × 10^{-6} $Å^{-2}$**	**SLD (heads) / × 10^{-6} $Å^{-2}$**	**VFS (heads)**
15	12.16	135.96	6.62	125.68	18.78	3.24	2.46	0.08
20	13.46	77.88	7.24	72.17	20.70	3.24	2.46	0.13
30	11.84	97.75	7.89	87.66	19.73	3.24	2.46	0.07
20	12.93	50.80	4.84	49.99	17.77	3.24	2.46	0.03
15	13.38	74.63	6.54	70.25	19.92	3.24	2.46	0.15
With PG-1:								
15	19.45	28.52	2.41	33.11	21.86	3.24	2.46	0.25
20	14.69	39.79	5.56	39.16	20.25	3.24	2.46	0.15
30	12.04	143.77	6.89	131.88	18.93	3.24	2.46	0.05
20	14.82	60.32	7.27	56.78	22.09	3.24	2.46	0.26

Table 5.11: Fitting results of the d_{31}POPG monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	**Percentage change in total layer thickness**	**Change in VFS**
15	116	+0.17
20	98	+0.02
30	96	+0.08
20	124	+0.02

Table 5.12: Percentage changes in the thickness and VFS (heads) between the d_{31}POPG monolayers with and without PG-1.

The reflectivity profiles of the d_{31}POPG:POPC (80:20 mol%) monolayers were recorded in both NRW and D_2O-based buffers (Figure 5.19-5.22). Overall, no significant changes in reflectivity were observed with the addition of PG-1 at the measured surface pressures in both buffers. There were no systematic changes in the thicknesses of the lipid heads and tails across the monolayers with and without PG-1 added. In the monolayer without PG-1, the VFS of the heads remained around 0.12 to 0.13 at lower surface pressures and decreased to 0.004 at a surface pressure of 30 mN m^{-1} (Table 5.13). There was a positive change in percentage thickness when PG-1 was added, except when the monolayer was compressed to 20 mN m^{-1}. Similarly, the VFS showed an increase with the addition of PG-1, except at 20 mN m^{-1} where a negative percentage change was recorded (Table 5.14).

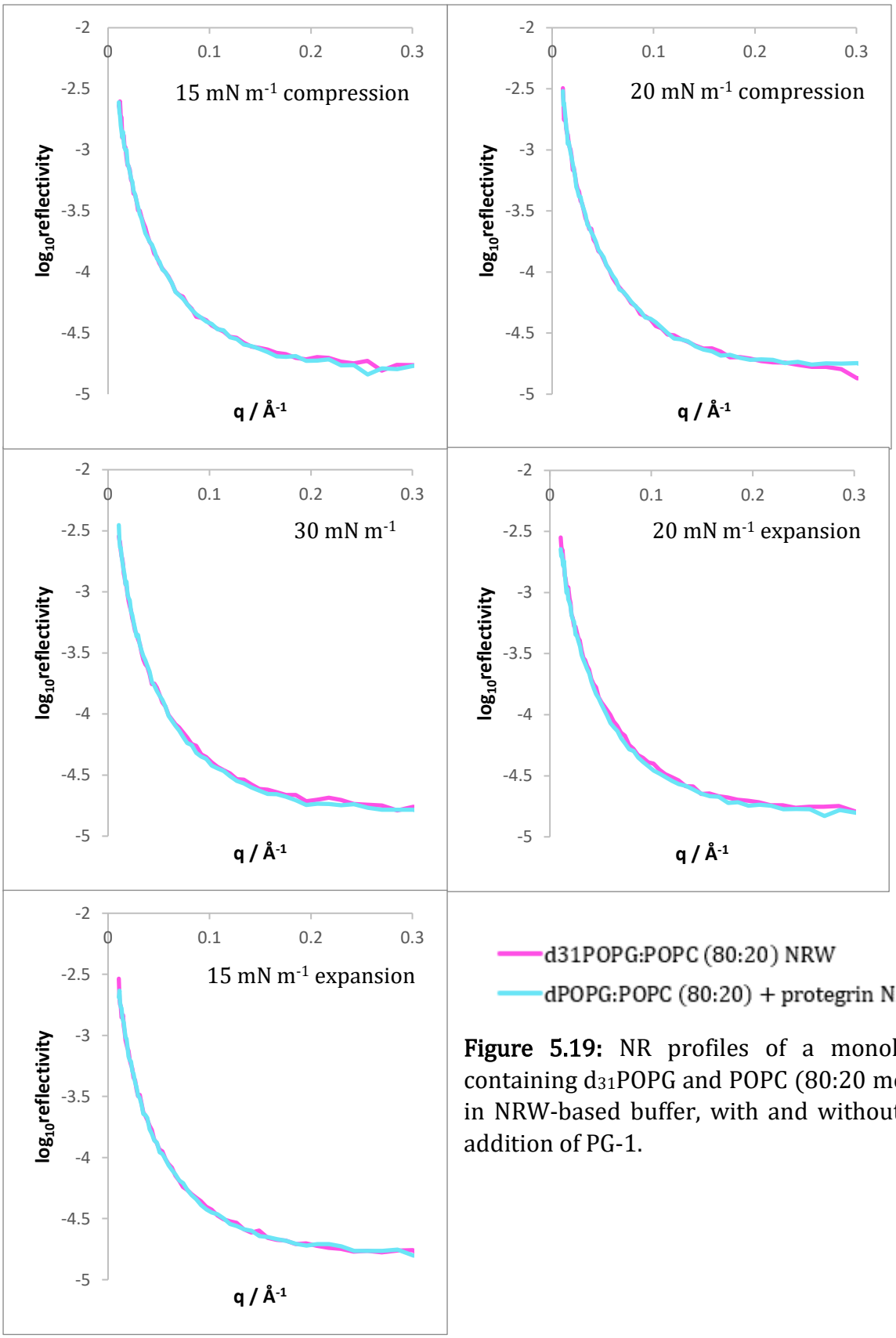


Figure 5.19: NR profiles of a monolayer containing d_{31}POPG and POPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

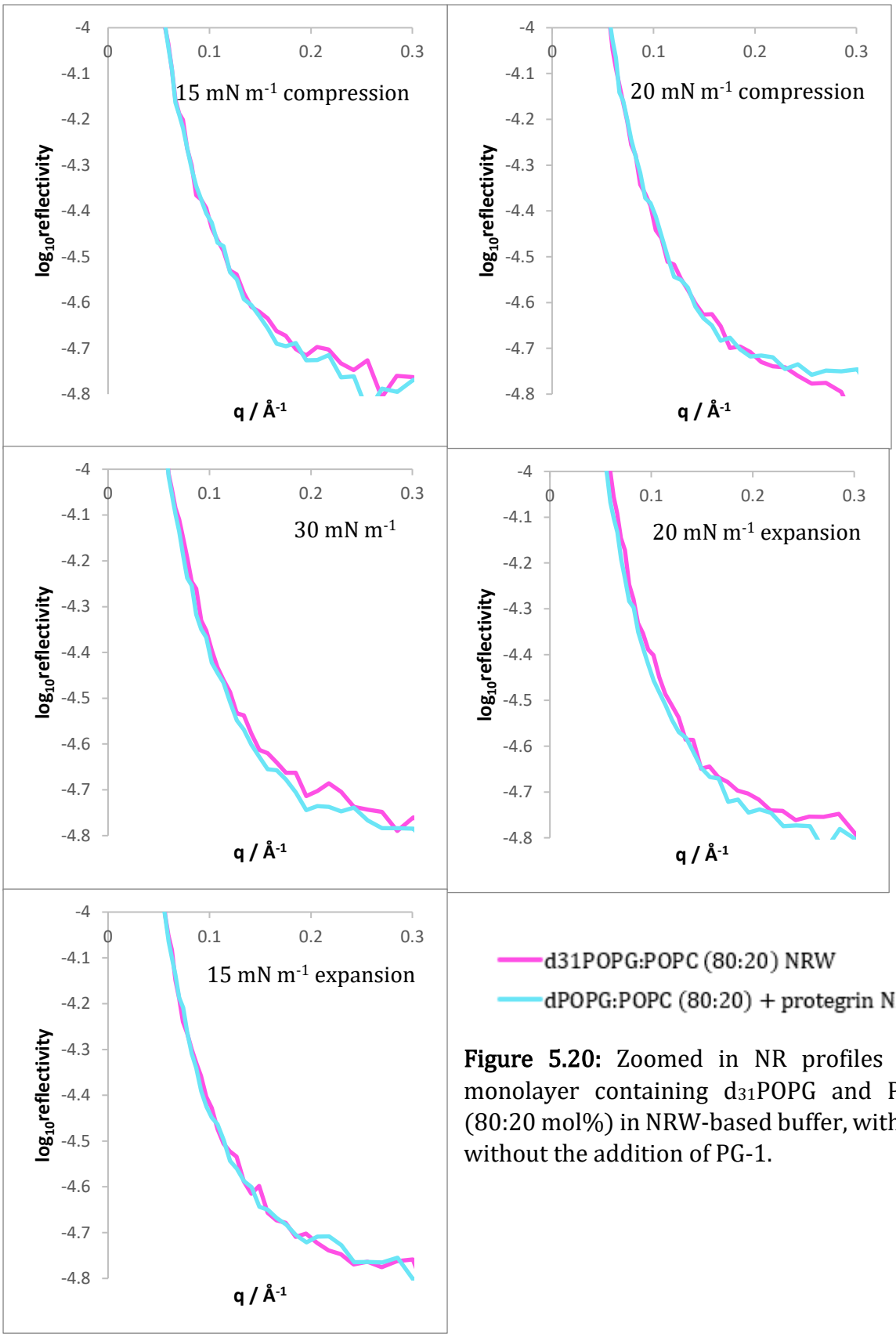


Figure 5.20: Zoomed in NR profiles of a monolayer containing d_{31}POPG and POPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

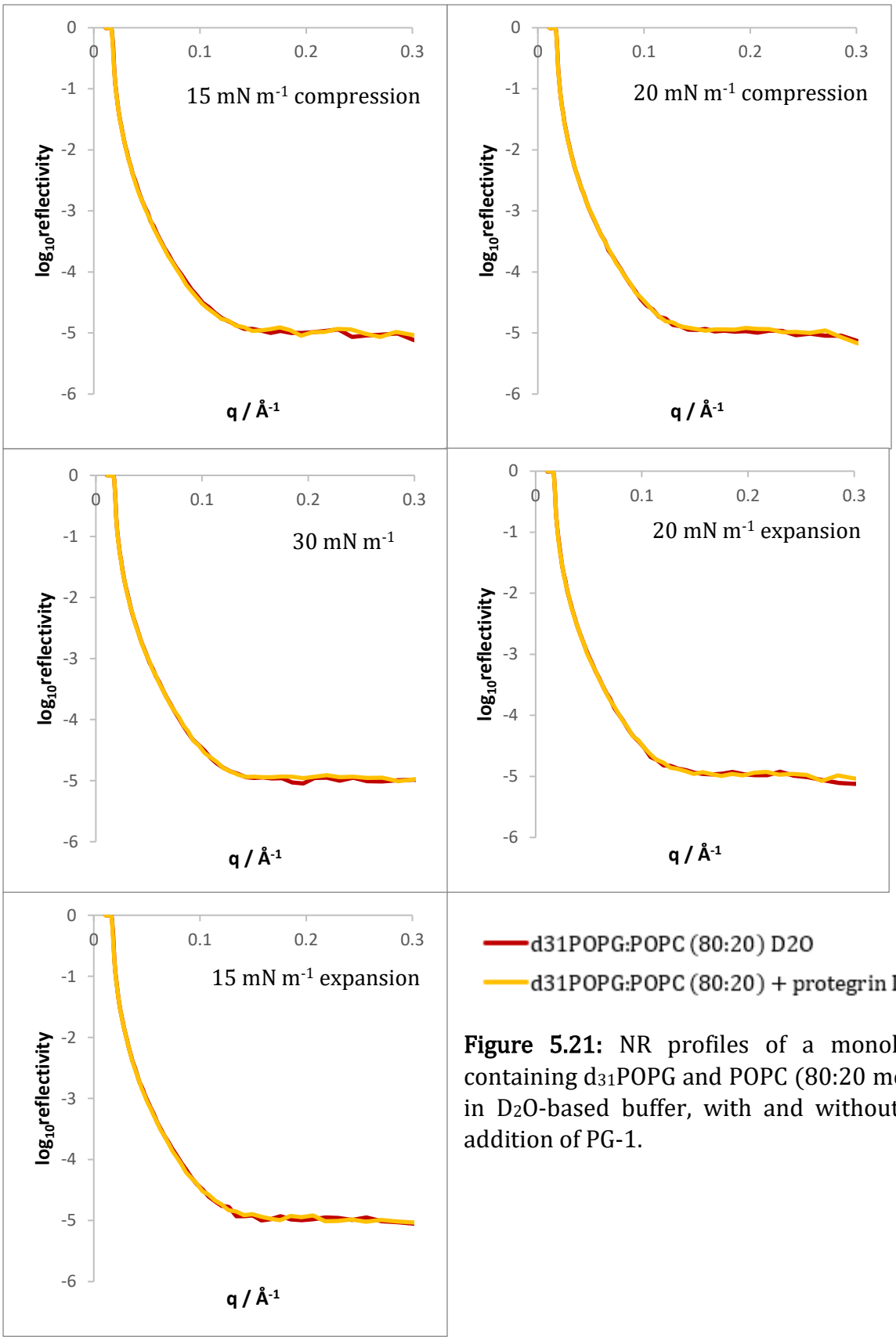


Figure 5.21: NR profiles of a monolayer containing d_{31}POPG and POPC (80:20 mol%) in D_2O-based buffer, with and without the addition of PG-1.

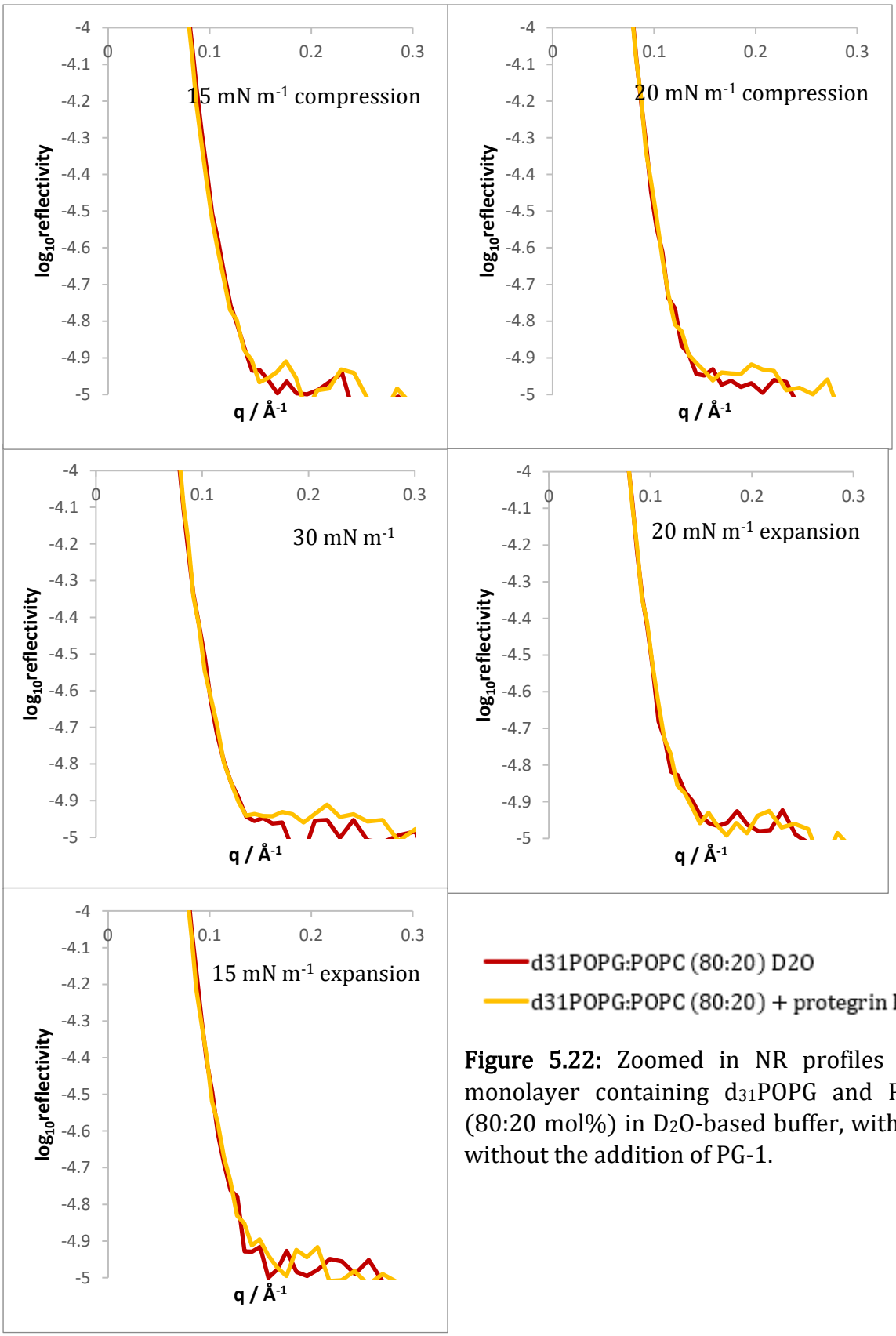


Figure 5.22: Zoomed in NR profiles of a monolayer containing d_{31}POPG and POPC (80:20 mol%) in D_2O-based buffer, with and without the addition of PG-1.

D_{31}POPG with POPC (80:20 mol%)								
Surface pressure / mN^{-1}	Thickness (tails) / Å	Sigma / Å	Thickness (heads) / Å	Sigma in / Å	Thickness (total) / Å	SLD (tails) / $\times 10^{-6}$ $Å^{-2}$	SLD (heads) / $\times 10^{-6}$ $Å^{-2}$	VFS (heads)
15	16.33	4.29	2.53	4.75	18.86	2.25	2.31	0.12
20	14.92	1.68	4.96	1.85	19.87	2.23	2.31	0.12
30	15.93	0.60	2.96	0.73	18.90	2.54	2.31	0.004
20	16.95	2.83	3.36	3.13	20.31	2.23	2.31	0.13
15	15.83	3.02	3.40	3.35	19.23	2.22	2.31	0.13
With PG-1:								
15	18.02	0.35	1.26	0.28	19.28	2.15	2.31	0.16
20	11.43	1.03	7.67	1.13	19.09	2.37	2.31	0.07
30	17.84	2.40	3.61	2.67	21.45	2.19	2.31	0.14
20	15.35	1.10	6.55	1.21	21.89	2.02	2.31	0.21
10	14.79	1.73	5.04	1.91	19.83	2.17	2.31	0.15

Table 5.13: Fitting results of the d_{31}POPG:POPC (80:20 mol%) monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	Percentage change in total layer thickness	Change in VFS
15	102	+0.04
20	96	-0.05
30	113	+0.14
20	108	+0.08
15	103	+0.02

Table 5.14: Percentage changes in the thickness and VFS (heads) between the d_{31}POPG:POPC (80:20 mol%) monolayers with and without PG-1.

Finally, the reflectivity profiles of both d_{31}POPG:POnPC (80:20 mol%) monolayers with and without PG-1 did not show systematic changes in the total lipid thicknesses and VFS of the heads, when measured at the surface pressures. Results for the d_{31}POPG:POnPC (80:20 mol%) monolayer with PG-1 expanded to surface pressures of 20 mN m^{-1} and 15 mN m^{-1} could not be recorded (Figure 5.23-26). In the monolayer without PG-1, the VFS of the lipid heads were higher at lower surface pressures, and decreases as the surface pressure increased (Table 5.15). A positive change in both the percentage thickness and VFS of the lipid heads were observed at all surface pressures during the compression of the monolayers from 15 to 30 mN m^{-1}, when PG-1 was added (Table 5.16).

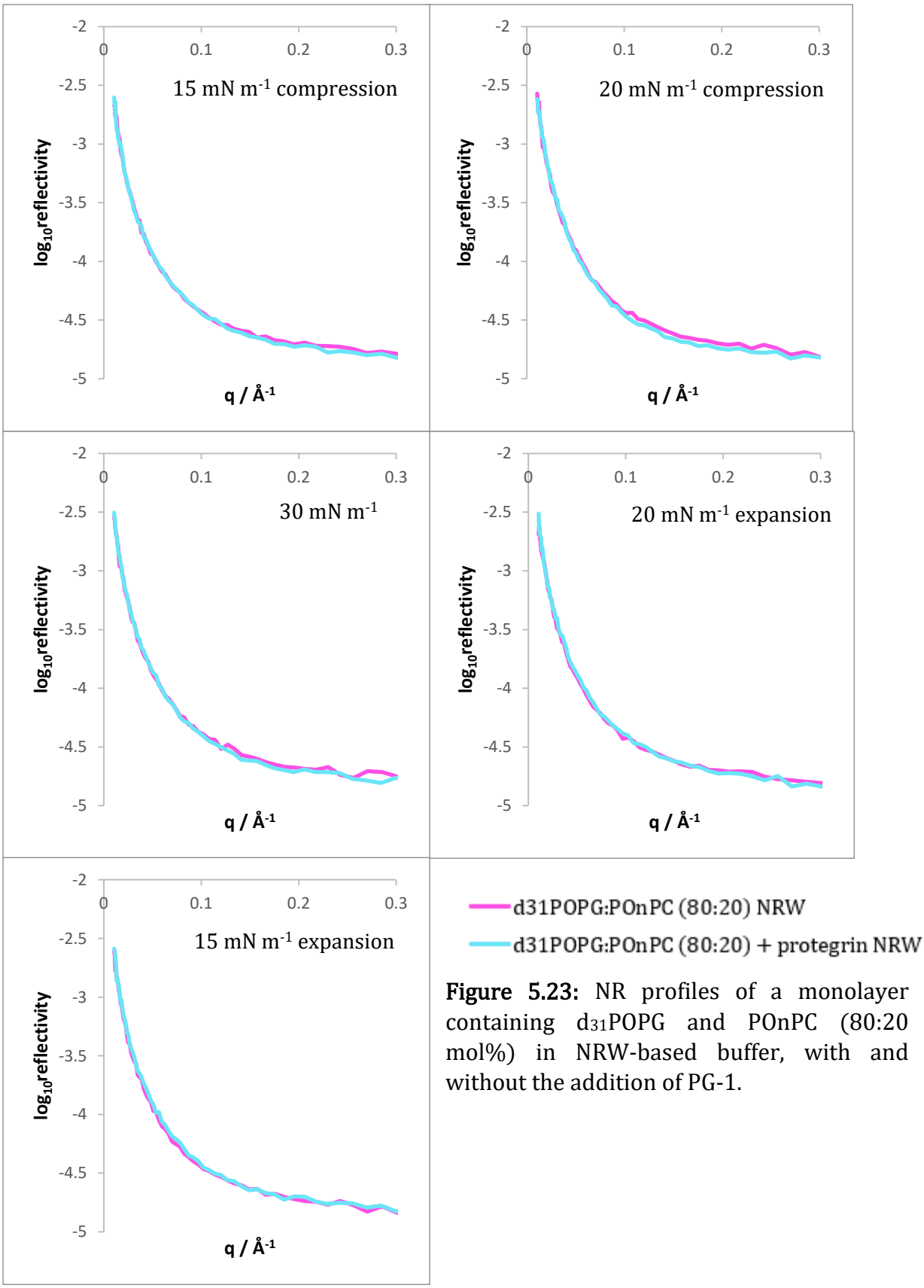


Figure 5.23: NR profiles of a monolayer containing d_{31}POPG and POnPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

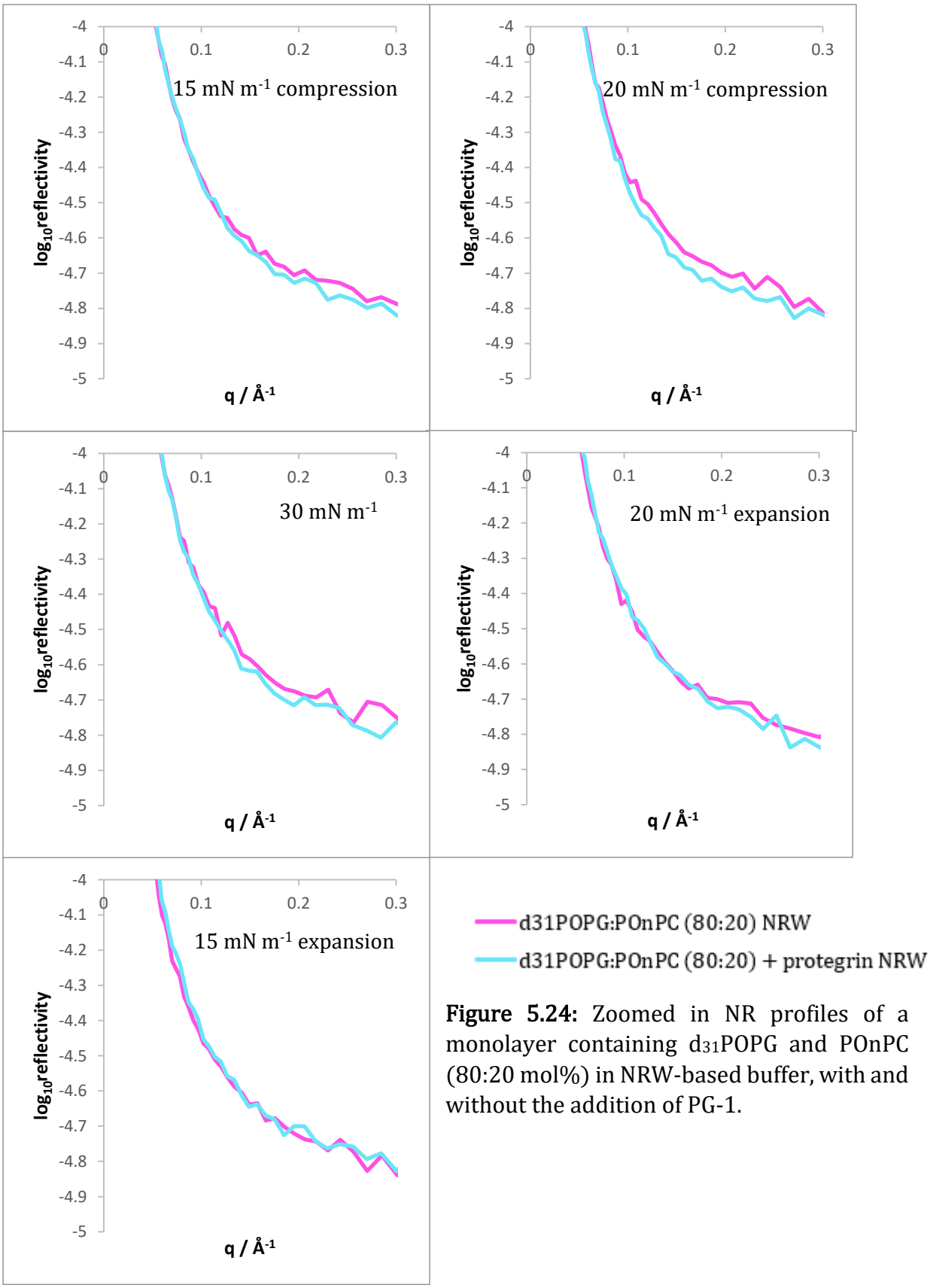


Figure 5.24: Zoomed in NR profiles of a monolayer containing d_{31}POPG and POnPC (80:20 mol%) in NRW-based buffer, with and without the addition of PG-1.

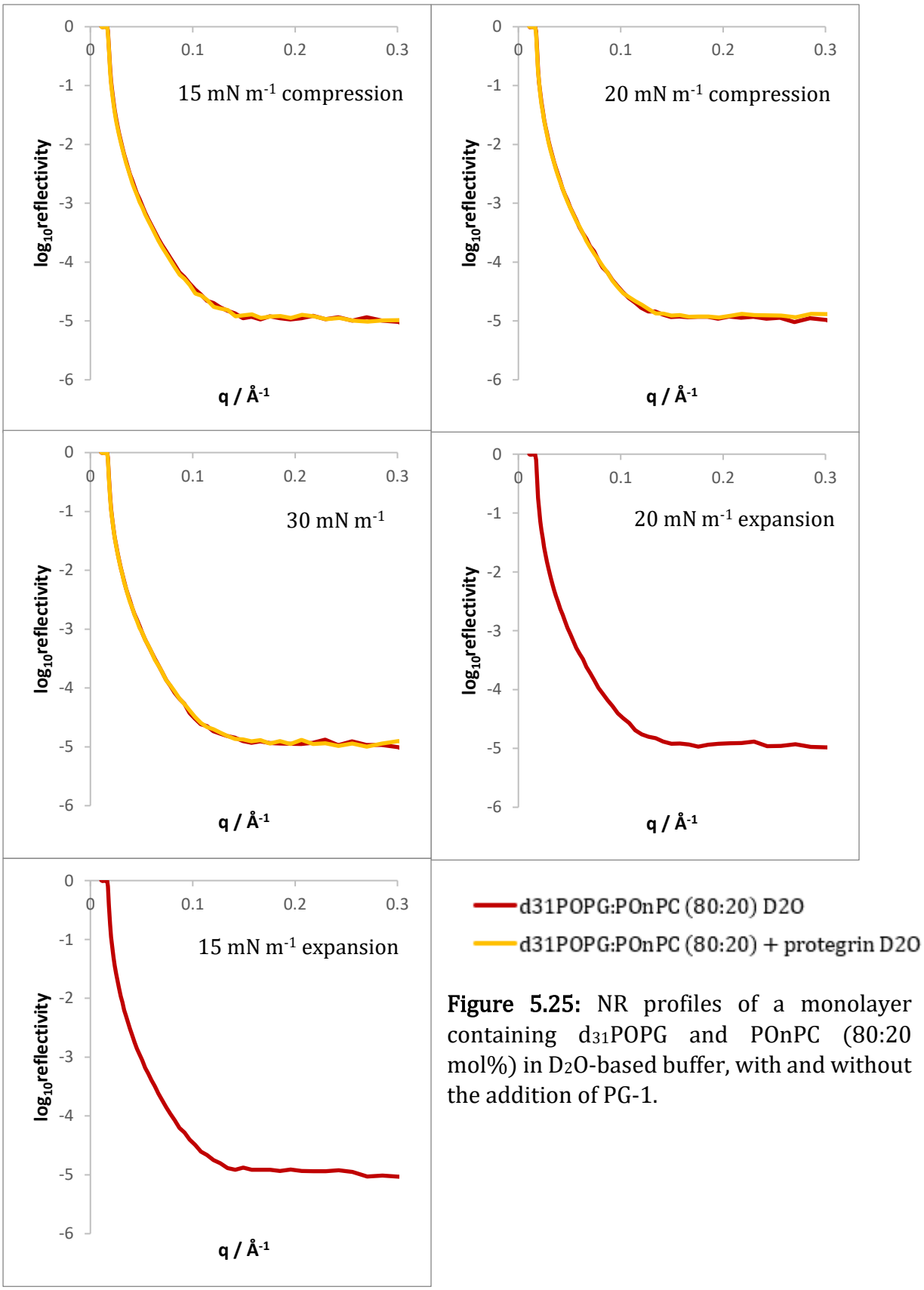


Figure 5.25: NR profiles of a monolayer containing d_{31}POPG and POnPC (80:20 mol%) in D_2O-based buffer, with and without the addition of PG-1.

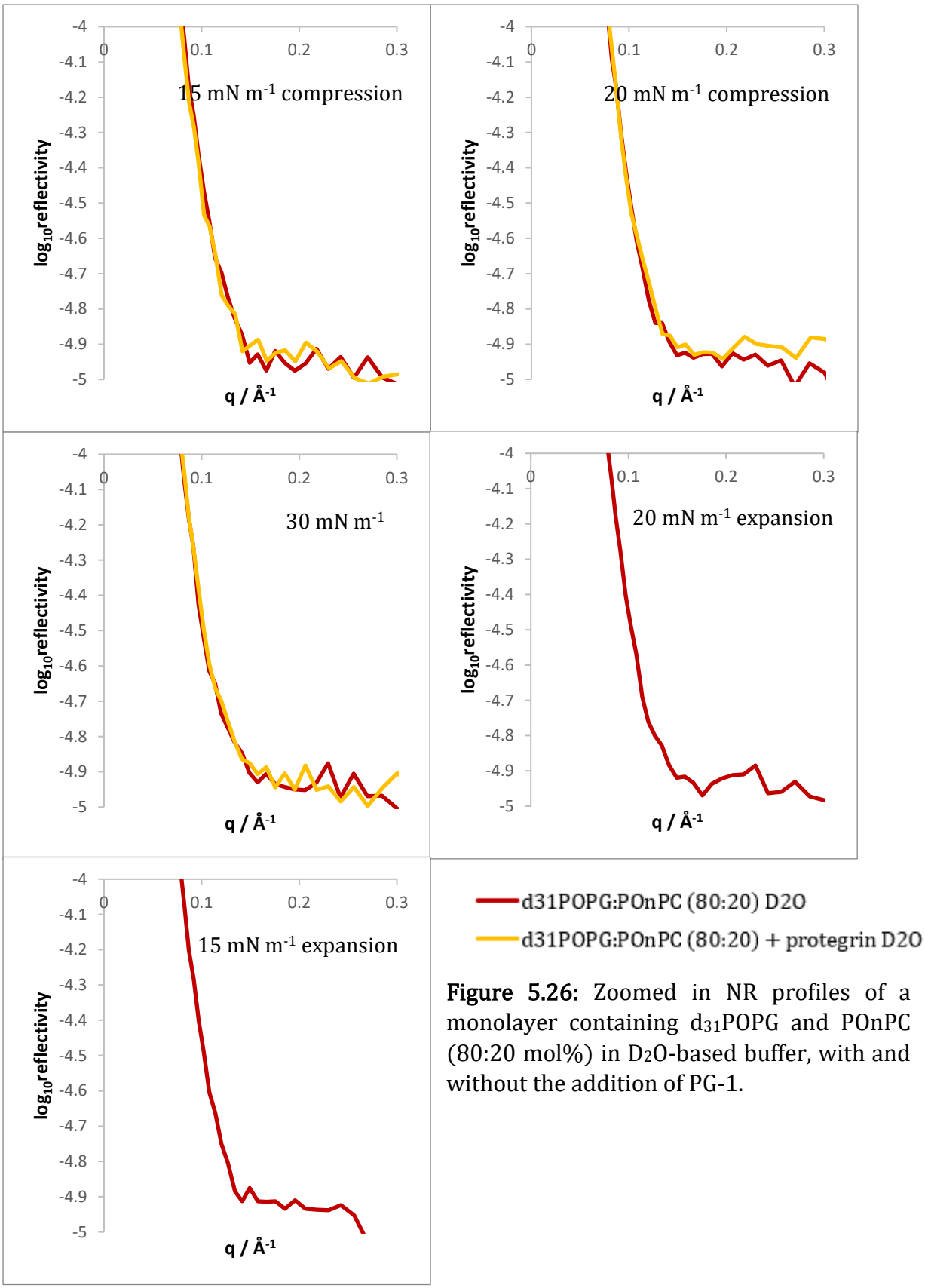


Figure 5.26: Zoomed in NR profiles of a monolayer containing d_{31}POPG and POnPC (80:20 mol%) in D_2O-based buffer, with and without the addition of PG-1.

d_{31}POPG with POnPC (80:20 mol%)								
Surface pressure / mN^{-1}	Thickness (tails) / Å	Sigma / Å	Thickness (heads) / Å	Sigma in / Å	Thickness (total) / Å	SLD (tails) / × 10^{-6} $Å^{-2}$	SLD (heads) / × 10^{-6} $Å^{-2}$	VFS (heads)
15	14.43	1.14	2.88	1.32	17.31	2.40	2.31	0.11
20	9.09	0.60	9.57	0.68	18.66	2.47	2.31	0.09
30	11.32	0.46	6.53	0.56	17.84	2.99	2.31	0.05
20	13.91	0.91	4.96	1.05	18.87	2.35	2.31	0.13
15	8.31	0.50	10.97	0.56	19.29	2.32	2.31	0.15
With PG-1:								
15	12.55	0.68	7.59	0.77	20.14	2.17	2.31	0.20
20	19.27	0.91	3.88	1.10	23.15	1.88	2.31	0.31
30	8.82	0.30	9.37	0.39	18.20	2.87	2.31	0.09

Table 5.15: Fitting results of the d_{31}POPG:POnPC (80:20 mol%) monolayers with and without PG-1 at the measured surface pressures, fitted using *refnx* and *motofit*.

Surface pressure / mN^{-1}	Percentage change in total layer thickness	Change in VFS
15	116	+0.09
20	124	+0.22
30	102	+0.04

Table 5.16: Percentage changes in the thickness and VFS (heads) between the d_{31}POPG:POnPC (80:20 mol%) monolayers with and without PG-1.

5.7 Discussion

This chapter aimed to investigate how the interactions between PG-1 and anionic PG lipids are affected by the presence of unoxidised and oxidised PC lipids (POPC and POnPC) in modelled bacterial cell membranes at different surface pressures, by comparing the differences in reflectivity between the modelled lipid membranes with and without the addition of PG-1 and looking at the changes in the fitted lipid thickness and VFS of the lipid heads.

NR was used to study precise changes in lipid thickness and lipid head VFS in the monolayers. The addition of PG-1 to the d_{62}DPPG, d_{62}DPPG:POPC (80:20 mol%) and d_{31}POPG monolayers in NRW-based buffer led to a decrease in reflectivity compared to the monolayers with no PG-1 added. No significant change in reflectivity was observed in all monolayers recorded in D_2O-based buffer, when POPC or POnPC was incorporated into the lipid mixtures.

Overall, the data shows that PG-1 embeds into the lipid membranes. As can be seen in Figure 5.27, the addition of PG-1 to the lipid monolayers tested showed an increase in the total thickness of the lipid layer. In the case of the mixed lipid monolayers, the increase in thickness was generally observed as the surface pressure was increased after the addition of the peptide. The increase in the thickness of the monolayer can most readily be interpreted as the peptide embedding in the lipid monolayer. This increase was larger when PG-1 is added to monolayers containing d_{31}POPG than d_{62}DPPG. In the case of mixed lipid layers, a larger initial increase in thickness was seen when the oxidised lipid POnPC was present in place of POPC. The difference between the POnPC and POPC monolayers are rather slight and should be interpreted with care as if the mixed monolayers were not homogenous. This means that the POnPC/POPC lipids may occur in small domains, whilst the peptide interacts mainly with the surrounding negatively charged PG lipids. In the case of the pure d_{31}POPG and d_{62}DPPG monolayers, although an initial increase in thickness was observed, when compressing the monolayer to give higher surface pressures of 30 mN m^{-1} slightly thinner monolayers were formed with PG-1, than without it. This may suggest that PG-1 was squeezed out of the monolayer, or it was changing the ordering of lipids in the monolayer. For instance, if PG-1 was lying flat in the monolayer, it would be thinner than the lipid tails.

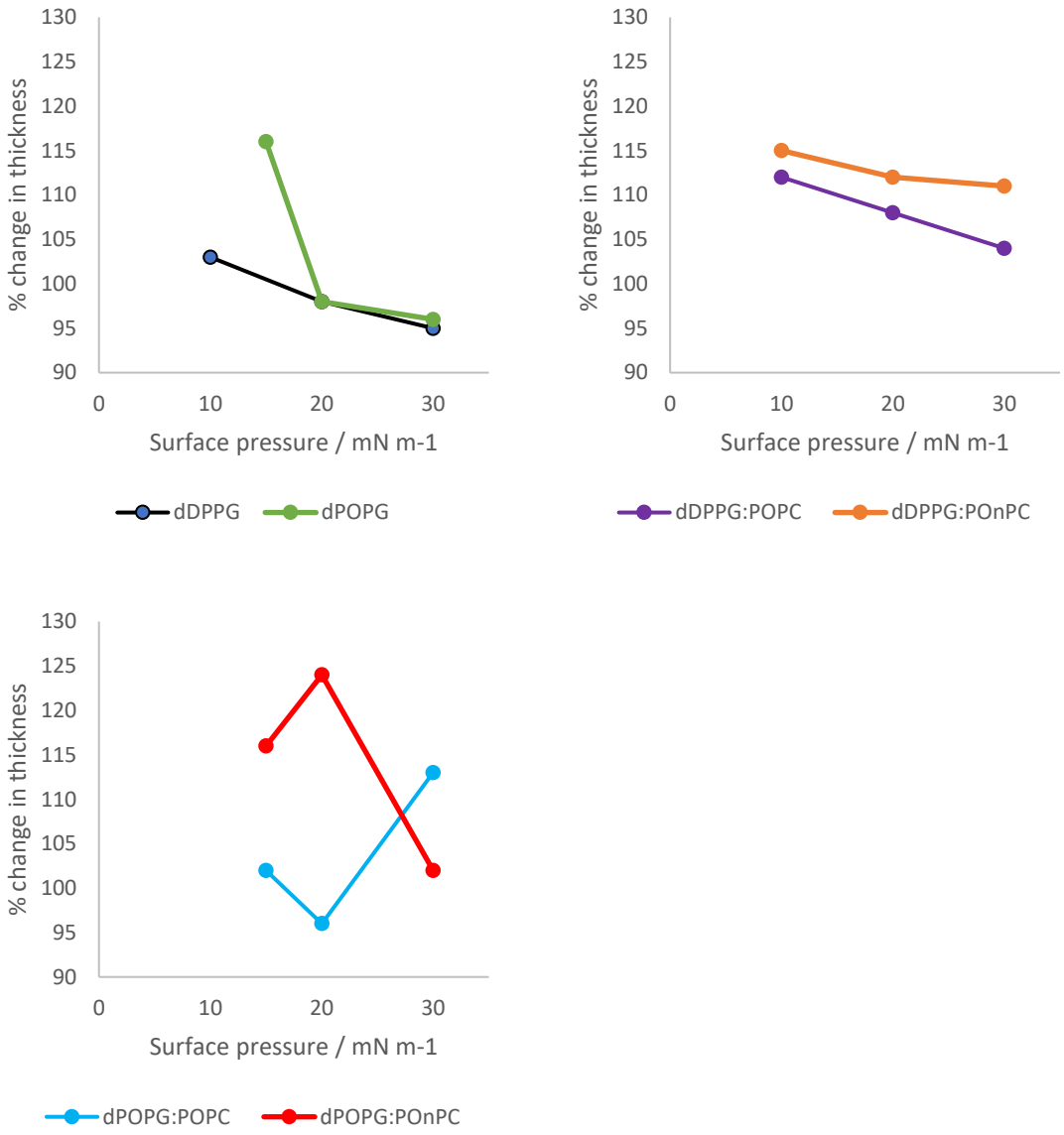


Figure 5.27: Percentage changes in total thickness of the lipid monolayers at the measured surface pressures. Top to bottom, left to right: d_{62}DPPG and d_{31}POPG, d_{62}DPPG:POPC (80:20 mol%) and d_{62}DPPG:POnPC, d_{31}POPG:POPC and d_{31}POPG:POnPC.

To fully understand the interactions of PG-1 with lipid monolayers, further experiments could be performed if a deuterated version of PG-1 was available to improve the reflectivity signal. Deuterated PG-1 may improve the signal for PG-1, where a smaller change in lipid thickness would suggest that the peptide is positioned flat across the membrane, and a larger change in thickness would imply that the peptide is positioned perpendicular to the normal of the monolayer. However, there are limitations to the types of residues that can be deuterated, and whether deuterated PG-1 is commercially available. For example, the N-terminal of lysine residues cannot be deuterated as the

deuterium switches back to hydrogen atoms in deuterated or NRW-based buffers. The neutron reflection data would also benefit from interpretation alongside complementary MD simulations similar to those previously described in Chapter 4.

Interpretation of the fitted VFS is not straightforward. It is unclear what happens with the VFS of the lipid heads, but the increase in VFS of the lipid heads may imply a decrease in the SLD of the heads, due to the presence of PG-1 associating in the head region. In most monolayers, the VFS decreases with increasing surface pressure. For instance, the d_{62}DPPG monolayers, d_{62}DPPG:POPC (80:20 mol%) monolayers, d_{31}POPG monolayer with PG-1, and the d_{31}POPG:POPC (80:20 mol%) and d_{31}POPG:POnPC (80:20 mol%) monolayers without PG-1. An increase in VFS may not imply an increase in the amount of solvent in the lipid heads.

The SLD of PG-1 is lower than the SLD of the lipid heads, which was 2.31×10^{-6} Å^{-2}, where the insertion of PG-1 would decrease the SLD of the lipid heads that in turn caused the fitting program to fit the data as having an increase in VFS. The decrease in SLD of the lipid heads could be due to more NRW-based buffer or hydrogenous peptide in the headgroup region. Compared to the NRW-based solvent, D_2O-based solvents have a high SLD, where the addition of PG-1 would cause a drop in VFS of the lipid heads, however, due to time constraints at the ISIS Neutron and Muon Source, not all datasets were fitted with both NRW and D_2O-based buffers.

It should be noted that there are different, and more sophisticated, approaches to fitting neutron reflectivity data, which may yield more accurate results. The reader should consult the fitting methods as described by Braun *et al.* and Niga *et al.* as alternatives to the methods described in this thesis (Braun et al., 2017; Niga et al., 2018).

5.8 Conclusions

There is evidence of interactions between PG-1 and the modelled lipid membranes, with structural reorganisation occurring. The presence of the oxidatively damaged lipid POnPC appeared to lead to slightly increased disruption caused by PG-1 than when POPC was present in the mixture, however, this change was modest. In future experiments, studying how PG-1 inserts itself into oxidatively damaged PG lipids by replacing POPG with POnPG, would be useful but POnPG is not readily commercially available. Techniques such as Brewster angle microscopy could be used in future work to visualise the presence of phase separation in lipid monolayers.

No definitive conclusion could be made on the exact positioning of PG-1 within the lipid membrane. Currently, the challenge would be to achieve highly precise analysis of the modelled lipid membranes. Ideally, it would be useful to correlate changes in reflectivity, thickness, or other parameters of the bacterial membrane, to the addition of PG-1. Additionally, determining the specific positioning of PG-1 relative to the bacterial membrane could reveal information on the mechanisms of PG-1 on the membrane's structure and dynamics. A large change in thickness between monolayers with and without PG-1 may imply that PG-1 inserts perpendicularly to the membrane, and a small change in thickness would imply that the peptide may be in a transmembrane position to the monolayer. Further experiments with deuterated PG-1 combined with complementary MD simulations could provide valuable insights. In addition to this, Brewster angle microscopy could be used to detect any changes to the phase separation of the lipids, since the presence of a condensed monolayer phase leads to a change in the refractive index and subsequently changes in reflectivity.

Compared to PG and PE lipids, PC lipids are less common in bacterial membranes. A future set of experiments could include inserting PG-1 into monolayers with d_{31}POnPG instead of d_{31}POPG, to investigate the effectiveness of PG-1 on oxidatively damaged PG lipids commonly found in bacterial membranes. Since d_{31}POnPG is not commercially available, the lipid would have to be subject to oxidative attack and this could not be done during the timeframe of this research.

Chapter 6 – Overall Conclusions and Future Work

Peptide mimics such as Mini-B and Super Mini-B have been successfully produced by peptide synthesis (Walther, Gordon, and Waring, 2016), however, many research attempts that have been made to express recombinant full length Sp-B in bacteria, have been largely without success. There were also attempts made by Waring, Baatz and Perez-Gil to ligate two peptide fragments together to synthesize full length Sp-B, however this had been unsuccessful (Asrat, 2021).

The work done in Chapter 2 aimed to investigate methods to express recombinant Sp-B using BL21 (DE3) *E. coli*. It was concluded that relatively pure samples of recombinant His6 Sp-B and Sp-B were successfully obtained through affinity purification alone. Further purification using size exclusion chromatography was deemed unnecessary, as most contaminants were removed during the initial purification. Furthermore, additional purification steps resulted in a higher loss of protein sample. Removal of the His6 tag by TEV caused Sp-B to be less stable in solution, due to its hydrophobicity and size. As a result, samples of Sp-B had to be kept at 4˚C and only at low concentrations below 0.1 mg mL^{-1}. It was discovered that lowering the pH of the buffer that both His6 Sp-B and Sp-B was kept in, prevented protein loss and precipitation, in cases where the samples had to be kept overnight at 4˚C or concentrated for mass spectrometry experiments. Generally, for small proteins below 10 kDa in size, a higher concentration of about 20-50 mg mL^{-1} is recommended for crystallisation studies (McPherson, 1990). The concentration of Sp-B obtained from work in Chapter 2 was insufficient for carrying out crystallography experiments. As explained earlier, the tendency of Sp-B to precipitate at high concentrations also poses a challenge to express it on a much larger scale to study its crystal structure.

CD spectroscopy experiments revealed that both His6 Sp-B and Sp-B would form predominantly alpha helical structures in solution containing 300 mM NaF. It also revealed that the residues on both proteins had the propensity to form a more helical structure when exposed to an environment mimicking a lipid membrane environment, as seen from the addition of increasing concentrations of TFE, which led to an increasing proportion of helical structure as the concentration of TFE was increased. This agrees with experimentally determined data from the PDB of Sp-B peptide mimics like $Sp\text{-}B_{1\text{-}25}$ and Mini-B (Gordon et al., 2000; Waring et al., 2005), results from structural predictions

of Sp-B by AlphaFold2 in Chapter 1, and the MD simulations on Sp-B_{1-25} described in Chapter 4.

Chapter 4 explored how a peptide mimic of Sp-B, Sp-B_{1-25}, positioned itself and interacted within lipid monolayers containing oxidised and unoxidised lipids. In all monolayers, an increase in APL led to a decrease in surface pressure and increasing disorder of the lipid tails, as expected and in agreement with previous MD simulations done by Baoukina and Olżyńska *et al.* (Baoukina et al., 2007; Olżyńska et al., 2016). It was also found that oxidative damage to unsaturated lipid tails significantly increased the compressibility of the monolayer and changed the profile of the surface pressure as a function of lipid area. The density profiles revealed that Sp-B_{1-25} gradually moves from being positioned parallel to the monolayer normal, between the headgroup and acyl chain regions, to embedding itself in the headgroup region as the monolayer expands. In the presence of oxidatively damaged lipids, it fully embeds in the headgroup region regardless of the APL. The presence of oxidised or unoxidised unsaturated lipids also impacted the helical structure of Sp-B_{1-25}, where oxidative damage to unsaturated lipid tails caused some unwinding of the alpha helix. This suggests that damage to the unsaturated lipid tails in turn, destabilises the structure of the Sp-B_{1-25} alpha helix, causing some residues in the peptide to preferentially form hydrogen bonds with the surrounding lipids.

Overall, the MD simulations revealed that oxidative damage to unsaturated surfactant lipids in a monolayer impacted surfactant function, which suggests potential implications for lung health when exposed to ozone. Further work could incorporate other surfactant lipids, such as neutral lipids, simulating a monolayer with a lipid composition similar to the one described in Chapter 1 and the literature, as well as investigating how cholesterol maintains membrane fluidity and influences the positioning of surfactant proteins and lipids. If an experimentally determined structure of full-length Sp-B is available, it would be a priority to understand the peptide-lipid and lipid-lipid interactions through simulations.

In Chapter 5, reflectivity data was used to investigate the effects of oxidative damage on the insertion of PG-1 into model bacterial membranes containing PG and PC lipids. It was found that the addition of PG-1 to different monolayers led to a decrease in reflectivity, which implies that PG-1 embedded into the lipid membranes. Generally, there was an increase in total lipid thickness, especially in the mixed monolayers, as the surface

pressure increased. Monolayers containing purely PG lipids showed an initial increase in thickness, with thinning at higher surface pressure upon the addition of PG-1, suggesting possible peptide reorganisation. The exact positioning of PG-1 within these lipid membranes remains inconclusive, which prompts future experiments with deuterated PG-1 if this was commercially available. Additionally, the presence of oxidative damage in PC lipids only modestly changed the insertion of PG-1 into the membrane. This showed that PG-1 did not preferentially interact with POPC or POnPC lipids.

In conclusion, this research was able to produce a recombinant form of Sp-B and confirming that the expressed protein had a predominantly alpha helical structure. However, the protocol was unable to produce large amounts of Sp-B due to its hydrophobic nature, potentially making a focus of future work to scale up the production of recombinant Sp-B. Although the MD simulations had limited availability to full-length Sp-B, future simulations could incorporate other lipids such as PI and neutral lipids, subsequently looking at the effects of oxidative damage on other unsaturated lipids in the monolayer. Simulating the parts of the unsaturated lipid tails cleaved off by oxidation would help in understanding whether they have an impact on the positioning and interactions with other lipids and peptides in the monolayer. Additionally, including the experimental structures of other surfactant proteins in the simulation may shed further insights into the peptide-lipid interactions in lung surfactant. Since POnPG was not a commercially available product, this restricted the understanding of how damage to both PG and PC lipids in bacterial membranes may affect PG-1 insertion. Similarly, the lack of deuterated PG-1 availability meant that the degree and positioning of PG-1 insertion into the modelled membranes could not be studied in detail. To supplement the research done in Chapter 5, future work could include further MD simulation experiments of PG-1 in model bacterial membranes, like those done in Chapter 4, to understand how PG-1 inserts into the membrane as well as the peptide-lipid interactions in the presence or absence of oxidative damage.

Chapter 7 Appendix

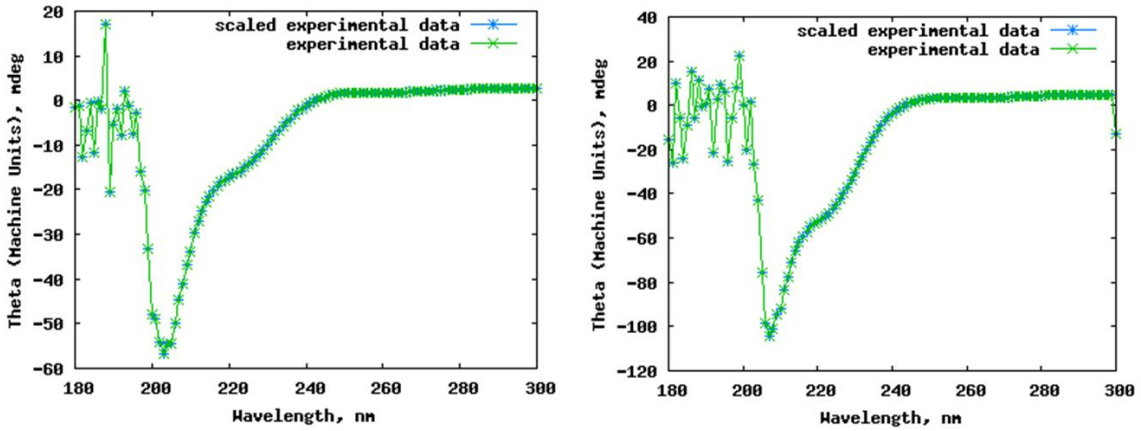


Figure 7.1: CD spectra of the scaled experimental data and experimental data of diluted His6 Sp-B (left) and Sp-B (right) generated in DichroWeb.

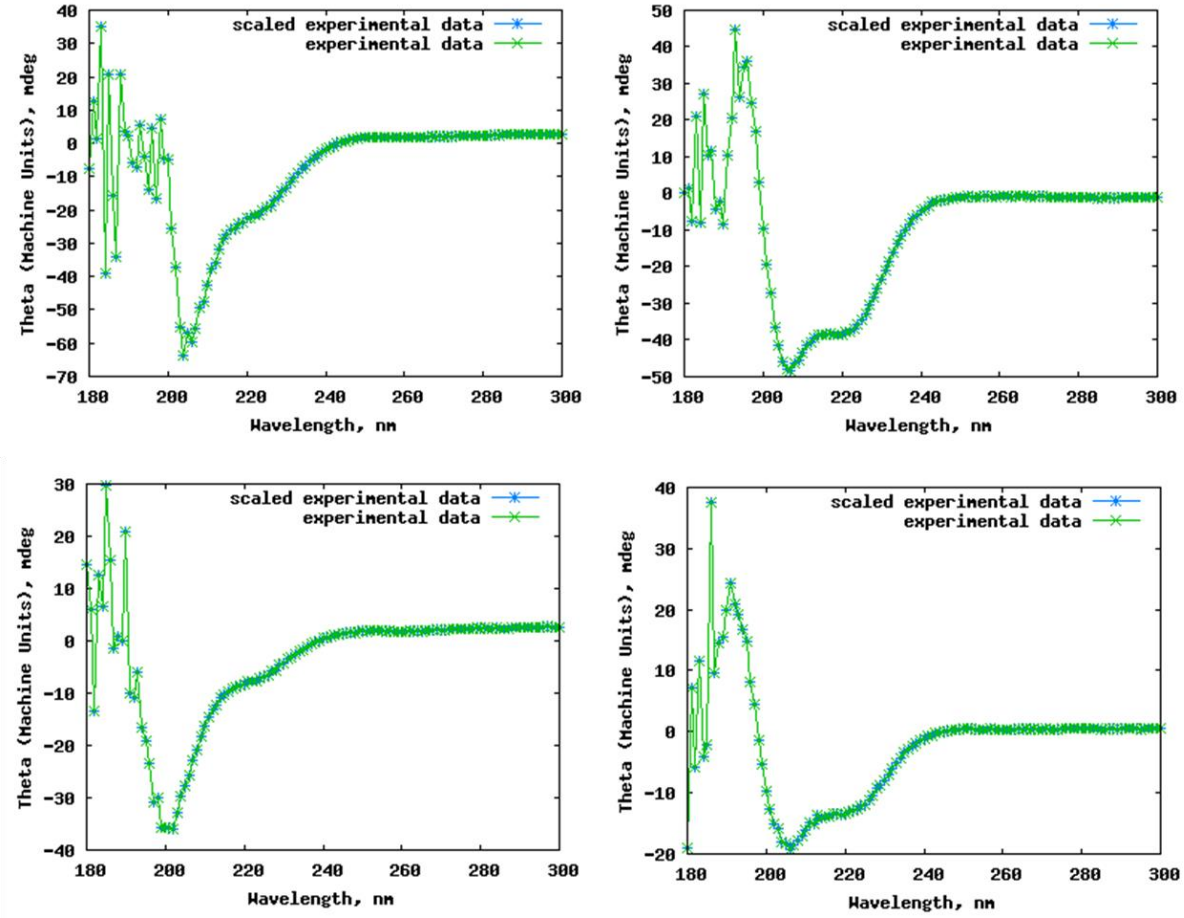


Figure 7.2: CD spectra of the scaled experimental data and experimental data of His6 Sp-B and Sp-B, with and without TFE, generated in DichroWeb. Top left: His6 Sp-B. Top right: His6 Sp-B with 40% TFE. Bottom left: Sp-B. Bottom right: Sp-B with 40% TFE.

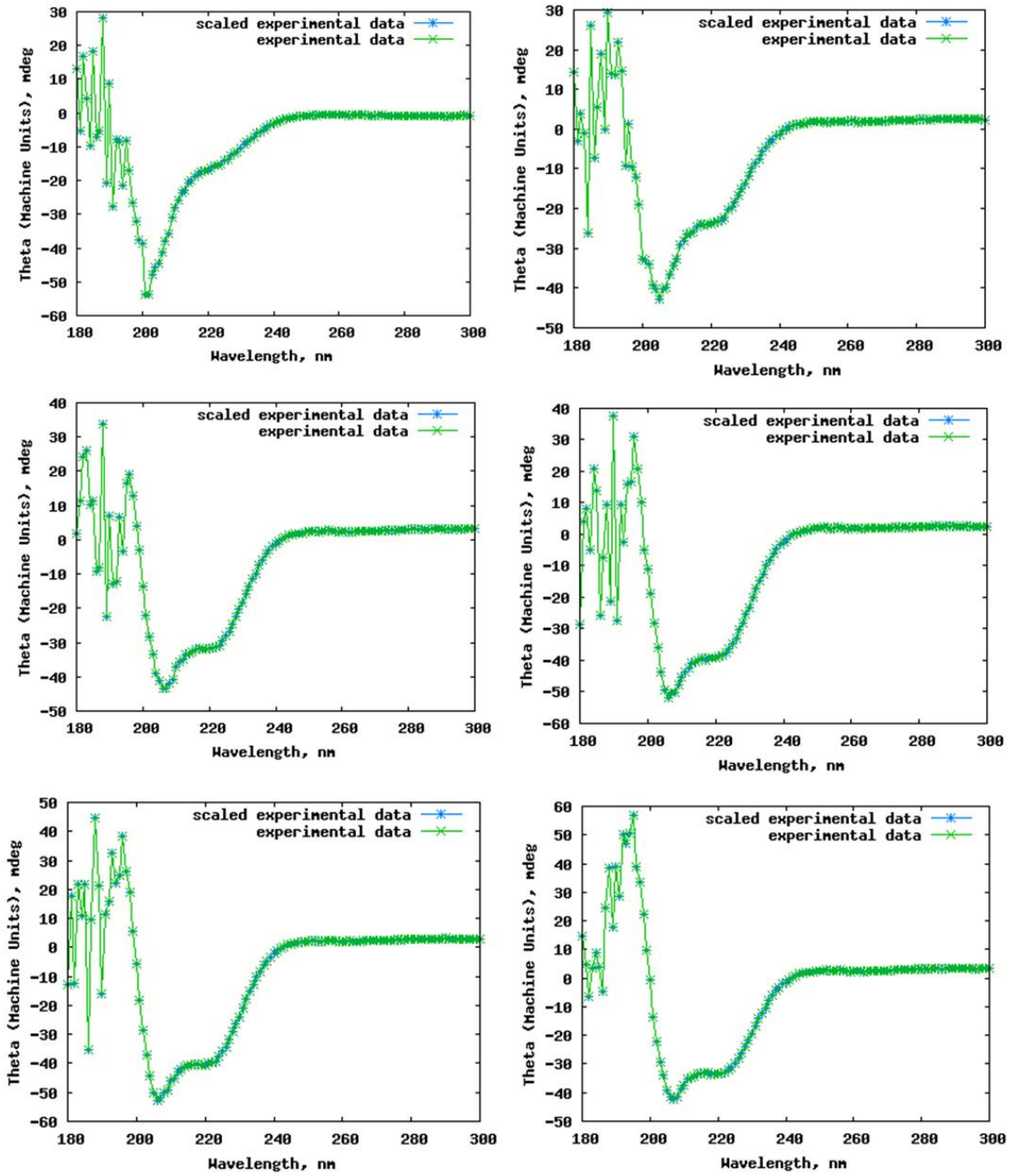


Figure 7.3: CD spectra of the scaled experimental data and experimental data of His6 Sp-B, with varying concentrations of added TFE, generated in DichroWeb. **Left to right, top to bottom:** His6 Sp-B + 0% TFE, His6 Sp-B + 10% TFE, His6 Sp-B + 20% TFE, His6 Sp-B + 30% TFE, His6 Sp-B + 40% TFE, His6 Sp-B + 50% TFE.

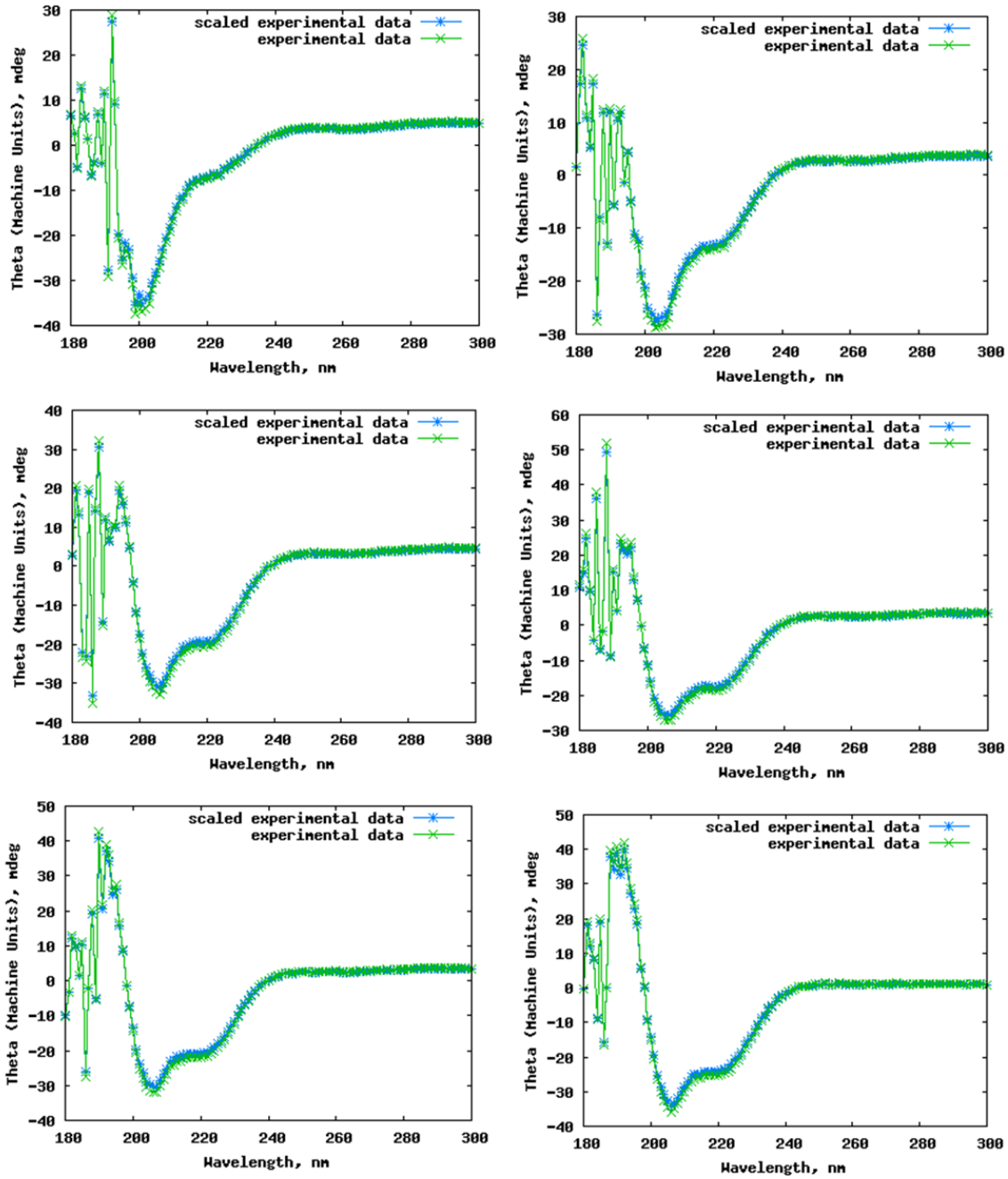


Figure 7.4: CD spectra of the scaled experimental data and experimental data of Sp-B, with varying concentrations of added TFE, generated in DichroWeb. **Left to right, top to bottom:** Sp-B + 0% TFE, Sp-B + 10% TFE, Sp-B + 20% TFE, Sp-B + 30% TFE, Sp-B + 40% TFE, Sp-B + 50% TFE.

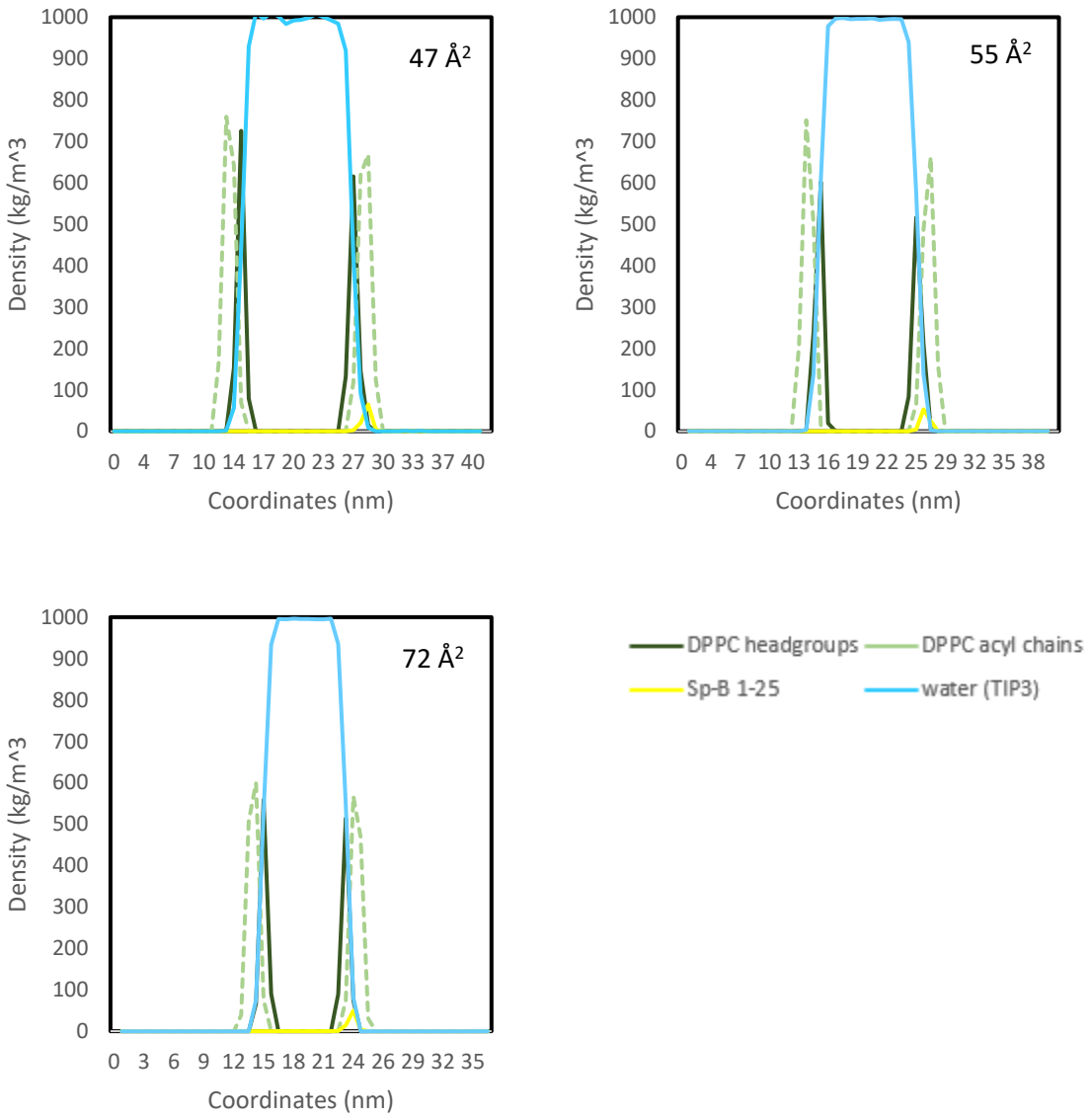


Figure 7.5: Full density profiles (upper and lower leaflets) of Sp-B_{1-25} in DPPC, measured at 47 Å², 55 Å² and 72 Å². Sp-B_{1-25} was simulated in the lower leaflet.

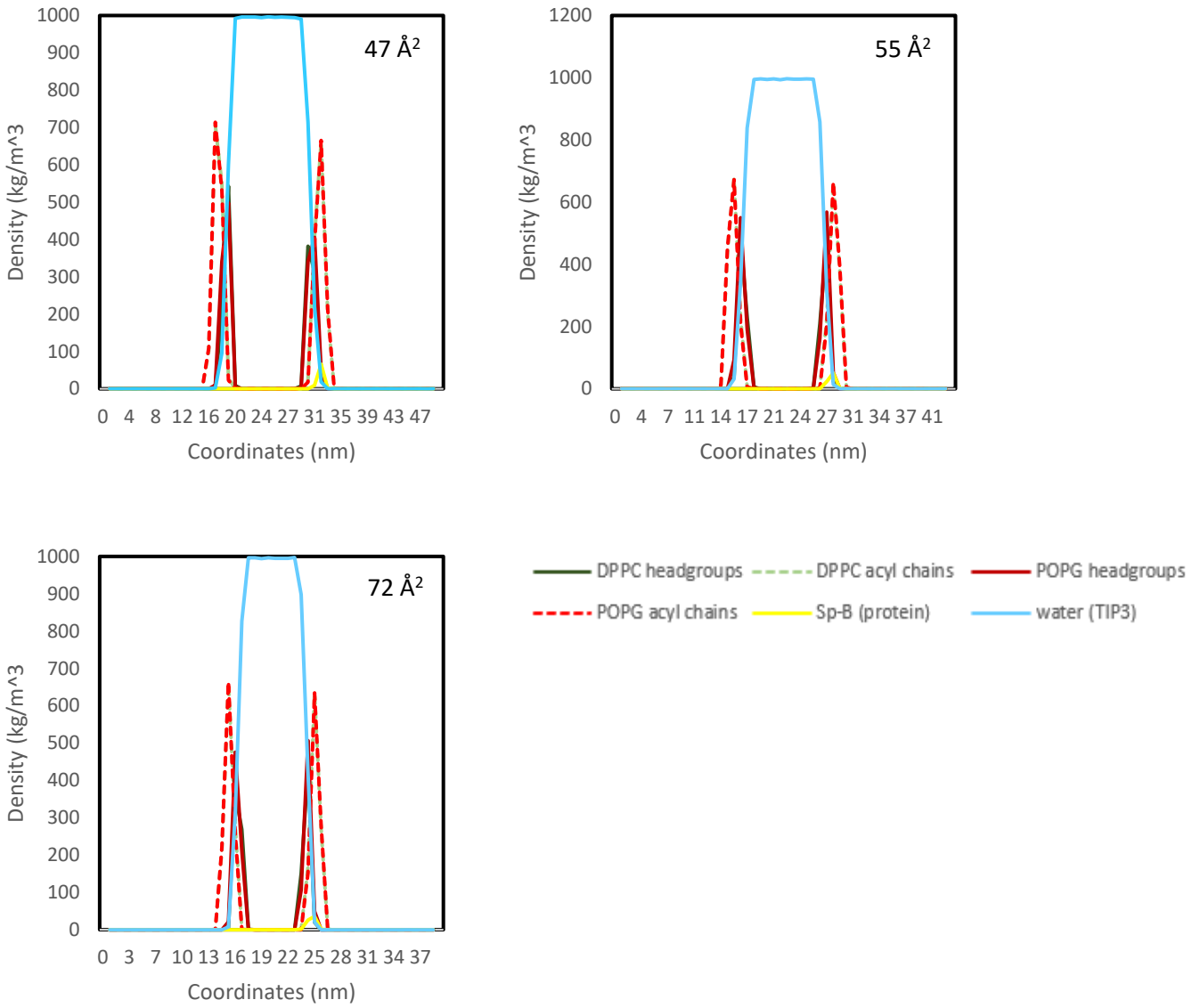


Figure 7.6: Full density profiles (upper and lower leaflets) of Sp-$B_{1\text{-}25}$ in DPPC:POPG, measured at 47 Å², 55 Å² and 72 Å². Sp-$B_{1\text{-}25}$ was simulated in the lower leaflet.

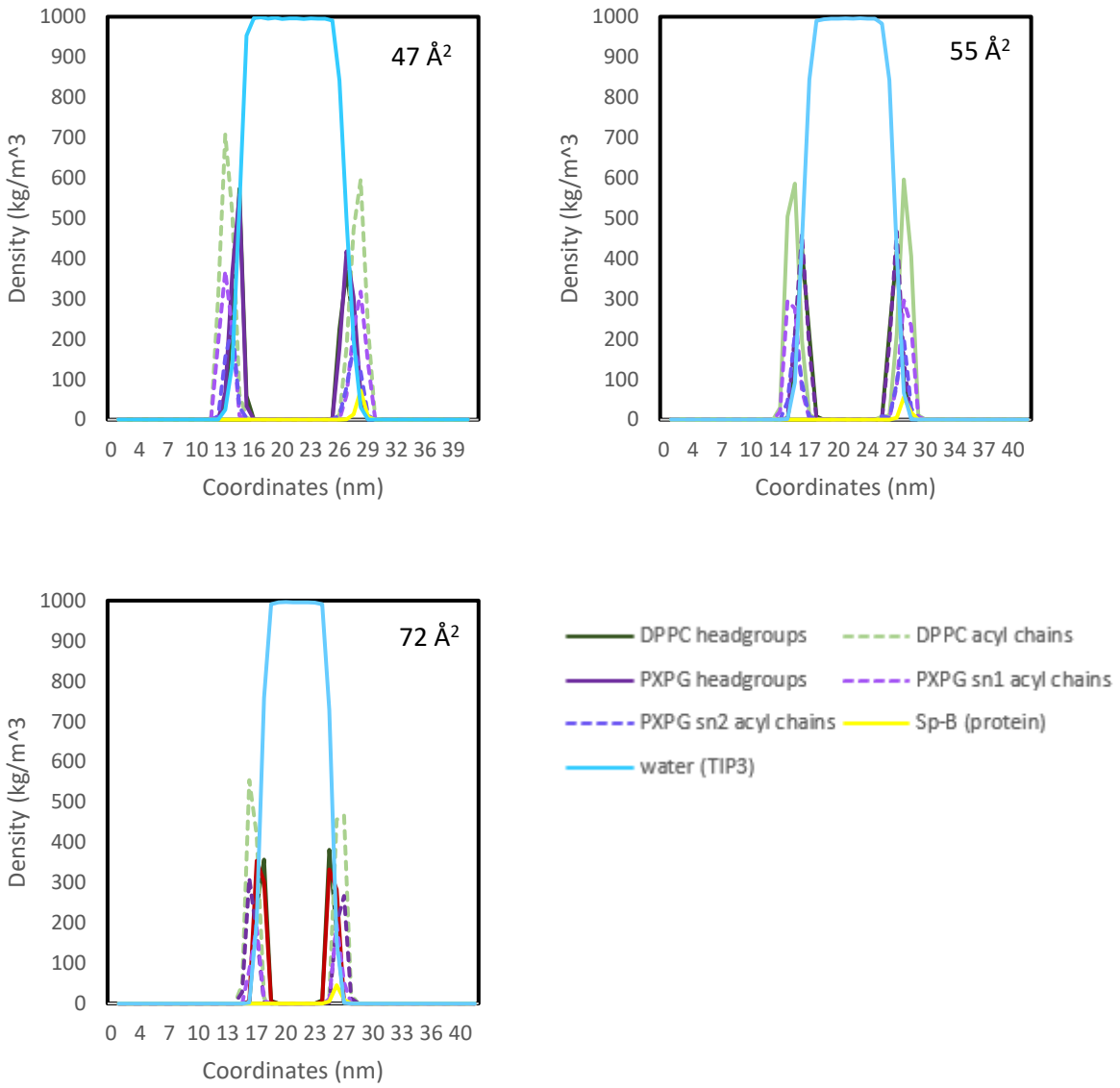


Figure 7.7: Full density profiles (upper and lower leaflets) of Sp-$B_{1\text{-}25}$ in DPPC:PXPG, measured at 47 Å², 55 Å² and 72 Å². Sp-$B_{1\text{-}25}$ was simulated in the lower leaflet.

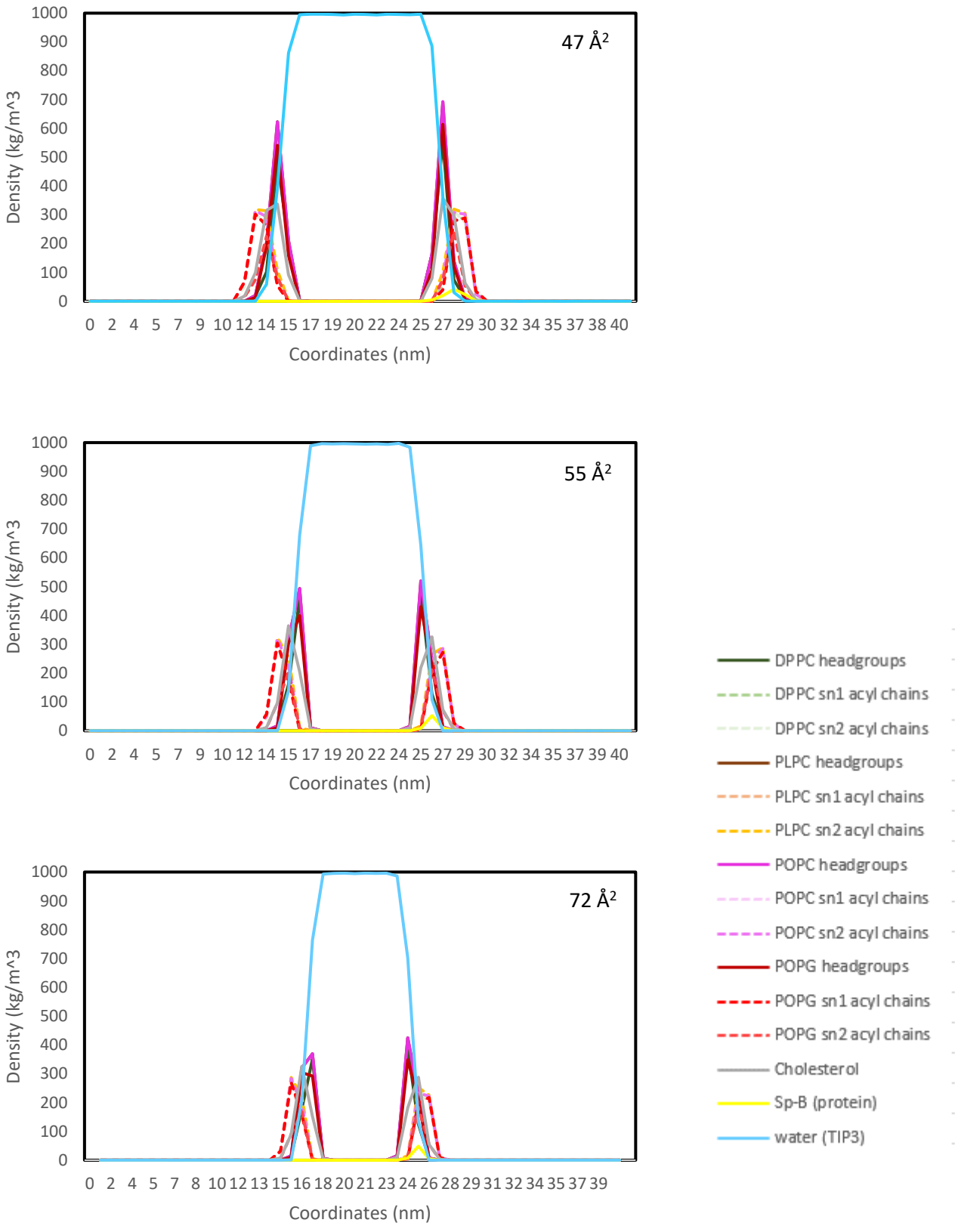


Figure 7.8: Full density profiles (upper and lower leaflets) of Sp-$B_{1\text{-}25}$ in the mixed unoxidised monolayer, measured at 47 Å², 55 Å² and 72 Å². Sp-$B_{1\text{-}25}$ was simulated in the lower leaflet.

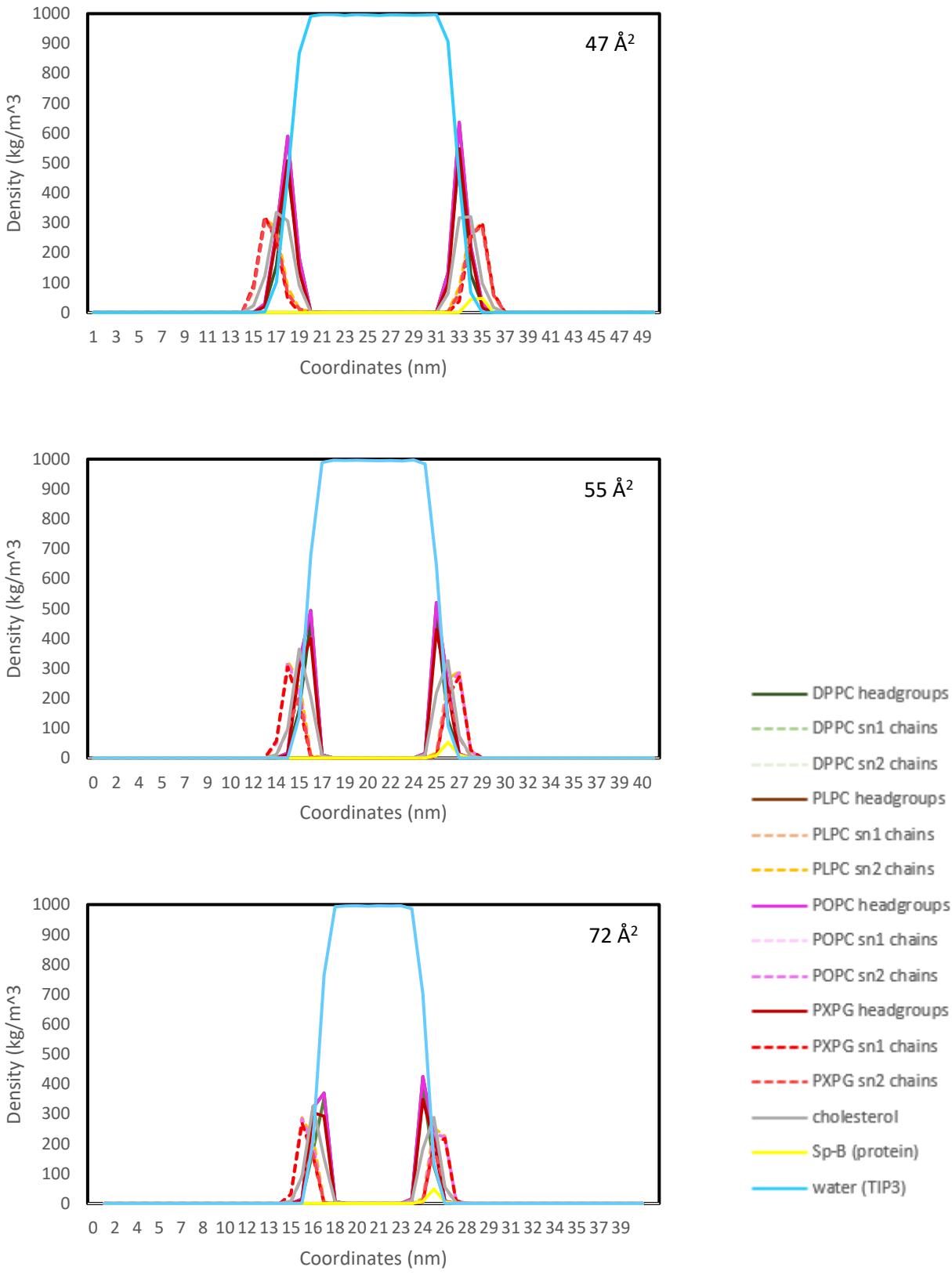
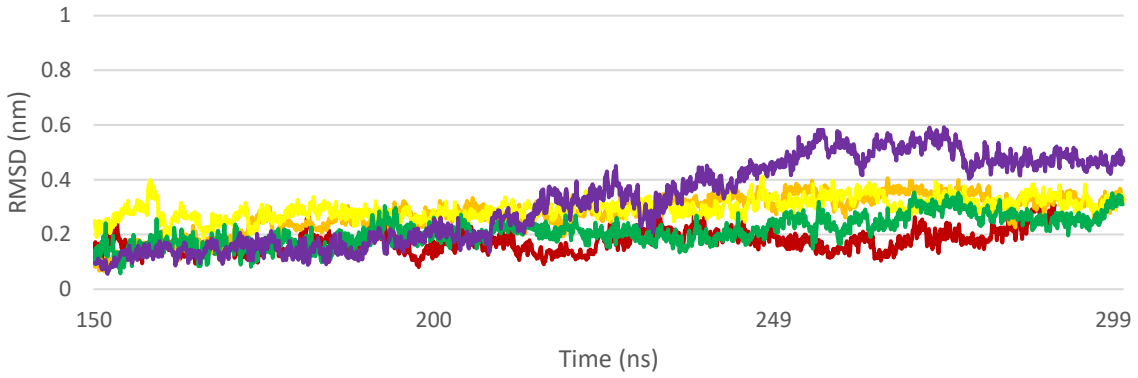
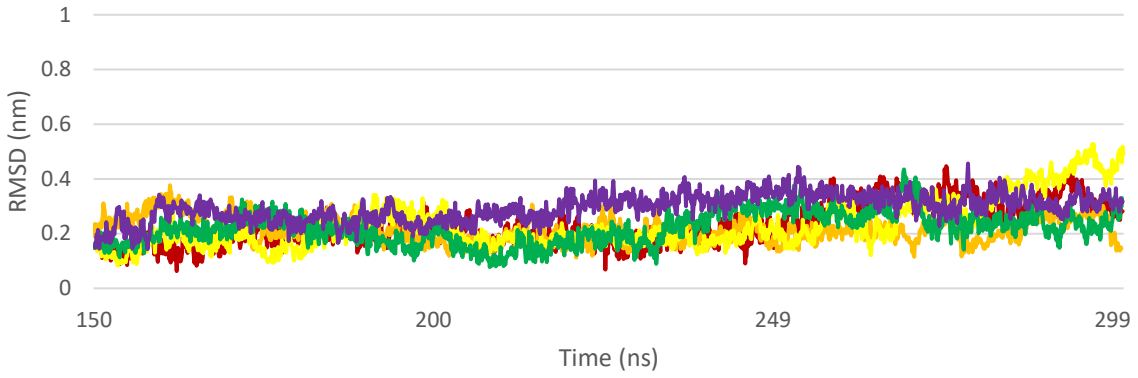


Figure 7.9: Full density profiles (upper and lower leaflets) of Sp-$B_{1\text{-}25}$ in the mixed oxidised monolayer, measured at 47 Å², 55 Å² and 72 Å². Sp-$B_{1\text{-}25}$ was simulated in the lower leaflet.

A



B



C

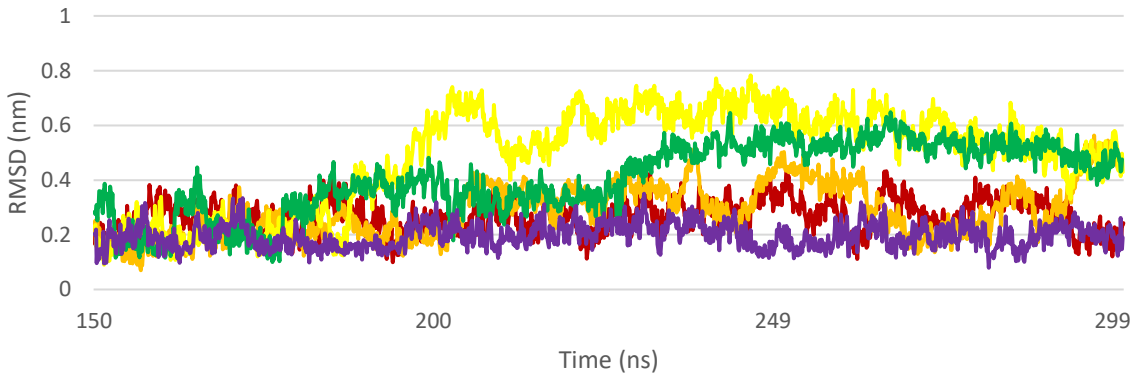


Figure 7.10: RMSD plots of the final 150 ns simulation of the monolayer systems at **(A)** 47 Å^2, **(B)** 55 Å^2 and **(C)** 72 Å^2.

References

Abdul-Gader, A., Miles, A.J. and Wallace, B.A. (2011) 'A reference dataset for the analyses of membrane protein secondary structures and transmembrane residuesusing circular dichroism spectroscopy', *Bioinformatics*, 27(12), pp. 1630–1636. Available at: https://doi.org/10.1093/bioinformatics/btr234.

Abraham, M.J. *et al.* (2015) 'Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers', *SoftwareX*, 1–2, pp. 19–25. Available at: https://doi.org/10.1016/j.softx.2015.06.001.

Agassandian, M. and Mallampalli, R.K. (2013) 'Surfactant phospholipid metabolism', *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*. NIH Public Access, pp. 612–625. Available at: https://doi.org/10.1016/j.bbalip.2012.09.010.

Ahn, V.E. *et al.* (2006) 'Crystal structures of saposins A and C', *Protein Science*, 15(8), pp. 1849–1857. Available at: https://doi.org/10.1110/ps.062256606.

Anderson, D.H. *et al.* (2003) 'Granulysin crystal structure and a structure-derived lytic mechanism', *Journal of Molecular Biology*, 325(2), pp. 355–365. Available at: https://doi.org/10.1016/S0022-2836(02)01234-2.

Anderson, I.S. *et al.* (2016) 'Research opportunities with compact accelerator-driven neutron sources', *Physics Reports*, 654, pp. 1–58. Available at: https://doi.org/10.1016/J.PHYSREP.2016.07.007.

Andersson, J.M. *et al.* (2017) 'Effect of cholesterol on the molecular structure and transitions in a clinical-grade lung surfactant extract', *Proceedings of the National Academy of Sciences of the United States of America*, 114(18), pp. E3592–E3601. Available at: https://doi.org/10.1073/pnas.1701239114.

Arbe, A. *et al.* (2020) 'Coherent structural relaxation of water from meso- To intermolecular scales measured using neutron spectroscopy with polarization analysis', *Physical Review Research*, 2(2), p. 022015. Available at: https://doi.org/10.1103/PHYSREVRESEARCH.2.022015/FIGURES/3/MEDIUM.

Arizumi, N. and Bond, S.D. (2012) 'On the estimation and correction of discretization error in molecular dynamics averages', *Applied Numerical Mathematics*, 62(12), pp. 1938–1953. Available at: https://doi.org/10.1016/j.apnum.2012.08.005.

Arnold, O. et al (2013) *Mantid-Data Analysis and Visualization Package for Neutron Scattering and mu-SR Experiments*. Available at: https://www.mantidproject.org/ (Accessed: 26 October 2023).

Asrat, T.G. (2021) 'The expression, purification and characterization of recombinant human lung surfactant protein-B'. Available at: https://doi.org/10.48336/WNEK-MN45.

Banfi, C. and Agostoni, P. (2016) 'Surfactant protein B: From biochemistry to its potential role as diagnostic and prognostic marker in heart failure', *International Journal of Cardiology*, 221, pp. 456–462. Available at: https://doi.org/10.1016/J.IJCARD.2016.07.003.

Baoukina, S. *et al.* (2007) 'Pressure-area isotherm of a lipid monolayer from molecular dynamics simulations', *Langmuir*, 23(25), pp. 12617–12623. Available at: https://doi.org/10.1021/la702286h.

Bauer, G.S. (2001) 'Physics and technology of spallation neutron sources', *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 463(3), pp. 505–543. Available at: https://doi.org/10.1016/S0168-9002(01)00167-X.

Beck, D.A.C., Armen, R.S. and Daggett, V. (2005) 'Cutoff size need not strongly influence molecular dynamics results for solvated polypeptides', *Biochemistry*, 44(2), pp. 609–616. Available at: https://doi.org/10.1021/bi0486381.

Beck, D.C. *et al.* (2000) 'The role of homodimers in surfactant protein B function in vivo', *Journal of Biological Chemistry*, 275(5), pp. 3365–3370. Available at: https://doi.org/10.1074/jbc.275.5.3365.

Bilek, A.M., Dee, K.C. and Gaver, D.P. (2003) 'Mechanisms of surface-tension-induced epithelial cell damage in a model of pulmonary airway reopening', *Journal of Applied Physiology*, 94(2), pp. 770–783. Available at: https://doi.org/10.1152/japplphysiol.00764.2002.

Böhm, G., Muhr, R. and Jaenicke, R. (1992) 'Quantitative analysis of protein far UV circular dichroism spectra by neural networks', *Protein Engineering, Design and Selection*, 5(3), pp. 191–195. Available at: https://doi.org/10.1093/protein/5.3.191.

Bornhorst, J.A. and Falke, J.J. (2000) '[16] Purification of Proteins Using Polyhistidine Affinity Tags', *Methods in enzymology*, 326, p. 245. Available at: https://doi.org/10.1016/S0076-6879(00)26058-8.

Bradley, R. and Radhakrishnan, R. (2013) 'Coarse-grained models for protein-cell membrane interactions', *Polymers*, 5(3), pp. 890–936. Available at: https://doi.org/10.3390/polym5030890.

Brahms, S. and Brahms, J. (1980) 'Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism', *Journal of Molecular Biology*, 138(2), pp. 149–178. Available at: https://doi.org/10.1016/0022-2836(80)90282-X.

Braun, L. *et al.* (2017) 'Polymers and surfactants at fluid interfaces studied with specular neutron reflectometry', *Advances in Colloid and Interface Science*, 247, pp. 130–148. Available at: https://doi.org/10.1016/J.CIS.2017.07.005.

Brooks, C.L. (1989) 'Computer simulation of liquids', *Journal of Solution Chemistry*, 18(1), pp. 99–99. Available at: https://doi.org/10.1007/bf00646086.

Bruhn, H. *et al.* (2003) 'Amoebapores and NK-lysin, members of a class of structurally distinct antimicrobial and cytolytic peptides from protozoa and mammals: A comparative functional analysis', *Biochemical Journal*, 375(3), pp. 737–744. Available at: https://doi.org/10.1042/BJ20030250.

Bruhn, H. (2005) 'A short guided tour through functional and structural features of saposin-like proteins', *Biochemical Journal*. Portland Press Ltd, pp. 249–257. Available at: https://doi.org/10.1042/BJ20050051.

Chelulei Cheison, S. *et al.* (2011) 'Analysis of the effect of temperature changes combined with different alkaline pH on the β-lactoglobulin trypsin hydrolysis pattern using MALDI-TOF-MS/MS', *Journal of agricultural and food chemistry*, 59(5), pp. 1572–1581. Available at: https://doi.org/10.1021/JF1039876.

Christmann, U. *et al.* (2009) 'Role of Lung Surfactant in Respiratory Disease: Current Knowledge in Large Animal Medicine', *Journal of Veterinary Internal Medicine*, 23(2), pp. 227–242. Available at: https://doi.org/10.1111/J.1939-1676.2008.0269.X.

Colmorten, K.B., Nexoe, A.B. and Sorensen, G.L. (2019) 'The Dual Role of Surfactant Protein-D in Vascular Inflammation and Development of Cardiovascular Disease', *Frontiers in Immunology*, 10, p. 480436. Available at: https://doi.org/10.3389/FIMMU.2019.02264/BIBTEX.

Cousin, F. and Fadda, G. (2020) 'An introduction to neutron reflectometry', *EPJ Web of Conferences*, 236, p. 04001. Available at: https://doi.org/10.1051/EPJCONF/202023604001.

Cruz, A., Casals, C. and Perez-Gil, J. (1995) 'Conformational flexibility of pulmonary surfactant proteins SP-B and SP-C, studied in aqueous organic solvents', *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*, 1255(1), pp. 68–76. Available at: https://doi.org/10.1016/0005-2760(94)00210-P.

Cunha, K.C. *et al.* (2015) 'Assessing protein conformational sampling and structural stability via de novo design and molecular dynamics simulations', *Biopolymers*, 103(6), pp. 351–361. Available at: https://doi.org/10.1002/bip.22626.

Desjardins, P., Hansen, J.B. and Allen, M. (2010) 'Microvolume protein concentration determination using the NanoDrop 2000c Spectrophotometer', *Journal of Visualized Experiments*, (33), p. 33. Available at: https://doi.org/10.3791/1610.

Domon, B. and Aebersold, R. (2006) 'Mass spectrometry and protein analysis', *Science*. Science, pp. 212–217. Available at: https://doi.org/10.1126/science.1124619.

Dulbecco, R. and Vogt, M. (1954) 'Plaque formation and isolation of pure lines with poliomyelitis viruses', *The Journal of experimental medicine*, 99(2), pp. 167–182. Available at: https://doi.org/10.1084/JEM.99.2.167.

Van Dyke, M.W., Sirito, M. and Sawadogo, M. (1992) 'Single-step purification of bacterially expressed polypeptides containing an oligo-histidine domain', *Gene*, 111(1), pp. 99–104. Available at: https://doi.org/10.1016/0378-1119(92)90608-R.

Dyson, M.R. *et al.* (2004) 'Production of soluble mammalian proteins in Escherichia coli: Identification of protein features that correlate with successful expression', *BMC Biotechnology*, 4(1), p. 32. Available at: https://doi.org/10.1186/1472-6750-4-32.

Eite Tiesinga *et al.* (2019) *The NIST Reference on Constants, Units and Uncertainty*. Available at: https://physics.nist.gov/cuu/Constants/bibliography.html (Accessed: 24 October 2023).

Feller, S.E. *et al.* (1997) 'Molecular dynamics simulation of unsaturated lipid bilayers at low hydration: parameterization and comparison with diffraction studies', *Biophysical journal*, 73(5), pp. 2269–2279. Available at: https://doi.org/10.1016/S0006-3495(97)78259-6.

Feller, S.E., Gawrisch, K. and MacKerell, A.D. (2002) 'Polyunsaturated fatty acids in lipid bilayers: Intrinsic and environmental contributions to their unique physical properties', *Journal of the American Chemical Society*, 124(2), pp. 318–326. Available at: https://doi.org/10.1021/JA0118340/SUPPL_FILE/JA0118340_S.PDF.

Feller, S.E. and MacKerell, A.D. (2000) 'An improved empirical potential energy function for molecular simulations of phospholipids', *Journal of Physical Chemistry B*, 104(31), pp. 7510–7515. Available at: https://doi.org/10.1021/JP0007843/SUPPL_FILE/JP0007843_SA.PDF.

Gasteiger, E. *et al.* (2003) 'ExPASy: The proteomics server for in-depth protein knowledge and analysis', *Nucleic acids research*, 31(13), pp. 3784–3788. Available at: https://doi.org/10.1093/NAR/GKG563.

Gidalevitz, D. *et al.* (2003) 'Interaction of antimicrobial peptide protegrin with biomembranes', *Proceedings of the National Academy of Sciences*, 100(11), pp. 6302–6307. Available at: https://doi.org/10.1073/PNAS.0934731100.

Gordon, L.M. *et al.* (2000) 'Conformational mapping of the N-terminal segment of surfactant protein B in lipid using 13C-enhanced fourier transform infrared spectroscopy', *Journal of Peptide Research*, 55(4), pp. 330–347. Available at: https://doi.org/10.1034/J.1399-3011.2000.00693.X.

Greenfield, N.J. (1996) 'Methods to estimate the conformation of proteins and polypeptides from circular dichroism data', *Analytical Biochemistry*. Academic Press Inc., pp. 1–10. Available at: https://doi.org/10.1006/abio.1996.0084.

Griese, M. *et al.* (2016) 'Surfactant proteins in pediatric interstitial lung disease', *Pediatric Research*, 79(1), pp. 34–41. Available at: https://doi.org/10.1038/pr.2015.173.

Grund, F. (1979) 'Forsythe, G. E. / Malcolm, M. A. / Moler, C. B., Computer Methods for Mathematical Computations. Englewood Cliffs, New Jersey 07632. Prentice Hall, Inc., 1977. XI, 259 S', *ZAMM - Zeitschrift für Angewandte Mathematik und Mechanik*, 59(2), pp. 141–142. Available at: https://doi.org/10.1002/zamm.19790590235.

Guo, C.J. *et al.* (2008) 'S-nitrosylation of surfactant protein-D controls inflammatory function', *PLoS Biology*, 6(11), pp. 2414–2423. Available at: https://doi.org/10.1371/journal.pbio.0060266.

Guvench, O. and MacKerell, A.D. (2008) 'Comparison of protein force fields for molecular dynamics simulations', *Methods in Molecular Biology*, 443, pp. 63–88. Available at: https://doi.org/10.1007/978-1-59745-177-2_4/TABLES/1.

Haiko, J. *et al.* (2009) 'Invited review: Breaking barriers — attack on innate immune defences by omptin surface proteases of enterobacterial pathogens', *http://dx.doi.org/10.1177/1753425909102559*, 15(2), pp. 67–80. Available at: https://doi.org/10.1177/1753425909102559.

Han, S.H. and Mallampalli, R.K. (2015) 'The role of surfactant in lung disease and host defense against pulmonary infections', *Annals of the American Thoracic Society*, 12(5), pp. 765–774. Available at: https://doi.org/10.1513/ANNALSATS.201411-507FR/SUPPL_FILE/DISCLOSURES.PDF.

Hecht, O. *et al.* (2004) 'Solution Structure of the Pore-forming Protein of Entamoeba histolytica', *Journal of Biological Chemistry*, 279(17), pp. 17834–17841. Available at: https://doi.org/10.1074/jbc.M312978200.

Heenan, R. (2010) *ISIS Neutron Training Course Reflectivity Practical Worksheet*. Available at: www.isis.stfc.ac.uk (Accessed: 26 October 2023).

Hernández, G., Anderson, J.S. and Lemaster, D.M. (2012) 'Experimentally assessing molecular dynamics sampling of the protein native state conformational distribution', *Biophysical Chemistry*, 163–164, pp. 21–34. Available at: https://doi.org/10.1016/j.bpc.2012.02.002.

Hoover, W.G. (1985) 'Canonical dynamics: Equilibrium phase-space distributions', *Physical Review A*, 31(3), pp. 1695–1697. Available at: https://doi.org/10.1103/PhysRevA.31.1695.

Horibe, H. and Goto, Y. (2012) 'Development of an environmentally friendly resist-removal process using wet ozone', *International Journal of Polymer Science*, 2012. Available at: https://doi.org/10.1155/2012/937928.

Humphrey, W., Dalke, A. and Schulten, K. (1996) 'VMD: Visual molecular dynamics', *Journal of Molecular Graphics*, 14(1), pp. 33–38. Available at: https://doi.org/10.1016/0263-7855(96)00018-5.

Huynh, L. *et al.* (2014) 'Structural properties of POPC monolayers under lateral compression: Computer simulations analysis', *Langmuir*, 30(2), pp. 564–573. Available at: https://doi.org/10.1021/LA4043809/SUPPL_FILE/LA4043809_SI_002.AVI.

Ikegami, M. *et al.* (2001) 'The Collagen-like Region of Surfactant Protein A (SP-A) Is Required for Correction of Surfactant Structural and Functional Defects in the SP-A Null Mouse', *Journal of Biological Chemistry*, 276(42), pp. 38542–38548. Available at: https://doi.org/10.1074/jbc.M102054200.

Ikegami, M., Takabatake, N. and Weaver, T.E. (2002) 'Intersubunit disulfide bridge is not required for the protective role of SP-B against lung inflammation', *Journal of Applied Physiology*, 93(2), pp. 505–511. Available at: https://doi.org/10.1152/japplphysiol.01137.2001.

Jen Chen, I., Yin, D. and MacKerell, A.D. (2002) 'Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds', *Journal of Computational Chemistry*, 23(2), pp. 199–213. Available at: https://doi.org/10.1002/JCC.1166.

Jeong, H., Kim, H.J. and Lee, S.J. (2016) 'Complete genome sequence of Escherichia coli strain BL21', *Genome Announcements*, 3(2). Available at: https://doi.org/10.1128/genomeA.00134-15.

Johansson, J. (1998) 'Structure and properties of surfactant protein C', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier, pp. 161–172. Available at: https://doi.org/10.1016/S0925-4439(98)00065-9.

Johnson, W.C. (1999) 'Analyzing protein circular dichroism spectra for accurate secondary structures', *Proteins: Structure, Function and Genetics*, 35(3), pp. 307–312. Available at: https://doi.org/10.1002/(SICI)1097-0134(19990515)35:3<307::AID-PROT4>3.0.CO;2-3.

Jumper, J. *et al.* (2021) 'Highly accurate protein structure prediction with AlphaFold', *Nature 2021 596:7873*, 596(7873), pp. 583–589. Available at: https://doi.org/10.1038/s41586-021-03819-2.

Kandasamy, S.K. and Larson, R.G. (2005) 'Molecular Dynamics Study of the Lung Surfactant Peptide SP-B1–25 with DPPC Monolayers: Insights into Interactions and Peptide Position and Orientation', *Biophysical Journal*, 88(3), pp. 1577–1592. Available at: https://doi.org/10.1529/BIOPHYSJ.104.038430.

Khatami, M.H., Saika-Voivod, I. and Booth, V. (2016) 'All-atom molecular dynamics simulations of lung surfactant protein B: Structural features of SP-B promote lipid reorganization', *Biochimica et Biophysica Acta - Biomembranes*, 1858(12), pp. 3082–3092. Available at: https://doi.org/10.1016/j.bbamem.2016.09.018.

Kim, K.H. *et al.* (2013) 'Effect of cholesterol nanodomains on monolayer morphology and dynamics', *Proceedings of the National Academy of Sciences of the United States of America*, 110(33), pp. E3054–E3060. Available at: https://doi.org/10.1073/pnas.1303304110.

King, R.J. and MacBeth, M.C. (1981) 'Interaction of the lipid and protein components of pulmonary surfactant Role of phosphatidylglycerol and calcium', *BBA - Biomembranes*, 647(2), pp. 159–168. Available at: https://doi.org/10.1016/0005-2736(81)90242-X.

King, S.D. and Chen, S.Y. (2020) 'Recent progress on surfactant protein A: Cellular function in lung and kidney disease development', *American Journal of Physiology - Cell Physiology*, 319(2), pp. C316–C320. Available at: https://doi.org/10.1152/AJPCELL.00195.2020.

Klont, F. *et al.* (2019) 'Quantification of surfactant protein D (SPD) in human serum by liquid chromatography-mass spectrometry (LC-MS)', *Talanta*, 202, pp. 507–513. Available at: https://doi.org/10.1016/J.TALANTA.2019.05.028.

Kmiecik, S. *et al.* (2016) 'Coarse-Grained Protein Models and Their Applications', *Chemical Reviews*. American Chemical Society, pp. 7898–7936. Available at: https://doi.org/10.1021/acs.chemrev.6b00163.

Kröner, C. *et al.* (2015) 'Genotype alone does not predict the clinical course of SFTPC deficiency in paediatric patients', *European Respiratory Journal*, 46(1), pp. 197–206. Available at: https://doi.org/10.1183/09031936.00129414.

Kurt, O.K., Zhang, J. and Pinkerton, K.E. (2016) 'Pulmonary Health Effects of Air Pollution', *Current opinion in pulmonary medicine*, 22(2), p. 138. Available at: https://doi.org/10.1097/MCP.0000000000000248.

Kuwajima, K. (1996) 'Stopped-Flow Circular Dichroism', in *Circular Dichroism and the Conformational Analysis of Biomolecules*. Springer US, pp. 159–182. Available at: https://doi.org/10.1007/978-1-4757-2508-7_5.

De La Serna, J.B. *et al.* (2004) 'Cholesterol rules: Direct observation of the coexistence of two fluid phases in native pulmonary surfactant membranes at physiological temperatures', *Journal of Biological Chemistry*, 279(39), pp. 40715–40722. Available at: https://doi.org/10.1074/jbc.M404648200.

Laskay, Ü.A. *et al.* (2013) 'Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments', *Journal of Proteome Research*, 12(12), pp. 5558–5569. Available at: https://doi.org/10.1021/pr400522h.

Lees, J.G. *et al.* (2006) 'A reference database for circular dichroism spectroscopy covering fold and secondary structure space', *Bioinformatics*, 22(16), pp. 1955–1962. Available at: https://doi.org/10.1093/bioinformatics/btl327.

Leonenko, Z. *et al.* (2007) 'An elevated level of cholesterol impairs self-assembly of pulmonary surfactant into a functional film', *Biophysical Journal*, 93(2), pp. 674–683. Available at: https://doi.org/10.1529/biophysj.107.106310.

Lindahl *et al.* (2021) 'GROMACS 2021.1 Source code'. Available at: https://doi.org/10.5281/ZENODO.4561626.

Lindahl, E., Hess, B. and van der Spoel, D. (2001) 'GROMACS 3.0: A package for molecular simulation and trajectory analysis', *Journal of Molecular Modeling*. Springer, pp. 306–317. Available at: https://doi.org/10.1007/S008940100045.

Manzer, R. *et al.* (2012) 'Ozone Exposure of Macrophages Induces an Alveolar Epithelial Chemokine Response through IL-1α', *https://doi.org/10.1165/rcmb.2007-0250OC*, 38(3), pp. 318–323. Available at: https://doi.org/10.1165/RCMB.2007-0250OC.

Markart, P. *et al.* (2007) 'Patients with ARDS show improvement but not normalisation of alveolar surface activity with surfactant treatment: Putative role of neutral lipids', *Thorax*, 62(7), pp. 588–594. Available at: https://doi.org/10.1136/thx.2006.062398.

Martínez-Calle, M. *et al.* (2019) 'Native supramolecular protein complexes in pulmonary surfactant: Evidences for SP-A/SP-B interactions', *Journal of Proteomics*, 207, p. 103466. Available at: https://doi.org/10.1016/j.jprot.2019.103466.

Martinez-Calle, M. *et al.* (2020) 'Pulmonary surfactant protein SP-B nanorings induce the multilamellar organization of surfactant complexes', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1862(6), p. 183216. Available at: https://doi.org/10.1016/J.BBAMEM.2020.183216.

McCormack, F.X. (1998) 'Structure, processing and properties of surfactant protein A', *Biochimica et Biophysica Acta - Molecular Basis of Disease*. Elsevier, pp. 109–131. Available at: https://doi.org/10.1016/S0925-4439(98)00062-3.

McPHERSON, A. (1990) 'Current approaches to macromolecular crystallization', *European journal of biochemistry*, 189(1), pp. 1–23. Available at: https://doi.org/10.1111/J.1432-1033.1990.TB15454.X.

Miles, A.J., Ramalli, S.G. and Wallace, B.A. (2022) 'DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data', *Protein Science : A Publication of the Protein Society*, 31(1), p. 37. Available at: https://doi.org/10.1002/PRO.4153.

Miles, A.J. and Wallace, B.A. (2016) 'Circular dichroism spectroscopy of membrane proteins', *Chemical Society Reviews*, 45(18), pp. 4859–4872. Available at: https://doi.org/10.1039/C5CS00084J.

Min, L., Choe, L.H. and Lee, K.H. (2015) 'Improved protease digestion conditions for membrane protein detection', *Electrophoresis*, 36(15), pp. 1690–1698. Available at: https://doi.org/10.1002/elps.201400579.

Molina de la Peña, I., Calvo, M.L. and F. Álvarez-Estrada, R. (2019) 'Focalizing slow neutron beams at and below micron scales and discussion on BNCT (II)', *Phosphorus, Sulfur, and Silicon and the Related Elements*, 194(10), pp. 956–966. Available at: https://doi.org/10.1080/10426507.2019.1631309.

Nagle, J.F. *et al.* (2019) 'Revisiting Volumes of Lipid Components in Bilayers', *Journal of Physical Chemistry B*, 123(12), pp. 2697–2709. Available at: https://doi.org/10.1021/ACS.JPCB.8B12010/SUPPL_FILE/JP8B12010_SI_001.PDF.

National Center for Biotechnology Information (2024) *1-Palmitoyl-2-(9-oxo-nonanoyl)-sn-glycero-3-phosphocholine*, *https://pubchem.ncbi.nlm.nih.gov/compound/46907874*.

Nelson, A. (2006) 'Co-refinement of multiple-contrast neutron/X-ray reflectivity data using MOTOFIT', *Journal of Applied Crystallography*, 39(2), pp. 273–276. Available at: https://doi.org/10.1107/S0021889806005073.

Nelson, A.R.J. and Prescott, S.W. (2019) 'Refnx: Neutron and X-ray reflectometry analysis in python', *Journal of Applied Crystallography*, 52(1), pp. 193–200. Available at: https://doi.org/10.1107/S1600576718017296/RG5158SUP5.ZIP.

Niga, P. *et al.* (2018) 'Interactions between model cell membranes and the neuroactive drug propofol', *Journal of Colloid and Interface Science*, 526, pp. 230–243. Available at: https://doi.org/10.1016/J.JCIS.2018.03.052.

Nogee, L.M. *et al.* (1994) 'A mutation in the surfactant protein B gene responsible for fatal neonatal respiratory disease in multiple kindreds', *Journal of Clinical Investigation*, 93(4), pp. 1860–1863. Available at: https://doi.org/10.1172/JCI117173.

Nosé, S. and Nose, S. (1984) 'A Unified Variational Formulation of Classical and Quantum Dynamics. I', *Citation: The Journal of Chemical Physics*, 81, p. 42. Available at: https://doi.org/10.1063/1.447334.

Ollila, O.H.S. and Pabst, G. (2016) 'Atomistic resolution structure and dynamics of lipid bilayers in simulations and experiments', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(10), pp. 2512–2528. Available at: https://doi.org/10.1016/J.BBAMEM.2016.01.019.

Olmeda, B., García-Álvarez, B. and Pérez-Gil, J. (2013) 'Structure-function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins', in *European Biophysics Journal*. Springer, pp. 209–222. Available at: https://doi.org/10.1007/s00249-012-0858-9.

Olżyńska, A. *et al.* (2016) 'Mixed DPPC/POPC Monolayers: All-atom Molecular Dynamics Simulations and Langmuir Monolayer Experiments', *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858(12), pp. 3120–3130. Available at: https://doi.org/10.1016/J.BBAMEM.2016.09.015.

Orgeig, S. *et al.* (2010) 'Recent advances in alveolar biology: Evolution and function of alveolar proteins', *Respiratory Physiology and Neurobiology*. NIH Public Access, p. S43. Available at: https://doi.org/10.1016/j.resp.2010.04.023.

Pallav L. Shah, Stradling, J.R. and Craig, S.E. (2020) 'The upper respiratory tract', in *Oxford Textbook of Medicine*. 6th edn, p. C18.1.1-C18.1.1.P29.

Pan, J. *et al.* (2012) 'Molecular structures of fluid phase phosphatidylglycerol bilayers as determined by small angle neutron and X-ray scattering', *Biochimica et biophysica acta*, 1818(9), pp. 2135–2148. Available at: https://doi.org/10.1016/J.BBAMEM.2012.05.007.

Parra, E. *et al.* (2013) 'Hydrophobic pulmonary surfactant proteins SP-B and SP-C induce pore formation in planar lipid membranes: Evidence for proteolipid pores', *Biophysical Journal*, 104(1), pp. 146–155. Available at: https://doi.org/10.1016/j.bpj.2012.11.014.

Parra, E. and Pérez-Gil, J. (2015) 'Composition, structure and mechanical properties define performance of pulmonary surfactant membranes and films', *Chemistry and Physics of Lipids*. Elsevier Ireland Ltd, pp. 153–175. Available at: https://doi.org/10.1016/j.chemphyslip.2014.09.002.

Pasenkiewicz-Gierula, M. *et al.* (1997) 'Hydrogen bonding of water to phosphatidylcholine in the membrane as studied by a molecular dynamics simulation: Location, geometry, and lipid-lipid bridging via hydrogen-bonded water', *Journal of Physical Chemistry A*, 101(20), pp. 3677–3691. Available at: https://doi.org/10.1021/JP962099V/ASSET/IMAGES/LARGE/JP962099VF00015.JPEG.

Paulin, L.M. *et al.* (2020) 'Association of Long-term Ambient Ozone Exposure With Respiratory Morbidity in Smokers', *JAMA Internal Medicine*, 180(1), pp. 106–115. Available at: https://doi.org/10.1001/JAMAINTERNMED.2019.5498.

Pettersen, E.F. *et al.* (2004) 'UCSF Chimera--a visualization system for exploratory research and analysis', *Journal of computational chemistry*, 25(13), pp. 1605–1612. Available at: https://doi.org/10.1002/JCC.20084.

Pinti, M. *et al.* (2016) 'Emerging role of Lon protease as a master regulator of mitochondrial functions', *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1857(8), pp. 1300–1306. Available at: https://doi.org/10.1016/J.BBABIO.2016.03.025.

Polgár, L. (2005) 'The catalytic triad of serine peptidases', *Cellular and Molecular Life Sciences*. Cell Mol Life Sci, pp. 2161–2172. Available at: https://doi.org/10.1007/s00018-005-5160-x.

Popovic, K. *et al.* (2012) 'Structure of saposin A lipoprotein discs', *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), pp. 2908–2912. Available at: https://doi.org/10.1073/pnas.1115743109.

Robinson, S.W., Afzal, A.M. and Leader, D.P. (2014) 'Bioinformatics: Concepts, Methods, and Data', *Handbook of Pharmacogenomics and Stratified Medicine*, pp. 259–287. Available at: https://doi.org/10.1016/B978-0-12-386882-4.00013-X.

Rodger, A. (2013) 'Far UV Protein Circular Dichroism', *Encyclopedia of Biophysics*, pp. 726–730. Available at: https://doi.org/10.1007/978-3-642-16712-6_634.

Roldan, N. *et al.* (2016) 'Effect of Lung Surfactant Protein SP-C and SP-C-Promoted Membrane Fragmentation on Cholesterol Dynamics', *Biophysical Journal*, 111(8), pp. 1703–1713. Available at: https://doi.org/10.1016/j.bpj.2016.09.016.

Romo, T.D. and Grossfield, A. (2014) 'Unknown unknowns: The challenge of systematic and statistical error in molecular dynamics simulations', *Biophysical Journal*. Biophysical Society, pp. 1553–1554. Available at: https://doi.org/10.1016/j.bpj.2014.03.007.

Rossmann, M. *et al.* (2008) 'Crystal Structures of Human Saposins C and D: Implications for Lipid Recognition and Membrane Interactions', *Structure*, 16(5), pp. 809–817. Available at: https://doi.org/10.1016/j.str.2008.02.016.

Ryan, M.A. *et al.* (2006) 'Antimicrobial activity of native and synthetic surfactant protein B peptides', *Journal of immunology (Baltimore, Md. : 1950)*, 176(1), pp. 416–425. Available at: https://doi.org/10.4049/JIMMUNOL.176.1.416.

Sarker, M., Jackman, D. and Booth, V. (2011) 'Lung surfactant protein A (SP-A) interactions with model lung surfactant lipids and an SP-B fragment', *Biochemistry*, 50(22), pp. 4867–4876. Available at: https://doi.org/10.1021/bi200167d.

Schlenkrich, M. *et al.* (1996) 'An Empirical Potential Energy Function for Phospholipids: Criteria for Parameter Optimization and Applications', *Biological Membranes*, pp. 31–81. Available at: https://doi.org/10.1007/978-1-4684-8580-6_2.

Schurch, S., Goerke, J. and Clements, J.A. (1976) 'Direct determination of surface tension in the lung', *Proceedings of the National Academy of Sciences of the United States of America*, 73(12), pp. 4698–4702. Available at: https://doi.org/10.1073/pnas.73.12.4698.

Sears, V.F. (1992) 'Neutron scattering lengths and cross sections', *Neutron News*, 3(3), pp. 26–37. Available at: https://doi.org/10.1080/10448639208218770.

Serrano, A.G. and Pérez-Gil, J. (2006) 'Protein-lipid interactions and surface activity in the pulmonary surfactant system', *Chemistry and Physics of Lipids*. Elsevier, pp. 105–118. Available at: https://doi.org/10.1016/j.chemphyslip.2006.02.017.

Seymour, J.F. and Presneill, J.J. (2002) 'Pulmonary alveolar proteinosis: Progress in the first 44 years', *American Journal of Respiratory and Critical Care Medicine*, pp. 215–235. Available at: https://doi.org/10.1164/rccm.2109105.

Shiraki, K., Nishikawa, K. and Goto, Y. (1995) 'Trifluoroethanol-induced Stabilization of the α-Helical Structure of β-Lactoglobulin: Implication for Non-hierarchical Protein Folding', *Journal of Molecular Biology*, 245(2), pp. 180–194. Available at: https://doi.org/10.1006/JMBI.1994.0015.

Skeel, R.D. (1999) 'Integration Schemes for Molecular Dynamics and Related Applications', in. Springer, Berlin, Heidelberg, pp. 119–176. Available at: https://doi.org/10.1007/978-3-662-03972-4_4.

Squires, G.L. (2012) *Introduction to the Theory of Thermal Neutron Scattering*. Cambridge University Press. Available at: https://doi.org/10.1017/CBO9781139107808.

Sreerama, N. and Woody, R.W. (2000) 'Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set', *Analytical Biochemistry*, 287(2), pp. 252–260. Available at: https://doi.org/10.1006/abio.2000.4880.

Stachowicz-Kuśnierz, A. *et al.* (2022) 'The lung surfactant activity probed with molecular dynamics simulations', *Advances in Colloid and Interface Science*, 304, p. 102659. Available at: https://doi.org/10.1016/J.CIS.2022.102659.

van Stokkum, I.H.M. *et al.* (1990) 'Estimation of protein secondary structure and error analysis from circular dichroism spectra', *Analytical Biochemistry*, 191(1), pp. 110–118. Available at: https://doi.org/10.1016/0003-2697(90)90396-Q.

Sych, T., Mély, Y. and Römer, W. (2018) 'Lipid self-assembly and lectin-induced reorganization of the plasma membrane', *Philosophical Transactions of the Royal Society B: Biological Sciences*. Royal Society Publishing. Available at: https://doi.org/10.1098/rstb.2017.0117.

Takada, S. (2012) 'Coarse-grained molecular simulations of large biomolecules', *Current Opinion in Structural Biology*, 22(2), pp. 130–137. Available at: https://doi.org/10.1016/J.SBI.2012.01.010.

Taneva, S.G. and Keough, K.M.W. (2000) 'Adsorption of pulmonary surfactant protein SP-A to monolayers of phospholipids containing hydrophobic surfactant protein SP-B or SP-C: Potential differential role for tertiary interaction of lipids, hydrophobic proteins, and SP-A', *Biochemistry*, 39(20), pp. 6083–6093. Available at: https://doi.org/10.1021/bi992074x.

Terpe, K. (2006) 'Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems', *Applied Microbiology and Biotechnology*, pp. 211–222. Available at: https://doi.org/10.1007/s00253-006-0465-8.

Thompson, K.C. *et al.* (2013) 'Degradation and rearrangement of a lung surfactant lipid at the air-water interface during exposure to the pollutant gas ozone', *Langmuir*, 29(14), pp. 4594–4602. Available at: https://doi.org/10.1021/LA304312Y/SUPPL_FILE/LA304312Y_SI_001.PDF.

Tropea, J.E., Cherry, S. and Waugh, D.S. (2009) 'Expression and purification of soluble His6-tagged TEV protease', *Methods in Molecular Biology*, 498, pp. 297–307. Available at: https://doi.org/10.1007/978-1-59745-196-3_19/COVER.

Tuckerman, M.E. and Martyna, G.J. (2000) 'Understanding Modern Molecular Dynamics: Techniques and Applications', *Journal of Physical Chemistry B*, 104(2), pp. 159–178. Available at: https://doi.org/10.1021/jp992433y.

Tuffery, P. and Derreumaux, P. (2012) 'Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: Limitations of current computational approaches', *Journal of the Royal Society Interface*. Royal Society, pp. 20–33. Available at: https://doi.org/10.1098/rsif.2011.0584.

University of Oxford (2019) *SEP-PAK C18 PURIFICATION*.

Vandenbussche, G. *et al.* (1992) 'Secondary Structure and Orientation of the Surfactant Protein SP-B in a Lipid Environment. A Fourier Transform Infrared Spectroscopy Study', *Biochemistry*, 31(38), pp. 9169–9176. Available at: https://doi.org/10.1021/bi00153a008.

Vargaftik, N.B., Volkov, B.N. and Voljak, L.D. (1983) 'International Tables of the Surface Tension of Water', *Journal of Physical and Chemical Reference Data*, 12(3), pp. 817–820. Available at: https://doi.org/10.1063/1.555688.

Venyaminov, S.Y. *et al.* (1993) 'Circular dichroic analysis of denatured proteins: Inclusion of denatured proteins in the reference set', *Analytical Biochemistry*, 214(1), pp. 17–24. Available at: https://doi.org/10.1006/abio.1993.1450.

Vieira, F., Kung, J.W. and Bhatti, F. (2017) 'Structure, genetics and function of the pulmonary associated surfactant proteins A and D: The extra-pulmonary role of these C type lectins', *Annals of Anatomy*. Elsevier GmbH, pp. 184–201. Available at: https://doi.org/10.1016/j.aanat.2017.03.002.

Vieira-Pires, R.S. and Morais-Cabral, J.H. (2010) '310 helices in channels and other membrane proteins', *The Journal of General Physiology*, 136(6), p. 585. Available at: https://doi.org/10.1085/JGP.201010508.

Wallace, B.A. (2003) 'Analyses of circular dichroism spectra of membrane proteins', *Protein Science*, 12(4), pp. 875–884. Available at: https://doi.org/10.1110/ps.0229603.

Wallace, B.A. *et al.* (2004) 'Biomedical applications of synchrotron radiation circular dichroism spectroscopy: Identification of mutant proteins associated with disease and development of a reference database for fold motifs', *Faraday Discussions*, 126(1), pp. 237–243. Available at: https://doi.org/10.1039/b306055c.

Walther, F.J. *et al.* (2016) 'Synthetic lung surfactants containing SP-B and SP-C peptides plus novel phospholipase-resistant lipids or glycerophospholipids'. Available at: https://doi.org/10.7717/peerj.2635.

Walther, F.J., Gordon, L.M. and Waring, A.J. (2016) 'Design of Surfactant Protein B Peptide Mimics Based on the Saposin Fold for Synthetic Lung Surfactants', *Biomedicine Hub*, 1(3), p. 1. Available at: https://doi.org/10.1159/000451076.

Walther, F.J., Gordon, L.M. and Waring, A.J. (2019) 'Advances in synthetic lung surfactant protein technology', *Expert Review of Respiratory Medicine*, 13(6), pp. 499–501. Available at: https://doi.org/10.1080/17476348.2019.1589372.

Wang, G. *et al.* (2019) 'Differential effects of human SP-A1 and SP-A2 on the BAL proteome and signaling pathways in response to Klebsiella pneumoniae and ozone exposure', *Frontiers in Immunology*, 10(MAR), p. 561. Available at: https://doi.org/10.3389/fimmu.2019.00561.

Wang, T. *et al.* (2022) 'Ground-level ozone pollution in China: A synthesis of recent findings on influencing factors and impacts', *Environmental Research Letters*, 17(6). Available at: https://doi.org/10.1088/1748-9326/AC69FE.

Wang, V. (2019) 'The Expression and Characterisation of Human Lung Surfactant Protein SP-B', MSc thesis, Birkbeck University of London.

Waring, A.J. *et al.* (2005) 'The role of charged amphipathic helices in the structure and function of surfactant protein B', *Journal of Peptide Research*, 66(6), pp. 364–374. Available at: https://doi.org/10.1111/j.1399-3011.2005.00300.x.

Waring, A.J. *et al.* (2023) 'Lung Surfactant Protein B Peptide Mimics Interact with the Human ACE2 Receptor', *International Journal of Molecular Sciences*, 24(13), p. 10837. Available at: https://doi.org/10.3390/IJMS241310837/S1.

Warr, R.G. *et al.* (1987) 'Low molecular weight human pulmonary surfactant protein (SP5): Isolation, characterization, and cDNA and amino acid sequences', *Proceedings of the National Academy of Sciences of the United States of America*, 84(22), pp. 7915–7919. Available at: https://doi.org/10.1073/pnas.84.22.7915.

Watson, A. *et al.* (2018) 'Surfactant proteins A and D: Trimerized innate immunity proteins with an affinity for viral fusion proteins', *Journal of Innate Immunity*. S. Karger AG, pp. 13–28. Available at: https://doi.org/10.1159/000492974.

Weaver, T.E. and Whitsett, J.A. (1991) 'Function and regulation of expression of pulmonary surfactant-associated proteins', *Biochemical Journal*. Portland Press Ltd, pp. 249–264. Available at: https://doi.org/10.1042/bj2730249.

Whitmore, L. and Wallace, B.A. (2004) 'DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data', *Nucleic Acids Research*, 32(WEB SERVER ISS.). Available at: https://doi.org/10.1093/nar/gkh371.

Whitmore, L. and Wallace, B.A. (2007) 'Protein Secondary Structure Analyses from Circular Dichroism Spectroscopy: Methods and Reference Databases'. Available at: https://doi.org/10.1002/bip.20853.

Wiegman, C.H. *et al.* (2020) 'Oxidative Stress in Ozone-Induced Chronic Lung Inflammation and Emphysema: A Facet of Chronic Obstructive Pulmonary Disease', *Frontiers in Immunology*, 11, p. 547173. Available at: https://doi.org/10.3389/FIMMU.2020.01957/BIBTEX.

Wüstneck, N. *et al.* (2003) 'Effects of oligomerization and secondary structure on the surface behavior of pulmonary surfactant proteins SP-B and SP-C', *Biophysical Journal*, 84(3), pp. 1940–1949. Available at: https://doi.org/10.1016/S0006-3495(03)75002-4.

Xiang, Z. (2006) 'Advances in Homology Protein Structure Modeling', *Current Protein & Peptide Science*, 7(3), pp. 217–227. Available at: https://doi.org/10.2174/138920306777452312.

Yang, L. *et al.* (2010) 'Surfactant Protein B Propeptide Contains a Saposin-Like Protein Domain with Antimicrobial Activity at Low pH', *The Journal of Immunology*, 184(2), pp. 975–983. Available at: https://doi.org/10.4049/JIMMUNOL.0900650.

Yasin, B. *et al.* (1996) 'Protegrins: structural requirements for inactivating elementary bodies of Chlamydia trachomatis', *Infection and immunity*, 64(11), pp. 4863–4866. Available at: https://doi.org/10.1128/IAI.64.11.4863-4866.1996.

Young, J.A. and Koppel, J.U. (1964) 'Slow Neutron Scattering by Molecular Hydrogen and Deuterium', *Physical Review*, 135(3A), p. A603. Available at: https://doi.org/10.1103/PhysRev.135.A603.

Young, S.L., Fram, E.K. and Larson, E.W. (1992) 'Three-dimensional reconstruction of tubular myelin', *Experimental Lung Research*, 18(4), pp. 497–504. Available at: https://doi.org/10.3109/01902149209064342.

Zaltash, S. *et al.* (2000) 'Pulmonary surfactant protein B: a structural model and a functional analogue.', *Biochimica et biophysica acta*, 1466(1–2), pp. 179–86. Available at: https://doi.org/10.1016/s0005-2736(00)00199-1.

Zhang, Z. *et al.* (2015) 'High-level production of membrane proteins in E. coli BL21(DE3) by omitting the inducer IPTG', *Microbial Cell Factories*, 14(1), p. 142. Available at: https://doi.org/10.1186/s12934-015-0328-z.

Zimmermann, L.J.I. *et al.* (2005) 'Surfactant metabolism in the neonate', in *Biology of the Neonate*, pp. 296–307. Available at: https://doi.org/10.1159/000084877.

Zoete, V. *et al.* (2011) 'SwissParam: A fast force field generation tool for small organic molecules', *Journal of Computational Chemistry*, 32(11), pp. 2359–2368. Available at: https://doi.org/10.1002/JCC.21816.

Zuo, Y.Y. *et al.* (2008) 'Current perspectives in pulmonary surfactant - Inhibition, enhancement and evaluation', *Biochimica et Biophysica Acta - Biomembranes*. Elsevier, pp. 1947–1977. Available at: https://doi.org/10.1016/j.bbamem.2008.03.021.