

**Some pages of this thesis may have been removed for copyright restrictions.**

If you have discovered material in Aston Research Explorer which is unlawful e.g. breaches copyright, (either yours or that of a third party) or any other law, including but not limited to those relating to patent, trademark, confidentiality, data protection, obscenity, defamation, libel, then please read our [Takedown policy](#) and contact the service immediately (openaccess@aston.ac.uk)

# RECOMBINANT SYNTHESIS OF PULMONARY SURFACTANT PROTEINS SP-B AND SP-C

By Anjana Patel

Doctor of Philosophy

Aston University

September 2018



©Anjana Patel, 2018

Anjana Patel asserts her moral right to be identified as the author of this thesis

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledgement.

**ASTON UNIVERSITY**

# RECOMBINANT SYNTHESIS OF SP-B AND SP-C

Anjana Patel

Doctor of Philosophy

2017

## Thesis Summary

SP-B and SP-C are two pulmonary surfactant proteins. They were isolated in the tear film in 2006 (Lukovic *et al.* 2006).

The molecular mechanism governing lipid spreading in the tear film is not well understood and, along with their production in yeast (as an alternative to animal derived material) provide the motives for this research.

To investigate the involvement of SP-B and SP-C in lipid spreading, SP-B and SP-C have been produced as recombinant proteins in *Pichia pastoris* yeast.

SP-B and SP-C are cleaved from their proproteins to produce mature active proteins and both forms were produced recombinantly.

The proteins were extracted with detergents, polymers and organic solvents to determine the best method of protein isolation and purification.

Purified proteins were subsequently tested for surface activity using a Langmuir trough; proSP-C10 and SP-B demonstrated surface tension activity.

The mechanism of SP-B and SP-C was examined through generation of *ab initio* structural models.

The behaviour of SP-B and SP-C was observed in membranes based on phospholipid compositions surrounding the SP-B and SP-C in the yeast plasma membrane using mass spectrometry analysis.

Computational modelling then demonstrated that SP-B induces lipid curvature, whilst SP-C may help to stabilise lipid vesicles by forming a molecular bridge between the membrane and vesicle.

SP-B is thought to help remodel the lung lipid layer on compression through the formation of vesicles that are connected to the membrane by SP-C.

## **Acknowledgements**

I'd like to give special thanks to a true unsung hero Professor Gerald Noone (OBE, MBE), for the continuous encouragement, the endless hours spent helping me improve my thesis and being a truly inspiring role model.

I'd like to thank my supervisors Professors Roslyn Bill, Brian Tighe (OBE) and Dr Alice Rothnie for giving me the opportunity to embark on a PhD and their guidance throughout. I'd also like to thank Professors David Poyner and Tim Dafforn for taking the time to examine this thesis.

My thanks go to Coopervision, in particular Andrew Broad and his team in Southampton for sponsoring the PhD project.

I'd like to also give special thanks to Dr John Simms for extensive help with the computational modelling and probably the most insightful scientific discussions I feel I'll ever have.

Other important acknowledgments are:

Val Franklin, for help with Langmuir trough experiments (Aston University)

Charles Moore-Kelly, for his help with Circular Dichroism (Birmingham University)

Ivana Milic, for her help with lipid mass spectrometry (Aston University)

Group members in particular Lina, Rav, Michelle, Pinar, Hoor, Romez, Aimen, David, Floren and Jaimin. My thanks also go to members of the BRU group-Aisling, Fiona and Simran.

I'd like to give an enormous thank you to Dr Sarah Routledge for her technical advice, unwavering support, patience and making working at Aston a real joy for me.

Finally I'd like to thank my family, especially my sister for being my cheerleader throughout this process and for being the best sister EVER!

## Contents

	Page
<b>Abbreviations</b>	8
<b>List of Tables</b>	10
<b>List of Figures</b>	11
<b>Chapter 1: Introduction</b>	
1.1.1 The Tear Film	16
1.1.2 The Mucoaqueous Layer	19
1.1.3 The Lipid Layer	21
1.1.4 Tear Film Spreading	27
1.1.5 Dry Eye	29
1.1.6 Surface Tension	30
1.1.7 Langmuir Trough and Surface Activity	31
1.2 Pulmonary Surfactant Proteins	
1.2.1 Discovery of SP-B and SP-C in the Tear Film	33
1.2.2 Pulmonary Surfactant	35
1.2.3 Pulmonary Surfactant Composition	36
1.2.4 Synthesis of SP-B and SP-C	38
1.2.5 Structure of SP-B	39
1.2.6 Structure of SP-C	43
1.2.7 The Role of SP-B and SP-C in the Lipid Layer of the Lung	45
1.2.8 DISCUSSION - Recombinant Synthesis of SP-B Within Previous Work	47
1.2.9 DISCUSSION - Recombinant Synthesis of SP-C	49
1.3 <i>Pichia Pastoris</i>	
1.3.1 DISCUSSION - <i>Pichia Pastoris</i> as a Recombinant Host Cell	52
1.3.2 Phospholipid Composition of <i>Pichia Pastoris</i>	54
1.4 Membrane Protein Extraction	
1.4.1 Solubilisation of Proteins from Membranes	56
1.4.2 SMA 2000P Solubilisation	56
1.4.3 DISCUSSION -Detergent Solubilisation	58
1.5 Aims and Objectives of the Work of this Thesis	62
<b>Chapter 2: Materials and Methods</b>	
2.1 Materials	
2.1.1 Buffers	
2.1.1.1 50 X TAE buffer	63
2.1.1.2 1% Agarose	63
2.1.1.3 Yeast Extract Peptone Dextrose Sorbitol Plates (YPDS)	63
2.1.1.4 Yeast Extract Peptone Dextrose Medium (YPD)	63
2.1.1.5 Buffered Glycerol-Complex Medium (BMGY) and Buffered Methanol-Complex Medium (BMMY)	64
2.1.1.6 Minimal Media	64
2.1.1.7 Breaking Buffer	64
2.1.1.8 Buffer A	64
2.1.1.9 Protein Buffer	65
2.1.1.10 Laemmli Buffer	65
2.1.1.11 Binding/Wash Buffer pH 8.0	65
2.1.1.12 Elution Buffer pH 8.0	65
2.1.1.13 Phosphate Buffered Saline (PBS)	65
2.1.1.14 Running Buffer	66
2.1.1.15 1 x TE Buffer (pH 8.0)	66
2.1.1.16 Transfer Buffer	66

2.1.1.17 Tris HCl (pH 8.0)	66
2.1.2 Reagents	
2.1.2.1 1 Kb O'GeneRuler DNA Ladder	66
2.1.2.2 Loading Dye	66
2.1.3 Restriction Enzymes	67
2.1.4 Competent Cells	
2.1.4.1 XL10 Gold	67
2.1.5 Kits and Equipment	
2.1.5.1 PCR Kit	67
2.1.5.2 Ligation Kit	67
2.1.5.3 Mini Prep Kit	67
2.1.5.4 Maxi Prep Kit	67
2.1.6 Strains	
2.1.6.1 <i>E.coli</i> XL10 Gold Competent Cells	67
2.1.6.2 <i>Pichia Pastoris</i> Strain X-33	67
2.1.7 Primers and Vectors	
2.1.7.1 Vectors	68
2.1.7.2 pPICZαA Vector Map	69
2.1.7.3 Primer Sequences	71
2.2 Methods	
2.2.1 Molecular Biology	
2.2.1.1 Vector Design	72
2.2.1.2 Primer Design	72
2.2.1.3 Transformation of Surfactant Genes in PUC19 Vector	72
2.2.1.4 MiniPrep	73
2.2.1.5 Nanodrop	73
2.2.1.6 Polymerase Chain Reaction (PCR)	74
2.2.1.7 Restriction Digest	75
2.2.1.8 Ligation	76
2.2.1.9 Transformation into XL10 Gold	76
2.2.1.10 Preparation of Competent X-33 Cells	77
2.2.2 Protein Expression	
2.2.2.1 Transformation of Pulmonary Surfactant Genes into <i>Pichia Pastoris</i> Strain X-33	78
2.2.2.2 Transformation Screen	79
2.2.2.3 Bioreactors	80
2.2.2.4 Membrane Preparation (large scale)	81
2.2.2.5 BCA Assay	81
2.2.2.6 Polymer Preparation	82
2.2.2.7 Polymer Solubilisation	83
2.2.2.8 Detergent Solubilisation	83
2.2.2.9 Protein Purification	84
2.2.2.10 Western Blot	85
2.2.2.11 Tris-Tricine Gel	87
2.2.2.12 Reconstitution of SP-B and SP-C in Synthetic Lipids	88
2.2.2.13 Sucrose Gradients	89
2.2.2.14 Determination of Protein Concentration	89
2.2.2.15 Densitometry	90
2.2.2.16 Langmuir Trough	91
2.2.2.17 Circular Dichroism	92
2.2.2.18 Statistical Analysis	92
2.3 Mass spectrometry	
2.3.1.1 Lipid Extraction	92
2.3.1.2 Mass Spectrometry	93
2.3.1.3 Data Analysis	93

2.3.1.4 Mass Spectrometry Confirming Protein Identification	94
2.4 Computational Modelling	95
<b>Chapter 3: Production and Purification of Recombinant ProSP-B and SP-B</b>	
3.1 Design and Construction of ProSP-B and SP-B Expression Constructs	97
3.2 <i>ProSFTPB</i> Synthesis	99
3.3 Isolation of High Yielding ProSP-B Producing Colonies	101
3.4 Identification of a High-Yielding Colony for Further Scale-Up	103
3.5 Solubilisation of SP-B with Organic Solvents	105
3.6 Polymer Solubilisation of ProSP-B	107
3.7 Detergent Solubilisation of ProSP-B	110
3.8 Solubilisation of ProSP-B with Guanidine Hydrochloride	112
3.9 Purification of ProSP-B using Guanidine Hydrochloride	113
3.10 The Recombinant Production of SP-B: <i>SFTPB</i> Vector Construction	118
3.11 Isolation of a High-Yielding SP-B Colony	119
3.12 Scale-Up of Highly-Expressing Colony 5	120
3.13 Solubilisation and Purification of SP-B with SMA2000P	121
3.14 SP-B-SMALP Secondary Structure Analysis	124
3.15 SP-B Detergent Solubilisation Screen	126
3.16 Purification using Anzergent	129
3.17 Effect of Recombinant SP-B on Surface Tension	131
3.18 Conclusions - Production and Purification of Recombinant ProSP-B and SP-B	141
<b>Chapter 4: Production and Purification of Recombinant ProSP-C and SP-C</b>	
4.1 Design and Construction of ProSP-C and SP-C Expression Constructs	143
4.2 Isolation of a High-Yielding ProSP-C Expressing Colony	145
4.3 Context of Molecular Weight Differences and Outcomes	149
4.4 Scale-Up of ProSP-C Expressing Cultures	150
4.5 Polymer Solubilisation of ProSP-C	152
4.6 Purification of ProSP-C-SMALP	154
4.7 ProSP-C Detergent Solubilisation	157
4.8 Purification of Detergent Solubilised ProSP-C	159
4.9 <i>ProSFTPC10</i> Synthesis	165
4.10 Solubilisation and Purification of ProSP-C10 using SMA2000P	169
4.11 Functional Characterisation of ProSPC-10 in SMALP	173
4.12 Solubilisation and Purification of ProSP-C10 using Conventional Detergents	176
4.13 Attempts to Reconstitute ProSP-C10 into Proteoliposomes	183
4.14 <i>SFTPC</i> Synthesis	184
4.15 Isolation of a High-Yielding SP-C Colony	186
4.16 Scale Up of a High Expressing Colony	187
4.17 SP-C Detergent Solubilisation Screen	188
4.18 Purification of Detergent Solubilised SP-C	190
4.19 Effect of SP-C Associated with POPC lipids on Surface Tension	195
4.20 Conclusions - Production and Purification of Recombinant ProSP-C and SP-C	198

## **Chapter 5: Analysis of The Lipid Context of Recombinant Surfactant Proteins: Using Mass Spectrometry to Inform Molecular Dynamics Simulations**

5.1 Lipid Extraction of ProSP-C, ProSP-C10, SP-C, ProSP-B and SP-B	200
5.2 Mass Spectrometric Analysis of the Lipid Compositions of SMALPs Containing ProSP-C, ProSP-C10, SP-C, ProSP-B, and SP-B	201
5.3 Lipids Associated with Recombinant, SMA-Extracted, ProSP-C: The Control Reference Condition	202
5.4 Lipids Associated with Recombinant, SMA-Extracted, Purified ProSP-C10	203
5.5 Some Implications of Lipids Associated with Recombinant, SMA-Extracted, Purified SP-C	203
5.6 Lipids Associated with Recombinant, SMA-Extracted, Purified ProSP-B	204
5.7 Lipids Associated with Recombinant, SMA-Extracted, Purified SP-B	205
5.8 Further Discussions and Evaluations of Lipid Associations with Recombinant Surfactant Proteins	207
5.9 Conclusions - Analysis of The Lipid Context of Recombinant Surfactant Proteins: Using Mass Spectrometry to Inform Molecular Dynamics Simulations	211
<b>Chapter 6: Molecular Dynamics Simulations of SP-B and SP-C</b>	
6.1 Modelling the Structural Dynamics of SP-B in an Implicit Membrane	214
6.2 Modelling the Structural Dynamics of SP-C in an Implicit Membrane	222
6.3 Modelling the Structural Dynamics of SP-B and SP-C in Explicit Membranes	228
6.4 The Function of SP-B and SP-C in Tear Film Lipid Spreading	236
6.5 Context of SP-B and SP-C in Tear Film Spreading	236
6.6 Conclusions - Molecular Dynamics Simulations of SP-B and SP-C	239
<b>Chapter 7: Discussion</b>	
7.1 Production of Recombinant ProSP-B and SP-B: ProSP-B Expression	240
7.2 SP-B Expression and Extraction	242
7.3 Mass Spectrometry Analysis and Modelling of Lipids Around SP-B	243
7.4 Computational Modelling: SP-B in an Implicit Membrane	244
7.5 SP-B in an Explicit Membrane	245
7.6 Production of Recombinant ProSP-C and SP-C: ProSP-C Expression, Extraction, Purification and Functional Analysis	246
7.7 SP-C Expression and Extraction	247
7.8 SP-C Mass Spectrometry	247
7.9 SP-C Computational Modelling	248
7.10 SP-C Implicit Model	248
7.11 SP-C Explicit Model	248
7.12 Summary and Key Conclusions	249
7.13 Future work	250
<b>References</b>	252

## Abbreviations

ABCA3	ATP-Binding Cassette Sub-Family A Member 3
Anz	Anzergent
AOX1	Alcohol Oxidase 1
APS	Ammonium Persulphate
ATII	Alveolar Type II Cells
BCA	Bicinchoninic Acid
BMGY	Buffered Glycerol Complex Medium
BMMY	Buffered Methanol Complex Medium
$\beta$ -OG	Octyl-Beta-Glucoside
CD	Circular Dichroism
CHAPS	3-((3-cholamidopropyl) Dimethylammonio)-1-Propanesulfonate
CHO	Chinese Hamster Ovary Cells
CID	Collision-Induced Dissociation
CMC	Critical Micelle Concentration
DAG	Diacyl-Glycerols
DM	Decyl Maltoside
DO	Dissolved Oxygen
DOC	Deoxycholate
DPPC	Dipalmitoylphosphatidylcholine
DTT	Dithiothreitol
<i>E. coli</i>	<i>Escherichia Coli</i>
EDTA	Ethylenediaminetetraacetic Acid
ER	Endoplasmic Reticulum
Fos-12	Fos-Choline 12
G-protein	Guanine Nucleotide-Binding Protein
HEKs	Human Embryonic Kidney Cells
h	Hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High-Performance Liquid Chromatography
IDA	Information Dependent Mode
LB	Lamellar Bodies
L	Litre
LB	Lysogeny Broth
LS	Lysophosphatidylcholine
mg	Milligram
MNG	Maltose Neopentyl Glycol-3
mL	Millilitre
MVB	Multivesicular Bodies
Ni-NTA	Nickel Nitrilotriacetic Acid
NLSS	N-lauroylsarcosine Sodium Salt
NMR	Nuclear Magnetic Resonance
MS	Mass Spectrometry
OAHFA	( <i>O</i> -acyl)- $\omega$ -hydroxy Fatty Acids
PA	Phosphatidic Acid
PBS	Phosphate Buffered Saline
PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction
PE	Phosphatidylethanolamine
PI	Phosphatidylinositol

PG	Phosphatidylglycerol
PKC	Protein Kinase C
POPC	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphoethanolamine
POPI	1-palmitoyl-2-oleoyl- <i>sn</i> -glycero-3-phosphoinositol
POPS	1-palmitoyl-2-oleoylglycero-3-phosphoserine
<i>P. pastoris</i>	<i>Pichia Pastoris</i>
ProSP-B	Precursor Pulmonary Surfactant Protein B
ProSP-C	Precursor Pulmonary Surfactant Protein C
PS	Phosphatidylserine
RDS	Respiratory Distress Syndrome
rpm	Revolutions Per Minute
RT-PCR	Reverse Transcription Polymerase Chain Reaction
REMC	Replica Exchange Monte Carlo
SALIP	Saposin-like Protein
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SMA	Styrene-Maleic Acid
SMALP	Styrene-alt-Maleic Acid Lipid Particles
SN	<i>Staphylococcus Nuclease</i>
SPB	Pulmonary Surfactant Protein B
SP-C	Pulmonary Surfactant Protein C
ProSP-C10	Precursor Pulmonary Surfactant Protein C with a His <sub>10</sub> tag
SM	Sphingomyelin
TAE	Tris-Acetate-EDTA
TAG	Triacyl-Glycerol
TEMED	N,N,N',N'-tetramethyl- ethane-1,2-diamine
TG	Triglyceride
TCA	Trichloroacetic Acid
YNB	Yeast Nitrogen Base
YPDS	Yeast Extract Peptone Dextrose Sorbitol

## List of Tables

	Page
Table 1.1: Structures of the Lipids Making up the Polar Layer and Non-Polar Layer of the Tear Film	24
Table 1.2: Composition of Lipids Making up the Polar Layer and Non-Polar Layer of the Tear Film	26
Table 1.3: Similarities and Differences Between Members of the SALIP Family	42
Table 1.4: Mutations in SP-C	45
Table 1.5: Comparison of the Methods Used to Synthesise and Extract Recombinant ProSP-B and SP-B	48
Table 1.6: Comparison of the Methods Used to Synthesis and Extract Recombinant Mature SP-C	51
Table 1.7: Phospholipid Abundances of <i>Pichia Pastoris</i> Identified Using Mass Spectrometry	55
Table 1.8: Examples of Detergents Used in Purification	61
Table 2.1: Restriction Enzymes Used for Cloning	67
Table 2.2: Commercial Vectors Used in Cloning	68
Table 2.3: Recombinant Protein Vectors	68
Table 2.4: Important Features Within the pPICZαA, B and C Vector and their Role	70
Table 2.5: Volumes and Final Concentrations of Reagents added to PCR Reactions	74
Table 2.6: PCR Cycling Times	74
Table 2.7: Restriction Enzyme Reaction Composition	75
Table 2.8: Enzymes and Reagents added to Ligation Reactions	76
Table 2.9: Reagents Required for 12% Separating Gel and 4% Stacking Gel in SDS-PAGE	85
Table 2.10: Reagents Required to make Tris-Tricine Separating and Stacking Gels	87
Table 3.1: Positive Identification of ProSP-B	116
Table 3.2: Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-B (70 kDa)	117
Table 3.3: Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-B (35 kDa)	117
Table 3.4: Identified SP-B Fragments from Mass Spectrometry Analysis	123
Table 3.5: Identified SP-B Fragments from Mass Spectrometry Analysis (40 kDa)	123
Table 3.6: Identified Peptide Fragments from Mass Spectrometry Analysis of SP-B (25 kDa)	124
Table 3.7: Statistical Significance of Different Detergents vs DM Using the Dunnet's Test	128
Table 4.1: Identified ProSP-C Peptide Fragments from Mass Spectrometry Analysis	156
Table 4.2: Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-C (35 kDa)	156
Table 4.3: Identified ProSP-C10 Fragments and their Location in the ProSP-C Sequence	182
Table 4.4: Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-C10 (40 kDa)	182
Table 5.1: Identified Phospholipids from Mass Spectrometry Analysis	209

## List of Figures

	Page
<b>Chapter 1</b>	
Figure 1.1: Schematic Representation of the Structure of the Eye and the Tear Film	16
Figure 1.2: Location of the Lacrimal Glands	17
Figure 1.3: Meibomian Gland Structure	18
Figure 1.4: Mucoaqueous Layer Structure	20
Figure 1.5: Tear Film Lipid Layer Structure	22
Figure 1.6: Molecular Models of Tear Film Lipid Spreading	28
Figure 1.7: Surface Tension	30
Figure 1.8: Schematic Illustration of the Langmuir Trough and Associated Components	33
Figure 1.9: Considered, Potential Model of the Interaction of SP-B and SP-C in the Tear Film	34
Figure 1.10: Pulmonary Surfactant Secretion	37
Figure 1.11: Primary Sequence of ProSP-B	39
Figure 1.12: Structure of SP-B	41
Figure 1.13: Primary Sequence of ProSP-C	43
Figure 1.14: The Structure of SP-C	44
Figure 1.15: Lipid Spreading in the Lung	46
Figure 1.16: SP-C Fusion Protein Construct used by Lukovic <i>et al.</i> 2006	49
Figure: 1.17: Structures of Styrene Maleic Anhydride and Styrene Maleic Acid	57
Figure 1.18: Protein Solubilisation using SMA2000P	58
Figure 1.19: Detergent Solubilisation	59
<b>Chapter 2</b>	
Figure 2.1: Invitrogen <i>Pichia Pastoris</i> pPICZ $\alpha$ A,B,C Expression Vector	69
<b>Chapter 3</b>	
Figure 3.1: <i>ProSFTP</i> B and <i>SFTP</i> B Vector Map Design	98
Figure 3.2: <i>ProSFTP</i> B Synthesis	99

Figure 3.3: Sequence Alignment of ProSP-B	100
Figure 3.4: ProSP-B Protein Expression Screen	102
Figure 3.5: Scale up of a High-Yielding ProSP-B Colony in a 2L Bioreactor	104
Figure 3.6: Extraction of ProSP-B using Chloroform and Methanol	106
Figure 3.7: Purification of Chloroform and Methanol Extracted ProSP-B	107
Figure 3.8: ProSP-B SMA Solubilisation Screen	109
Figure 3.9: ProSP-B Purification via Solubilisation with SMA2000P	109
Figure 3.10: Detergent Solubilisation of ProSP-B	111
Figure 3.11: Solubilisation of ProSP-B using GH	112
Figure 3.12: Purification of ProSP-B using GH	114
Figure 3.13: Diagnostic Digest of the Transformed <i>SFTP</i> B Vector	118
Figure 3.14: SP-B Protein Expression Screen	120
Figure 3.15: Large Scale Fermentation of a High SP-B Expressing Yeast Colony	121
Figure 3.16 Purification of SP-B after Solubilisation with SMA2000P	122
Figure 3.17: CD Trace of SP-B-SMALP and Accompanying HT	125
Figure 3.18: SP-B Detergent Solubilisation Screen	126
Figure 3.19: Statistical Analysis of SP-B Detergent Solubilisation	128
Figure 3.20: SP-B Purification using Anzergent	130
Figure 3.21: Surface Activity Measurement of SP-B using a Langmuir Trough	132
Figure 3.22: Maximum Surface Pressure of SP-B in Anzergent Micelles Plotted Against Volume	133
Figure 3.23: Behaviour of SP-B and Detergents in the Langmuir Trough	135
Figure 3.24: Maximum Surface Pressure of Precipitated SP-B	137
Figure 3.25: Maximum Pressure of SP-B Reconstituted in POPC Lipids	139
Figure 3.26: Examination of Surface Activity of SP-B in POPC Lipids	140
<b>Chapter 4</b>	
Figure 4.1: <i>ProSFTPC</i> and <i>SFTPC</i> Vector Map Design	144
Figure 4.2: <i>ProSFTPC</i> Synthesis	145
Figure 4.3: ProSP-C Sequence Alignment	146
Figure 4.4: ProSP-C Expression Screen	148

Figure 4.5: Scale-up of High-Yielding ProSP-C Colony in a 2 L Bioreactor	151
Figure 4.6: SMALP Extraction of ProSP-C	153
Figure 4.7: Purification of ProSP-C in SMALP	154
Figure 4.8: ProSP-C Detergent Screen	158
Figure 4.9: ProSP-C Purification Visualised on a Coomassie Stained Gel	160
Figure 4.10: Purification of ProSP-C using an AKTApurifier 10 System	162
Figure 4.11: ProSP-C Purification using Co <sup>2+</sup> -NTA	164
Figure 4.12: <i>ProSFTPC10</i> Vector Map Design	166
Figure 4.13: <i>ProSFTPC10</i> Synthesis	167
Figure 4.14: Transformation of the Linearised <i>ProSFTPC10</i> Vector into X-33	168
Figure 4.15: Purification of ProSP-C10-SMALP	170
Figure 4.16: Purification of ProSP-C10-SMALP	171
Figure 4.17: CD Spectrum and Accompanying HT Trace of ProSPC-10 in a SMALP	172
Figure 4.18: Langmuir Trough Isotherms of ProSP-C10 and BmrA (Control)	174
Figure 4.19: Maximum Surface Pressure of ProSP-C10 in SMALP	175
Figure 4.20: Detergent Solubilisation Screen of ProSP-C10	177
Figure 4.21: Coomassie Stained Gel of ProSP-C10 Purification	179
Figure 4.22: Solubilisation of ProSP-C10 with $\beta$ -OG	180
Figure 4.23: Coomassie Stained Gel of ProSP-C10 Purification with $\beta$ -OG	181
Figure 4.24: Sucrose Gradient of Reconstituted ProSP-C10	183
Figure 4.25: <i>SFTPC</i> Synthesis	185
Figure 4.26: SP-C Sequence Alignment	185
Figure 4.27: SP-C Expression Screen	186
Figure 4.28: Scale Up of a High SP-C Expressing Colony	187
Figure 4.29: SP-C Detergent Solubilisation Screen	189
Figure 4.30: Solubilisation and Purification of SP-C with Fos-choline 12	190
Figure 4.31: Purification Screen of SP-C Solubilised in $\beta$ -OG, DM, Fos and MNG	192
Figure 4.32: Solubilisation and Purification Gel of SP-C with $\beta$ -OG	194
Figure 4.33: Maximum Surface Pressure of SP-C Mixed with POPC Lipids	197

Figure 4.34: Langmuir Trough Isotherms of SP-C Mixed with POPC	198
<b>Chapter 5</b>	
Figure 5.1: Extraction Process of Lipids Encapsulated in a SMALP for Analysis by Mass Spectrometry	201
Figure 5.2: Mass Spectrometry Analysis of Lipids Around ProSP-C, ProSP-C10, ProSPB, SP-B and SP-C	206
Figure 5.3: Lipid species Identified by Mass Spectrometry Represented in Venn Diagrams	210
<b>Chapter 6</b>	
Figure 6.1: <i>Ab Initio</i> Model of SP-B	215
Figure 6.2: 2D Structure of SP-B	216
Figure 6.3: Amino Acid Interaction of Helices 1 and 3 with the Membrane	217
Figure 6.4: Snapshots of the Initial and Final Conformations of SP-B in an Implicit Membrane	219
Figure 6.5: SP-B Membrane Insertion	220
Figure 6.6 <i>Ab Initio</i> Model of SP-C	222
Figure 6.7: SP-C Membrane Orientation	223
Figure 6.8: Snapshots of the Initial and Final Conformations of SP-C in an Implicit Membrane	224
Figure 6.9: Close Amino Acid Interactions of SP-C in a 'V' Conformation	225
Figure 6.10: Schematic Illustration of SP-B and SP-C in Tear Film Lipid Spreading	227
Figure 6.11: Period Simulation of SP-B	229
Figure 6.12: Period Simulation of SP-C	230
Figure 6.13: Protein-Lipid Interactions of SP-B	231
Figure 6.14: Protein-Lipid Interactions of SP-C	232
Figure 6.15: An Explicit Bilayer Membrane in the Absence of SP-B	233
Figure 6.16: An Explicit Bilayer Membrane with SP-B	234
Figure 6.17: An Explicit Bilayer Membrane with SP-C	235

# Chapter 1

## Introduction

SP-B and SP-C are pulmonary surfactant proteins discovered in the eye's tear film in 2007 (Lars Brauer 2007).

Both SP-B and SP-C have long been implicated in Respiratory Distress Syndrome (RDS), (Avery and Mead 1959) and more recently in Dry Eye Disease (DED) (Posa *et al* 2017).

In the case of pulmonary surfactant production, RDS can occur during the last six weeks of a full term pregnancy, with babies born before this time possibly being deficient in surfactant. Inadequate levels of SP-B and SP-C have been associated with premature death in babies.

However, better diagnosis and treatment via artificially abstracted protein (animal sources) has added to a suite of treatments greatly increasing premature baby survival rates. One renowned case of loss, from the USA, was of one of the Kennedy Family's babies. This potentially gave stimulus to this area from over 50 years ago.

The mechanistic role of SP-B and SP-C is more established in the lung than in the tear film and there is growing speculation that the two processes are similar.

In dry eye disease (DED) an over-expression of both SP-B and SP-C has been observed (Posa *et al* 2017), but the link between overexpression of the surfactant proteins and dry eye is not well understood.

In addition to this, the absolutely precise role of SP-B and SP-C in tear film lipid spreading, at the molecular level, also remains unclear.

This thesis is focused on the synthesis and characterisation of recombinant forms of these proteins in *Pichia pastoris* yeast. This is to enable their biophysical and biochemical analysis, with a view to understanding their functional roles in the tear film lipid layer, as well as potentially providing an alternative source to animal derived surfactants which may have immunological effects.

### 1.1.1 The Tear Film

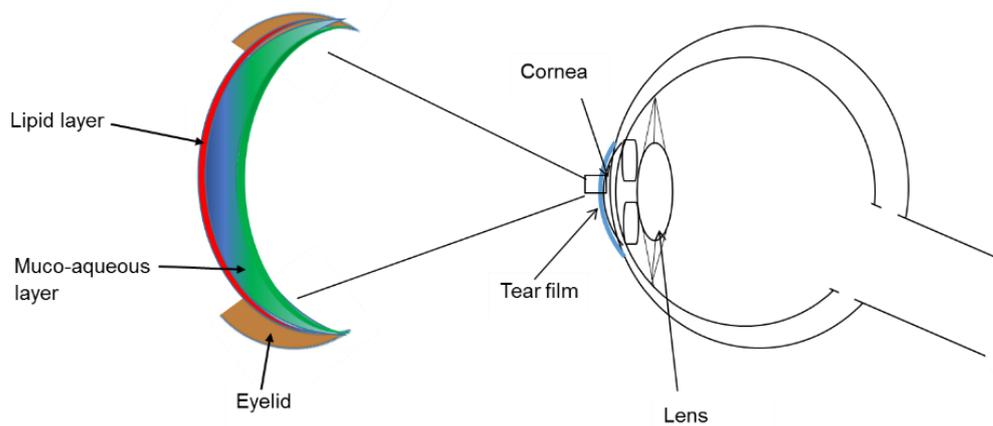
The tear film is a multi-layered structure that covers the surface of the eye (McCulley and Shine 1997).

From the time the tear film break up is initiated, the tear film structure is compressed and re-spread on each complete blink of the eye (Holly 1973).

Functions of the tear film therefore include: provision of an optically-high-quality surface to the cornea; removal of cellular debris and foreign matter; lubrication of the eyelids, conjunctiva and cornea: all so as to prevent mechanical damage from the high pressures of blinking.

The overall process also supplies the cornea with nutrients and oxygen as well as vitally preventing evaporation of the underlying aqueous layer (Holly and Lemp 1977).

The current accepted model of tear film structure is illustrated below in Figure 1.1.



**Figure 1.1 Schematic Representation of the Structure of the Eye and the Tear Film**

The tear film is the ocular fluid covering the cornea.

Previous models suggested that the tear film was made up of a distinct mucus, aqueous and lipid layer. The mucus layer and aqueous layers are more interspersed than previously thought. Mucoaqueous layer (green/blue); lipid layer (red).

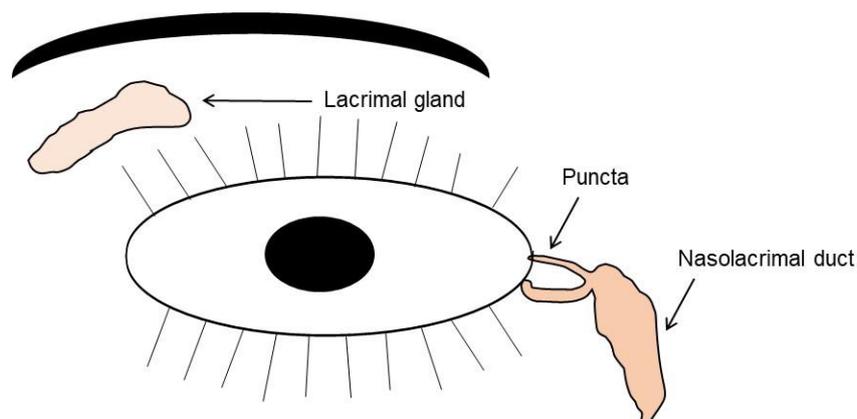
Tear film thickness is reported to range between 6-7  $\mu\text{m}$  and it adheres to the cornea with the help of a mucin rich layer.

Classical tear film models divided the tear film structure into distinct layers made up of a mucus layer, an aqueous layer and a lipid layer (Wolff 1954).

More recent theories, suggest that the mucus and aqueous layers are more dispersed and therefore less distinct than previously suggested by those earlier “classical” models (Green-Church *et al.* 2011). The different layers are secreted by glands and epithelial cells around the eye. See Figure 1.4 later for a more detailed illustration of the tear film.

The lacrimal gland, found above the eye secretes the aqueous layer; tears secreted from the lacrimal gland travel into lacrimal rivers along the lid margins and are secreted over the cornea.

After a blink, the tear fluid flows out through the puncta and continues down the nasolacrimal duct (Millodot 2009). Parasympathetic and sympathetic nerves innervate the lacrimal glands (Sibony *et al.* 1988). The lacrimal glands and puncta are illustrated below in Figure 1.2.



**Figure 1.2 Location of the Lacrimal Glands**

The lacrimal gland is found above the eye and secretes the aqueous layer.

Tear secretions from the lacrimal gland travel along the lacrimal rivers and are secreted onto the surface of the cornea.

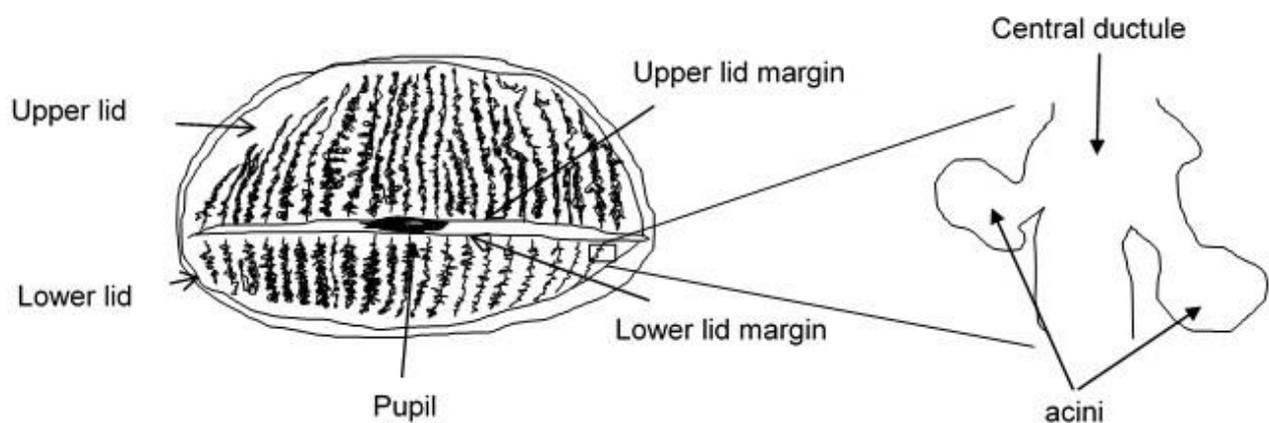
After a blink, the tears flow through to the puncta and continue to the nasolacrimal duct.

Tear fluid that covers the eye constitutively is known as basal tears.

These are different from reflex tears which can be produced as a result of introducing an irritant to the eye. There are also emotional tears, brought on by sadness, joy or stress.

The lipid layer is secreted from meibomian glands (Wolff 1954). Each meibomian gland is made up of small clusters of acinar cells connected by fine ductules to a main duct, which empty onto the lid margin, see Figure 1.3, below.

These meibomian lipids collect in marginal reservoirs and are then spread onto the pre-ocular surface with each blink (Bron *et al.* 2004).



**Figure 1.3 Meibomian Gland Structure**

The meibomian glands line the upper and lower eyelid.

A single meibomian gland is composed of multiple secretory acini that are arranged circularly around a central duct.

Meibomian lipids are produced in the acini and travel down the central ductules and deposited on the upper and lower margins.

### 1.1.2 The Mucoaqueous Layer

A mucus layer covers the surface of the corneal epithelium, and is also referred to as the glycocalyx layer.

The corneal surface is formed of corneal and conjunctival epithelium, it forms a highly folded surface which forms finger like projections. Membrane-tethered mucins are attached to these epithelial protrusions and this helps the tear film adhere to the corneal surface (Dilly 1985).

The tightly packed arrangement of epithelial cells and membrane tethered mucins, also helps to prevent pathogen adherence (Govindarajan and Gipson 2010).

Also present throughout the mucoaqueous layer are gel forming mucins (Gipson 2004).

The mucin family is made up of 20 different glycoproteins (Corfield 2015); these glycoproteins having tandem repeats of serine and threonine amino acids called 'mucin domains' and form sites for O-linked glycosylation (Gendler and Spicer 1995).

Membrane tethered mucins are produced by the conjunctival epithelia.

MUC1, MUC4 and MUC16 are the major tethered mucins found in the mucus layer (Argueso *et al.* 2003). The molecular weight of membrane tethered mucins ranges from: 120-300 kDa (MUC1), 900 kDa (MUC4) and 20MDa (MUC16).

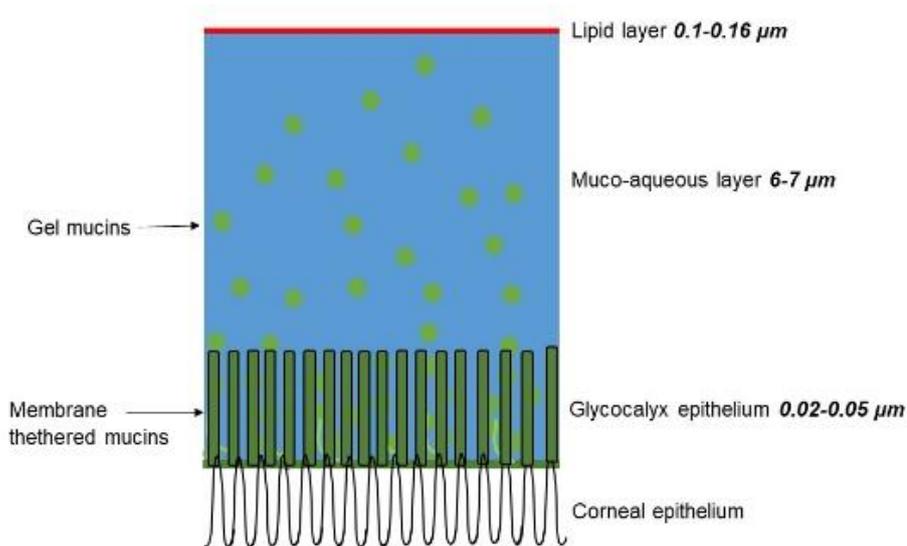
The difference in molecular weights stem from differences in genetic polymorphisms between individual mucins, the alleles being co-dominantly expressed (Govindarajan and Gipson 2010).

Gel forming mucins form goblet cells of the conjunctiva. MUC5A is the major gel mucin secreted by serous cells of the lacrimal gland (Nielsen *et al.* 1997).

Also present are MUC2, MUC6, and MUC19 (Hattrup and Gendler 2008).

Mucins are shown as part of a proposed model (Gipson 2009) which is the currently accepted view, see below in Figure 1.4.

A schematic model is shown below in Figure 1.4.



**Figure 1.4 Mucoaqueous Layer Structure**

The glycoalkyx is formed of columnar transmembrane mucins (dark green) associated with the corneal epithelium.

The soluble mucins are represented by circles in the Mucoaqueous layer.

[The size of the different tear film layers have been referenced from Holly and Lemp 1977; the illustration is not to absolute scale].

The mucoaqueous layer contains electrolytes such as sodium, potassium, calcium, magnesium, chlorine, phosphate and bicarbonate. Metabolites such as amino acids, urea, glucose, and lactate are also present (Munoz-Hernandez *et al.* 2016).

A large number of proteins also exist in the aqueous layer and have been reported to range in number from 491 (de Souza *et al.* 2006) to 1800 (Willcox *et al.* 2017).

The most common proteins in the tear film are lysozyme, lactoferrin, lipocalin, IgA and albumin (Li *et al.* 2005). Lysozyme is found at a concentration of 2.5 mg/ml and is found to be highly effective against gram positive bacteria (Nash *et al.* 2006). Lactoferrin has antimicrobial activity against gram positive, gram negative, fungi, yeasts, and parasites (García-Montoya *et al.* 2012, Flanagan and Wilcox 2009).

Moll's glands found in the eyelid, secrete lysozyme, lactoferrin,  $\beta$ -defensins and IgA (Bräuer *et al.* 2012).

Proteins in the tear fluid, and their concentrations, have been used as biomarkers for eye diseases. For example, lactoferrin concentrations have been found to be lower in Sjogren syndrome and dry eye (Willcox *et al.* 2017).

Sjogren syndrome is an autoimmune condition that is characterised by symptoms which include dry eyes and a dry mouth (reference: *BMJ* 2016)

### **1.1.3 The Lipid Layer**

The lipid layer is the outermost layer of the tear film.

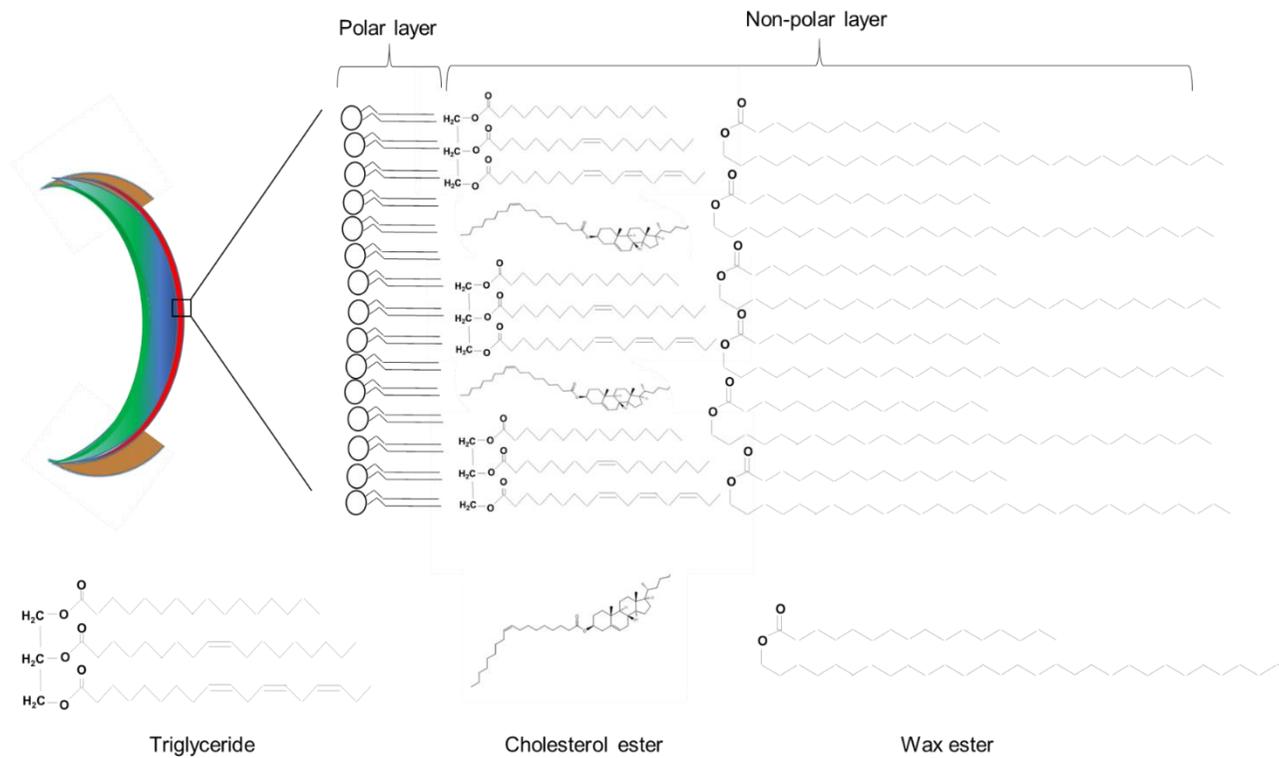
It has an important role in stability (Craig and Tomlinson 1997) and helps to prevent evaporation of the aqueous layer (McCulley and Shine 1997).

The lipid layer thickness is reported to range between 42 nm (King-Smith, Hinel and Nichols 2010) and 160 nm (Green-Church *et al.* 2011); there being varying thickness across the tear film (Willcox *et al.* 2017).

Changes to the lipid component of the tear film manifest in conditions such as dry eye (King-Smith *et al.* 2010).

The tear film lipid layer structure is composed of a polar and non-polar phase (McCulley and Shine 1997).

A deduced, potential model of the tear film lipid layer adapted from the model proposed by McCulley and Shine (McCulley and Shine 1997) is illustrated, see below in Figure 1.5.



**Figure 1.5 Tear Film Lipid Layer Structure**

This illustration is an adaptation of the model proposed by (McCulley and Shine 1997).

The polar layer is made up of phospholipids and fatty acids, while the non-polar layer is made up of hydrophobic lipids such as cholesterol and wax esters.

Details of the lipid composition of each of these layers can be found in Table 1.2 below. [Cholesterol Oleate Image reference: commons.wikimedia.org.]

The meibomian gland secretes lipids found in the lipid layer. Excreted lipids from the meibomian gland are termed 'meibum' or 'meibomian lipids.'

The excreted lipids form the lipid layer which itself is divided into a non-polar layer and a polar layer.

The non-polar layer is composed of triglycerides, wax esters and cholesterol esters. These hydrophobic lipids help to prevent evaporation of the underlying aqueous layer.

To achieve effective spreading, the non-polar layer sits above a 'lubricating' polar layer, this aiding a re-spread of the lipid layer after a blink (Rantamaki *et al* 2011).

The polar layer is composed of phospholipids such as phosphatidylcholine (PC), phosphatidylethanolamine (PE), sphingomyelin (SM), ceramides, and cerebroside.

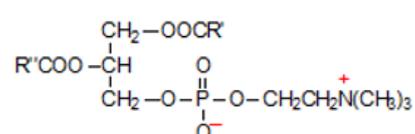
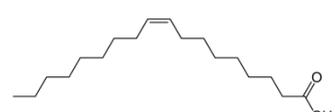
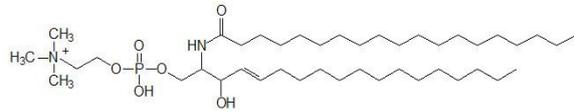
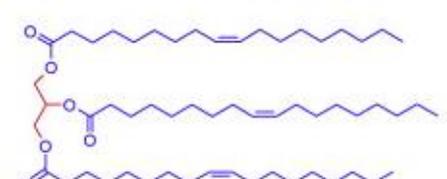
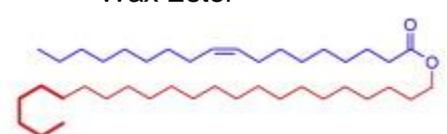
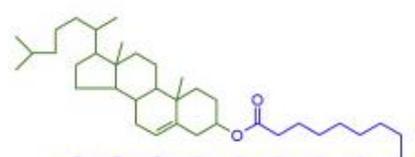
More recently, very long chain (*O*-acyl)- $\omega$ -hydroxy fatty acids (OAHFA) have been detected in meibomian lipid secretions, and these have been suggested to form the polar layer (Butovich, Wojtowicz and Molai 2009).

Details of the lipid composition and structures can be found below in Table 1.1.

**Table 1.1 Structures of the Lipids making up the Polar Layer and**

**Non-Polar Layer of the Tear Film**

The (O-acyl)- $\omega$ -hydroxy fatty acid, oleic acid, and the triglyceride, triolein are the most abundant polar lipids. A general structure is shown for wax esters but the most common fatty acids that were identified were C16, C17 and C18 (Butovich *et al.* 2009). Structures were taken from (Butovich 2013).

	<b>Location in lipid layer</b>	<b>Examples</b>
Glycerophospholipids e.g. POPC, POPE	Polar layer	Phosphatidylcholine 
Fatty acids OAHFA	Polar layer	Oleic acid 
Ceramide e.g. SM	Polar layer	Sphingomyelin 
Triglyceride e.g. Triolein	Non- polar layer	Triolein 
Wax esters C16, C17 and C18	Non-polar layer	Wax Ester 
Sterols	Non-polar layer	Cholesterol ester 

All of the lipids found in the tear film layer are derived from the meibomian glands.

However, differences have been observed in lipids derived from the meibum and the lipid composition of tear film lipids (Brown *et al.* 2013).

The main difference between the two was the presence of phospholipids in the tear film, these not having been previously detected in earlier studies.

It is recognised now that this is due to the phospholipid composition being extremely low, alongside differences in sample collection and analysis (Butovich, Wojtowicz and Molai 2009).

This lack of agreement on the phospholipid composition, posed a challenge when deciding which phospholipids to use in the membrane models later in Chapter 6.

Table 1.2 shown below, demonstrates the reported levels of lipid groups and their abundances from eight studies of tear film, covering from 1978 to 2014.

Perhaps unsurprisingly, there is a wide variation between the eight studies over these 36 years, much of this being due to differences in sampling techniques and improvements in the sensitivity of those techniques.

**Table 1.2 Composition of Lipids making up the Polar Layer and Non-Polar Layer of the Tear Film**

Earlier studies were based on techniques that were not as sensitive as the techniques available today, and as a result, were only able to detect lipid species and not their abundances.

Abbreviations: WE, wax esters; Chl Est, cholesterol ester; TAG, triglyceride; POPI; POPE, CER, cerebrosides; TLC, thin layer chromatography; GC, gas chromatography; MS, mass spectrometry; HPLC, high performance liquid chromatography; LC, liquid chromatography; ESI, electrospray ionization; UPLC, ultra-performance liquid chromatography; RP, reverse phase.

Study/techniques used for analysis	Lipid identity/content					
	Wax ester	Cholesterol ester	Triglycerides	Free fatty acids	Sphingolipids	Phospholipids
(Tiffany 1978) TLC ; variations found between samples	WE	Chl est	TG			
(Nicolaidis <i>et al.</i> 1981); GC-MS	32.32%	27.28%	3.70%	1.98%	Polar lipids: 14.83%	
(McCulley and Shine 1997); human meibum; samples not pooled	68%	16%	6%	1%	1.5%	4%
(Butovich <i>et al.</i> 2007); HPLC MS	WE	Chl-est	TAG			
(Campbell, Griffiths and Tighe 2011); LCMS ESI						No phospholipids
(Saville <i>et al.</i> 2011); phospholipids on contact lenses ESI/MS		Chl				PC SM
(Rantamaki <i>et al.</i> 2011); UPLC-MS			5%		3%	PC:70% PE:20% Cer:3%
(Man Lam <i>et al.</i> 2014); RP HPLC ESI MS	35.21%	44.82%	2.84%		1.77%	6.45%

#### 1.1.4 Tear Film Spreading

The tear film is a dynamic, multi-layered structure that is constantly compressed and expanded by blinking.

The stability of the tear film is compromised by soluble mucins traversing through to the lipid layer causing instability and tear film break-up (Holly and Lemp 1977; Brown and Dervichian 1969).

Blinking helps to provide more stability to the tear film, but the exact molecular level mechanism has been poorly understood.

Understanding the behaviour of the tear film lipid layer under compression and expansion has been challenging at the molecular level.

Molecular models have recently helped to bridge the gap in knowledge, helping us to understand more about the spreading mechanism.

Various computational models of the tear film have been developed, but these have been in isolation to each other; with many investigating solely either the polar layer or the non-polar layer, see below in Figure 1.6.

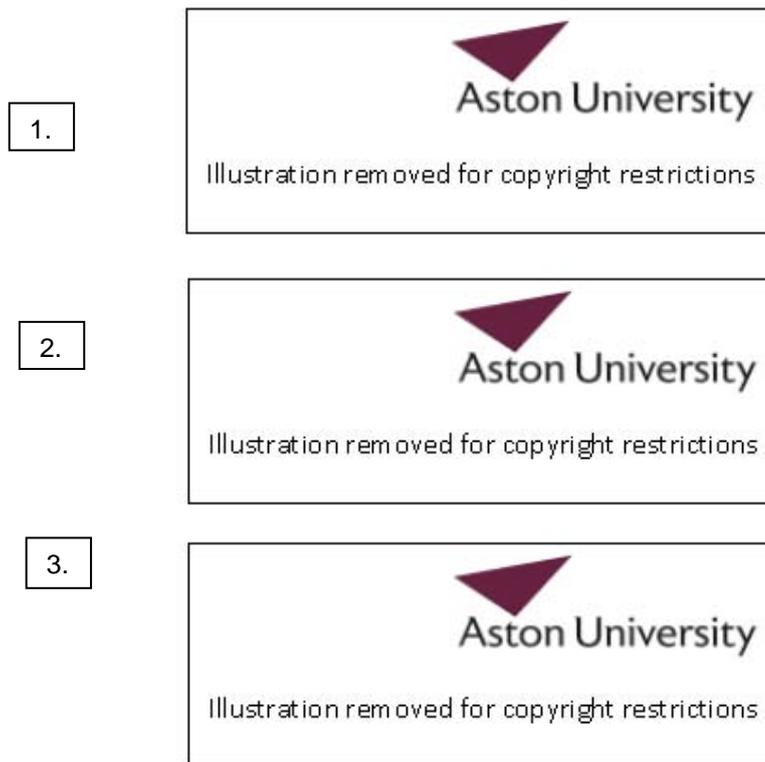
Compression of the lipid layer has been suggested to form inverse vesicles in the polar layer (Wizert, Iskander and Cwiklik 2014) (2) (Cwiklik 2016) (1); and multi-layered structures in the hydrophobic layer (Kulovesi *et al.* 2010; Patterson, Vogel and Prenner 2016) (3). The formation of multi-layered structures has also been demonstrated *in vitro* (Millar 2013).

The validity of such models e.g. Wizer, Iskander, Cwiklik 2014 should thus be questioned, especially in an area that has no standardised techniques for even tear sample collection or analysis.

This has meant that models, see below in Figure 1.6 have been developed from papers wherein the phospholipid concentrations also apparently varied greatly, thus making comparison/additive development particularly challenging.

The models also fail to include SP-B and SP-C.

Recognising the likely contributory surface activity of SP-B and SP-C could help to understand all the components which create the surface tension differences which facilitate lipid spreading.



**Figure 1.6 Molecular Models of Tear Film lipid Spreading**

Compression of the tear film polar layer has been suggested to form inverse vesicles (panels 1 and 2) and an 'island' structure in the non-polar layer (panel 3).

The relaxed state is reflected by column A in panels 1 and 2, as the polar layer is gradually compressed it causes curvature of lipids represented by column B and finally the formation of vesicles in panel C.

Compression of the non-polar layer then causes the lipids to form in a central 'island' (panel 3).

Images taken from (Cwiklik 2016) Wizert, Iskander and Cwiklik 2014 (panel 2) (Kulovesi *et al.* 2010 (panel 3).

### 1.1.5 Dry Eye

“Dry eye” is a multifactorial disease of the ocular surface characterized by a loss of homeostasis of the tear film.

The lipid layer has an important role in helping to prevent evaporation of the underlying aqueous layer.

It is important that the lipid layer spreads effectively after each blink.

An inadequate lipid spreading can cause increased tear evaporation, so leading to symptoms of ocular discomfort, known as dry eye.

Dry eye is accompanied by ocular symptoms, in which tear film instability and hyperosmolarity; ocular surface inflammation and damage; and neurosensory abnormalities all play etiological roles (the dry eye workshop report DEWS 2017)”

Post menopausal women, older adults and contact lens wearers can be particularly affected by dry eye (Schaumberg *et al.* 2003; Schaumberg *et al.* 2009).

Although there is a wealth of research that links dry eye to increased tear evaporation, the lipid composition of dry eye patients was **not found** to be different from the ‘normal’ lipid profile reported in other studies (Man Lam *et al.* 2014).

Although this is only one study, the problems causing dry eye may not however relate to lipid composition but rather to how those lipids are spreading.

Other factors that may have been overlooked until recently, include; the levels of SP-B and SP-C, they being found to be higher in patients, rather than at some lower optimum, with dry eye (Posa *et al.* 2017).

Understanding the role of SP-B and SP-C may well help to link some of the physiological symptoms present in dry eye to the underlying molecular mechanism. For instance the exposure of SP-B and SP-C at the lipid surface could make the proteins susceptible to oxidation; so causing changes to the secondary structure which in turn could have an effect on lipid spreading.

### 1.1.6 Surface Tension

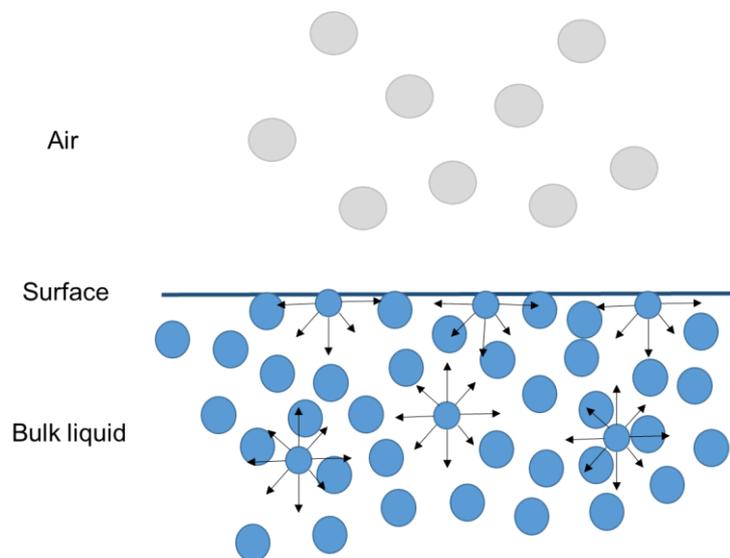
The lipid layer and the aqueous layer form a liquid-liquid (L-L) interphase, while the lipid layer is situated at a gas (air)-liquid interphase (G-L).

Differences in surface tension forces (L-L) are considered to be the driving forces that help the lipid layer spread (Berger and Corrsin 1974).

From a molecular perspective, the surface tension of water can be seen as the force that acts on water molecules at the surface compared to forces acting in the bulk, see below in Figure 1.7.

There are no net forces acting on molecules in the bulk i.e. the molecule is pulled in no particular direction.

Molecules at the surface will have attractive forces with neighbouring molecules and as a result, there is a tendency for those water molecules to be pulled together at the surface, thus seeking to minimize the area of the surface (Barnes and Gentle 2005).



**Figure 1.7 Surface Tension**

Water molecules in the bulk, are able to interact with each other, but those molecules at the surface of the water are unable to interact with anything above as there is nothing above.

The strength of the interactions at the surface cause the water molecules to be pulled into the surface and form a surface film.

### 1.1.7 Langmuir Trough and Surface Activity

SP-B and SP-C have an important role in lowering the surface tension in the alveoli, which helps reduce alveolar collapse (Avery and Mead 1959)

A Langmuir trough, detailed below, was used to test if the recombinant SP-B and SP-C, production described in Chapters 3 and 4 respectively, displayed surface activity.

Surface activity can be measured by surface tension and surface pressure. These reflect differences in the downward force (surface tension) and the outward pressure exerted by the molecules at the surface (surface pressure).

The Langmuir trough resembles a shallow tray, which is filled with HPLC-grade water to form a sub-phase; SP-B and SP-C were then injected on the surface of this bulk sub-phase.

The surface area of the sub-phase oscillates between an expanded and compressed state, these being achieved by outward or inward movement of computer controlled barriers.

These computer controlled barriers, on compression, gradually decrease the area available to the SP-B and SP-C molecules from 400 cm<sup>2</sup> to 100 cm<sup>2</sup>.

The changing surface tension and pressure are measured and recorded using a Wilhelmy plate balance and data logging.

A Wilhelmy plate, with dimensions ( $l \cdot w \cdot t \cdot d$ ), experiences a downward force ( $F$ ). This is due to a combination of the surface tension of the liquid and the weight of the plate. The relationship is shown in the equation below:

$$F = [g \cdot (P_{\text{plate}} \cdot l \cdot w \cdot t)] - [g \cdot (P_{\text{liquid}} \cdot d \cdot w \cdot t)] + [2 \cdot (w + d) \cdot \gamma \cdot \cos\theta]$$

**$P_{\text{plate}}$** : density of the plate;

**$P_{\text{liquid}}$** : density of the liquid;

**$\gamma$** : surface tension of the liquid;

**$\theta$**  the contact angle of liquid to the Wilhelmy plate;

**$g$** : acceleration due to gravity ( $9.8 \text{ m/s}^2$ ).

The downward force on the plate is affected by gravity ( $g$ ) and buoyancy ( $P_{\text{plate}} \cdot l \cdot w \cdot t$ ) -  $[g \cdot (P_{\text{liquid}} \cdot d \cdot w \cdot t)]$ .

The effect of gravitational force is negated if the plate is both tared and kept at a constant depth, so eliminating the effects of buoyancy.

To ensure the plate is perfectly wetted so that the contact angle can be set to 0 degrees, the Wilhelmy plate is submerged to 2mm below the sub-phase.

Resolution of  $F$ , then simplifies to the following:

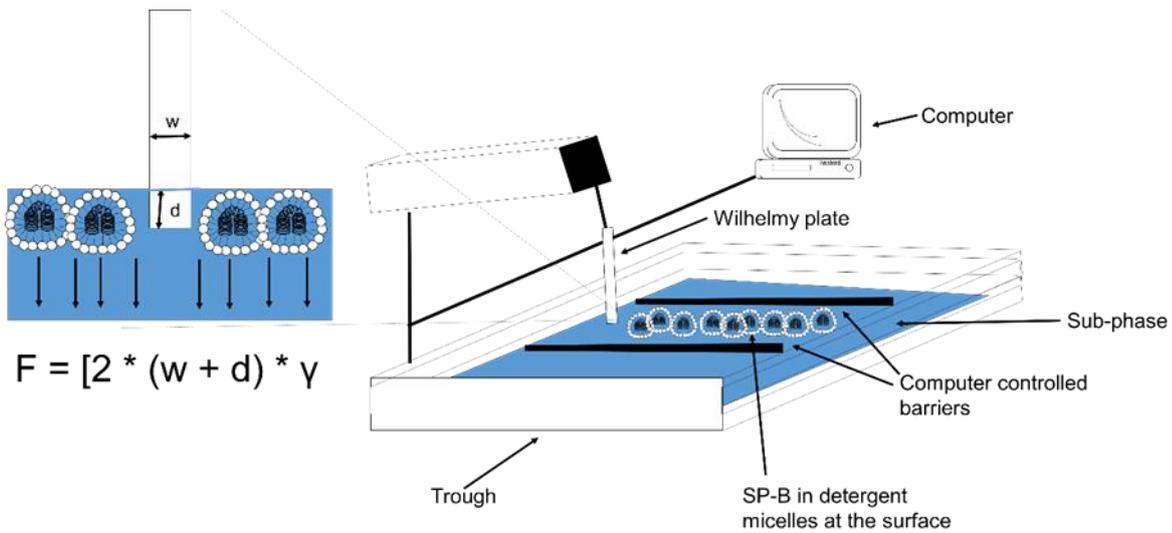
$$F = [2 \cdot (w + d) \cdot \gamma]$$

where  $\gamma$  = surface tension

Surface tension can thus also be expressed as a function of the downward force.

$$\gamma = F / [2 \cdot (w + d)]$$

The Wilhelmy plate and other components of the Langmuir trough are illustrated in Figure 1.8 below.



**Figure 1.8: Schematic Illustration of the Langmuir Trough and Associated Components**

A Langmuir trough was used to measure the effect on surface tension of SP-B and SP-C molecules.

SP-B and SP-C were carefully added on top of a sub-phase, here composed of HPLC grade water.

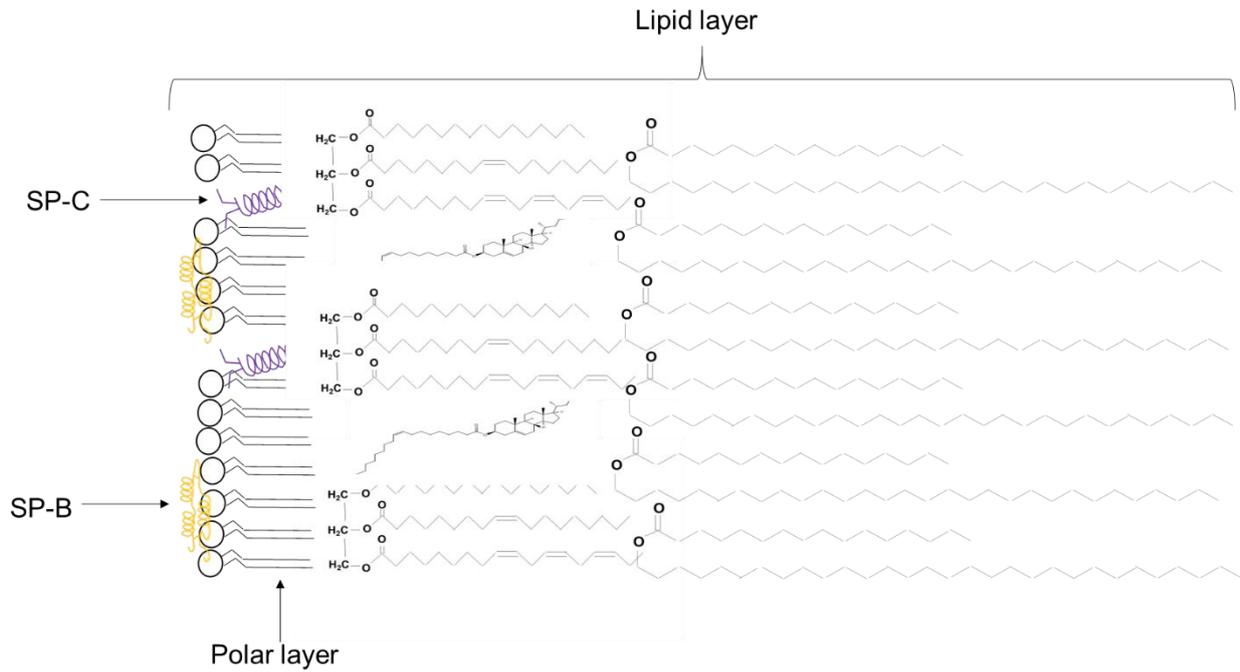
Computer controlled barriers increase and decrease the area available to the molecules, the resultant surface tension effects are measured by a Wilhelmy plate.

Collected pressure vs surface area data is represented in an isotherm.

### 1.2.1 Discovery of SP-B and SP-C in the Tear Film

SP-B and SP-C are pulmonary surfactant proteins that have recently been discovered in the tear film (Brauer *et al.* 2007).

A considered, potential model of the interaction of SP-B and SP-C with phospholipids is shown below in Figure 1.9.



**Figure 1.9** Considered, Potential Model of the Interaction of SP-B and SP-C in the Tear Film

The tear film is divided into a polar and non-polar layer, see above in Table 1.2 for a detailed composition of tear film lipids.

The polar layer is made up of phospholipids and fatty acids.

SP-B and SP-C are thought to be associated with the phospholipid head groups and the acyl tails respectively.

The interaction of SP-B with phospholipid head groups and SP-C with the acyl chains is how SP-B and SP-C are considered to interact in the (Johansson *et al.* 1991; Johansson 1991)

### 1.2.2 Pulmonary Surfactant

Pulmonary surfactant is a protein-lipid complex that covers the surface of the lungs.

It is made up of 90% lipid and 10% protein (Veldhuizen *et al.* 1998).

Four pulmonary surfactant proteins, SP-A (King *et al.* 1973), SP-B, SP-C and SP-D (Persson *et al.* 1989), make up the surfactant population.

Note; SP-A and SP-D have immunomodulatory functions.

As briefly mentioned earlier, the pulmonary surfactant lining is one of the last biological interfaces to develop *in utero*. Thus, pre-term infants born prematurely may lack pulmonary surfactant and require exogenous replacements; the condition being known as Respiratory Distress Syndrome (RDS).

Research into synthetic surfactant surged after the death of Patrick Bouvier Kennedy, the prematurely-born son of Jackie and John F Kennedy (Halliday 2008), who died of RDS.

Although the death of JFK's son helped drive research into pulmonary surfactants, the importance of pulmonary surfactants helping to lower the surface tension was a discovery which had been first made by Neergaard in 1929 (Neergaard 1929). He observed the ability of surfactants to reduce surface tension in excised lung.

It then took some 25 years after Neergaard made his discovery of the link between that absence of pulmonary surfactants and resultant difficulties in breathing to be more fully understood (Clements 1956, Pattle 1955).

Again, a further 5 years later, evidence was published that indicated preterm infants who had deficiencies in pulmonary surfactant were dying of RDS, (Avery and Mead 1959).

In RDS, pre-term infants fail to produce enough pulmonary surfactant to sustain breathing and need exogenous (to date animal-derived), surfactant replacements until they are able to start producing their own. The use of exogenous surfactant has greatly reduced the incidence of RDS (Phillip 1995) and the survival rate of babies with RDS is now much higher (Vorbroker *et al.* 1995).

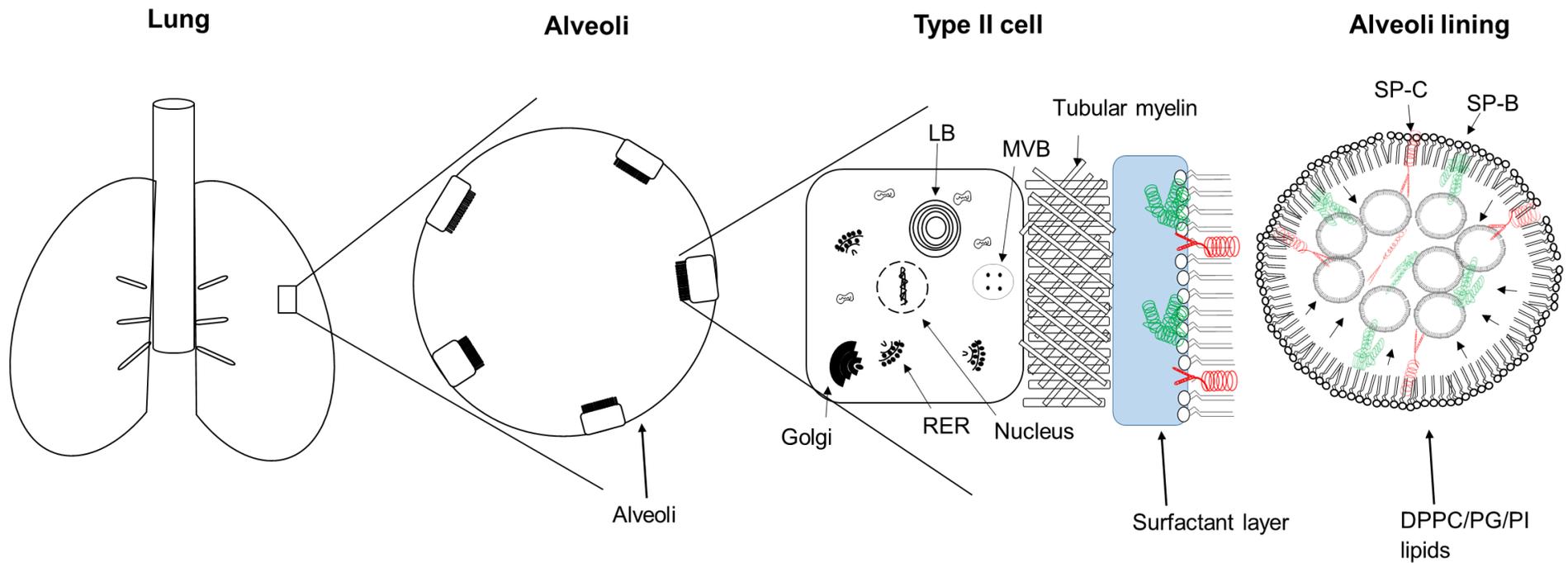
### 1.2.3 Pulmonary Surfactant Composition

80-90% of the pulmonary surfactant mass is composed of phospholipids, some 60-70% of which comprises; dipalmitoylphosphatidylcholine (DPPC), phosphatidylglycerol (PG) and phosphatidylinositol (PI).

Some 2% to 8% of the pulmonary surfactant mass is made up of neutral lipids such as cholesterol, with 5-10% composed of pulmonary surfactant proteins SP-A, SP-B, SP-C and SP-D (Possmayer *et al.* 1984, Goerke 1998).

Note: The two hydrophobic proteins SP-B and SP-C are responsible for reducing surface tension and are important for helping to prevent alveolar collapse.

It is the particular SP-B and SP-C proteins which are investigated in this thesis, see below in Figure 1.10 illustrating the surfactant layer physiology within the alveoli.



**Figure 1.10 Pulmonary Surfactant Secretion**

Pulmonary surfactant lines the alveolar walls; its function being to lower surface tension during inhalation and exhalation.

SP-B and SP-C transcripts, synthesised in the nucleus, are translated on the rough endoplasmic reticulum (RER), modified in the Golgi bodies before being trafficked to multi-vesicular bodies (MVB) and lamellar bodies (LB).

The protein-lipid complexes are secreted onto the cell surface through exocytosis to form tubular myelin, where the SP-B, SP-C proteins and lipids adsorb to the surface.

#### 1.2.4 Synthesis of SP-B and SP-C

SP-B is encoded by the *SFTPB* gene.

The primary translation product is 42 kDa 381 amino acids (aa) and this is cleaved to a 7-8 kDa (79 aa) so as to produce a mature peptide. A signal sequence made of 24 aa is located at the N-terminus of the peptide and mediates the translocation of SP-B into the lumen of the endoplasmic reticulum (ER) (Lin *et al.* 1996).

Following translocation into the ER, the SP-B protein is transported to the Golgi and then onto multivesicular bodies (MVB) (Voorhout *et al.* 1992).

SP-B is finally transported from the MVB to lamellar bodies (LB) (Lin *et al.* 1996) where the active protein is cleaved by napsin A (Brasch *et al.* 2003), cathepsin H (Guttentag *et al.* 2003) and pepsinogen (Foster *et al.* 2004).

The sorting of SP-B to LB was studied by expressing deletion constructs in pheochromocytoma (a rat cell line known as PC-12).

These studies indeed demonstrated that the N-terminus peptide was required for sorting.

Transgenic mice were used to confirm *in vivo* that an SP-B construct consisting of the signal peptide, the N-terminus peptide and the mature peptide was appropriately sorted, processed and secreted (Lin *et al.* 1996).

The synthesis and transport mechanism of surfactant lipid dipalmitoylphosphatidylcholine (DPPC), from the ER to the LB, occurs through an ABCA3 transporter on the LB membrane.

SP-B and SP-C are assembled with surfactant phospholipids into bilayer membranes that are stored in LB's.

The contents of LB's are secreted into the airway to produce tubular myelin and multi-layered surface films (Whitsett, Wert and Weaver 2015).

Mutations in ABCA3 have also been observed to cause RDS (Anandarajan, Paulraj and Tubman 2009).

Pulmonary surfactant is recycled, and alveolar macrophages play an important role in surfactant uptake and degradation.

SP-C is also synthesized in a precursor form, and is subject to proteolytic cleavage to yield a mature peptide.

It is thought SP-B and SP-C follow the same pathway as well as using the same processing enzymes.

This is shown by experiments where when SP-B is absent, the processing and sorting of SP-C is also altered (Vorbroker *et al.* 1995).

### 1.2.5 Structure of SP-B

SP-B is secreted as a propeptide see below in Figure 1.11.

The N-terminus and C-terminus of proSP-B are then cleaved to produce the active 79 amino acid SP-B peptide (Ueno *et al.* 2004).

MAESHLLQWLLLLLPTLCGPGTAAWTTSSLACAQGPEFWCQSLEQALQCRALGHCLQEY  
WGHVGGADDLCQECEDIVHILNKMAKEAIFQDTMRKFLEQECNVLPKLLMPQCNQVLLDDYF  
PLVIDYFQNQTDSNGICMHLGLCKSRQPEPEQEPGMSDPLPKPLRDPLDPLLDKLVLPVL  
PGALQARPGPHTQDLSEQQ[FPIPLPYCWL~~CRALIKRIQAMIPKGALAVAVAQVCRVVPLV~~  
~~AGGICQCLAERYSVILLDTLLGRMLPQLVCRVLRCSM~~DDSAGPRSPTGEWLPRDSECH  
LCMSVTTQAGNSSEQAIPQAMLQACVGSWLDREKCKQFVEQHTPQLLTLVPRGWDAAHTT  
CQALGVCGTMSSPLQCIHSPDL

#### Figure 1.11 Primary Sequence of ProSP-B

ProSP-B, with a molecular weight of 42kDa, is cleaved to a mature dimeric protein with a molecular weight of 18kDa.

The region highlighted, and between the square brackets above [ ] represents the SP-B protein, the N-terminus and C-terminus regions on either side of SP-B are cleaved.

**Both proSP-B and SP-B have been produced as recombinant proteins in this study.**

Some 52% of SP-B's sequence is composed of hydrophobic amino acids (Glasser 1987).

Because of this high hydrophobicity, and the associated challenges of producing recombinant forms of SP-B, its 3D structure is not yet completely understood. The only structural data available comes from protein samples purified from animal sources using solvents.

SP-B has been shown to contain a high  $\alpha$ -helical content (40-50%) according to circular dichroism and infrared spectroscopy studies (Andersson *et al.* 1995, Pérez-Gil and Cruz. 1993, Vandenbussche *et al.* 1992, Perez-Gil *et al.* 1993).

Structural information on SP-B has been derived from a peptide (mini-B) composed of the N- and C-terminus of SP-B in SDS micelles using NMR (Sarker *et al.* 2007), and FTIR (Waring *et al.* 2005).

Other structures deduced using NMR are of truncated versions of SP-B (Kurutz and Lee 2002, Booth *et al.* 2004).

Structural information relating to the secondary structure have also been derived from sequence similarity of SP-B against NK lysin (Andersson *et al.* 1995), a member of the Saposin superfamily of proteins.

SP-B contains seven cysteines, six of which are responsible for the formation of the 3 intramolecular disulphide bridges.

It is thought the role of the cysteine bonds is to help stabilize the structure (Johansson *et al.* 1991). The N-terminus of SP-B is important for insertion into the bilayer (Sharifahmadian *et al.* 2013).

Figure 1.12, see below, illustrates a structure of mini-B generated using NMR (Sarker *et al.* 2007).



**Figure 1.12 Structure of SP-B**

Structure of SP-B generated by NMR of a synthetic peptide composed of the N and C-terminus regions in SDS micelles (Sarker *et al.* 2007).

Image taken from the protein data bank (Sarker *et al.* 2007).

Gene deletions in animal studies of *SFTPB*, the gene encoding SP-B (Clark *et al.* 1995), or mutations in the gene, are associated with poor prognoses; children with this do not survive beyond the first few months of life (Nogee *et al.* 1993).

SP-B is a member of the Saposin-like Protein (SALIP) family.

Shared characteristics of the SALIP family include amphipathic helical sequences, and conserved cysteine residues that form disulphide bonds (Andersson *et al.* 1995).

Like other members of the Saposin family (Kishimoto, Hiraiwa and O'Brien' 1992), SP-B is synthesized as a 42 kDa proprotein that is later cleaved to an 18kDa active protein (Weaver 1998).

Shared similarities between members of the SALIP family are tabulated, see below in Table 1.3.

**Table 1.3 Similarities and Differences Between Members of the SALIP family**

Members of the SALIP family have conserved hydrophobic amino acid distribution (6 cysteines located at conserved positions) and a high helical content leading to 4-5 amphipathic  $\alpha$ -helices.

	SP-B	NK-Lysin	Saposin A	Saposin B	Granulysin
Identical residues to SP-B		19%	22%	22%	20%
No. of cysteine residues	7	6	6	6	6
Disulphide bridges	3	3	3	3	2
Helices	5	5	4	4	5
Charge	+4	+6	-8.2	-8	+11
Interact with lipids	Permanently associated	X	Intermittently Lipid transfer and antigen presentation		Mediates cytotoxicity
Conserved hydrophobic residues	✓	✓	✓	✓	✓
Location	Lungs, eyes, testis	T cells and NK cells (pigs)	Sphingolipid catabolism		T cells and NK cells (humans)

### 1.2.6 Structure of SP-C

The gene encoding SP-C, *SFTPC* is located on chromosome 8p and is organised into 6 exons and 5 introns.

Organisation of the pro-protein (proSP-C) is as follows: N-terminus, transmembrane domain, linker region and the 'BRI12 linked to familial British and Danish, Dementia, Chondromodulin-1' (BRICHOS) domain (Mulugeta, Nureki and Beers 2015).

These regions are highlighted in the primary sequence of proSP-C, see below in Figure 1.13.

**MDVGSKEVLMESPDPDYSAAPRGRFGIPCCPVHLKRLIVVVVVVLIVVVIVGALLMGLHMS**  
**QKHEMVLMSIGAPEAQQLALSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCY**  
**IMKIAPESIPSLEALNRKVHNFQMECSLQAKPAVPTSKLGQAEGRDAGSAPSGGDPFLG**  
**MAVNTLCG EVPLYYI**

**Figure 1.13 Primary Sequence of ProSP-C**

The region that is not highlighted is the N-terminus, the region in yellow represents the mature domain and the region in green is a linker region. The grey region is the BRICHOS domain.

SP-C is an alpha helical lipopeptide, 4.2 kDa in size (Johansson *et al.* 1994).

The primary sequence is evolutionarily conserved (Johansson 1998).

ProSP-C is cleaved at the N and C-terminus to produce a hydrophobic membrane protein made of 35 amino acids (Glasser *et al.* 1990).

The N-terminus of SP-C has a positive net charge, and includes two palmitoylated cysteine residues (Curstedt *et al.* 1990).

The C-terminal region is enriched in valine residues, forming a highly hydrophobic  $\alpha$ -helix.

The role of the palmitoylation is to help maintain the lowest surface tensions of phospholipid films at high compression rates (Baumgart *et al.* 2010).

SP-C has a transmembrane orientation and is tilted at a 70° angle (Vandenbussche *et al.* 1992, Johansson *et al.* 1994).

The structure of SP-C is shown below in Figure 1.14. The protein from which the structure was solved was derived from porcine lung lavages.



**Figure 1.14 The Structure of SP-C**

The SP-C protein was isolated from porcine lung lavages.

Analysis of SP-C structure, using NMR revealed SP-C, shows an  $\alpha$ -helical transmembrane protein.

The yellow residues represent the hydrophobic amino acids. Image taken from (Johansson 1998).

There are 60 known mutations of SP-C.

Mutations made to the N-terminus of proSP-C, through mutagenesis studies, cause a failure in cleavage of the N-terminus.

Mutations in the cysteine residues 120 and 148 result in endoplasmic reticulum retention and aggregation.

These mutations indicate that the BRICHOS domain may have a role in folding (Mulugeta *et al.* 2015).

Deletion of exon 4 in SP-C resulted in deletion of 37 amino acids (Nogee *et al.* 2001).

Table 1.4, see below, lists some of the mutations in SP-C and their effects.

**Table 1.4 Mutations in SP-C**

Mutations in SP-C lead to aggregation, ER retention and mistrafficking (Mulugeta *et al.* 2015).

Location	Mutation	Effect
N-terminus	Mutagenesis	Failure to cleave the NH <sub>2</sub> domain
Linker (59-93)		Mistrafficking to the MVB
COOH	Deletion of exon 4	Deletion of 37 amino acids
BRICHOS	Cysteine residues 120 and 148	ER retention and aggregation
BRICHOS	SPCL188Q	Intracellular SP-C protein aggregation

### 1.2.7 The Role of SP-B and SP-C in the Lipid Layer of the Lung

During exhalation, the lipid layer in the lung is compressed forming 3D 'vesicle like' structures which adsorb and desorb from the membrane (Cabre *et al.* 2009).

This behaviour of the lung lipid layer is attributed to the high phospholipid composition, which as a result, at high pressures, collapses into vesicles which are able to detach from the membrane during expiration, before re-adsorbing during expansion.

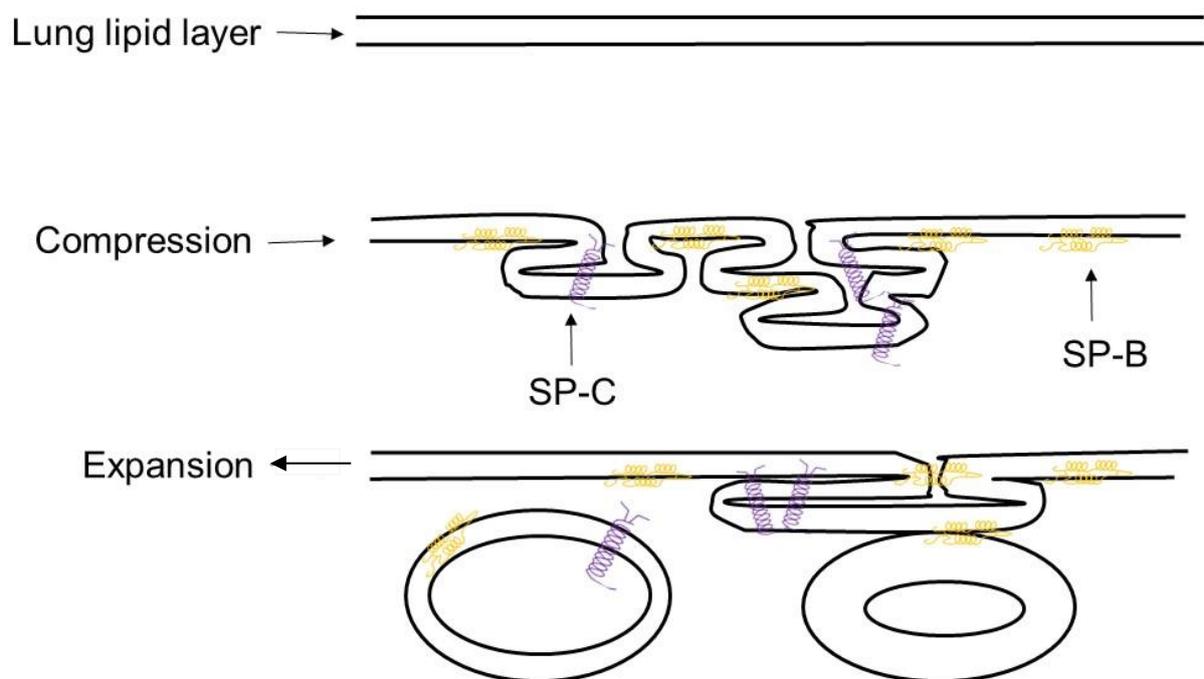
SP-B is involved in promoting interfacial adsorption of surfactant phospholipids (Serrano *et al.* 2007) and is thought to connect monolayers at the air/water interface to bilayers (simulations) (Baoukina and Tieleman 2011).

SP-C is palmitoylated at the amino terminus and is thought to be important for helping anchor the amino terminus to the membrane (Johansson *et al.* 1994) and for helping the vesicles adsorb to the lipid layer on expansion (Baumgart *et al.* 2010).

Involvement of SP-B and SP-C allows the surface tension to be lowered and facilitates gaseous exchange (Schurch *et al.* 2010).

The multi-layered structure formed on compression, has been demonstrated by electron microscopy (Bachofen *et al.* 2005).

A schematic illustration of the role of SP-B and SP-C during compression and expansion of the lung lipid layer is shown, see below in Figure 1.15.



**Figure 1.15 Lipid Spreading in the Lung**

The lung lipid layer is subject to compression and expansion during breathing. The cycle of events during the compression and expansion are considered to be as follows:

Compression of the lipid layer causes expulsion of lipid/lipid complexes from the interface and formation of 3D vesicles.

SP-B promotes establishment of membrane-membrane contacts leading to membrane vesicles that 'separate' away from the membrane

SP-B helps connect vesicle membranes, and SP-C anchors vesicles to the membrane.

SP-B and SP-C help with re-spreading.

### 1.2.8 DISCUSSION - Recombinant Synthesis of SP-B within Previous Work

A previous attempt to synthesize a truncated form of proSP-B in *Pichia pastoris* yeast was unsuccessful (Gil *et al.* 2009).

In that work, residues 24-31 of proSP-B were amplified using PCR, then re-amplified to include the  $\alpha$ -secretion factor and the His<sub>6</sub>-tag present in pPICZ $\alpha$ . The construct was ligated into pGEM-Teasy vector and *E. coli* cells were transformed. The vector was extracted, then linearized using restriction enzymes and transformed into X-33 yeast and grown on YPDS plates with zeocin.

Transformed cells were grown in 200 ml of BMGY and further grown overnight before being spun down and re-suspended until an A<sub>600</sub> of 1.0 into BMMY.

The culture medium was grown for 4 days and recombinant proteins were analysed by western blot - but no bands were found.

The sequence was transcribed and this was checked by reverse transcriptase polymerase chain reaction (RT-PCR)

The lack of components needed for translation in yeast cells, and improper folding, were cited as reasons for the lack of secreted expression (Gil *et al.* 2009).

More recent research on SP-B has focused on mimics of SP-B e.g. KL4 (Kronqvist *et al.* 2017).

A summary of the various methods used to synthesis recombinant SP-B is shown, see below in Table 1.5.

**Table 1.5 Comparison of the Methods Used to Synthesise and Extract Recombinant ProSP-B and SP-B**

Methods for the recombinant synthesis of SP-B and proSP-B are outlined below.

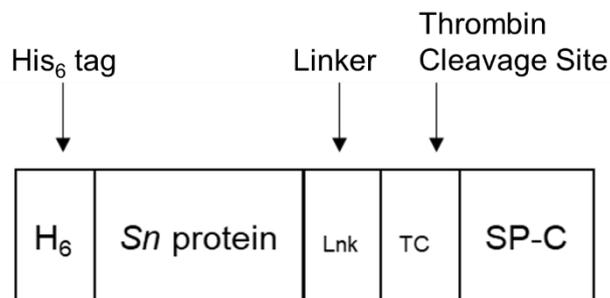
	<b>(Weaver 1996)</b>	<b>(Zaltash and Johansson 1998)</b>
Construct	SP-B	ProSP-B
Host	<i>E.coli</i>	<i>E.coli</i>
Vector	pProEx1	PET15b
Cloning	Fusion protein	Restriction sites included in primers. PCR of proSP-B DNA followed by digestion using enzymes
Transformation	DH10B	BL21 <i>E.coli</i>
Solubilisation	6M GH	SDS
Purification	Ni-NTA	Talon column
Yield	-	1.5 ml/L
Secondary structure analysis	-	CD
Functional activity	-	-
Limitations		No surface activity data

### 1.2.9 DISCUSSION - Recombinant Synthesis of SP-C

The production of recombinant SP-C was first established in 2006 (Lukovic *et al.* 2006).

The authors cited using a fusion protein to overcome the difficulties of producing recombinant SP-C.

From left to right, the construct had a His<sub>6</sub>-tag followed by the hydrophilic *Staphylococcus nuclease* (SN) protein, a linker protein and thrombin cleavage site. These are linked to SP-C, as illustrated in Figure 1.16 below.



**Figure 1.16 SP-C Fusion Protein Construct used by Lukovic *et al.* 2006**

The His<sub>6</sub>-tag is placed at the N-terminus and adjacent to the *Staphylococcus nuclease* (SN) protein. This is followed by a linker protein and thrombin cleavage site that allows cleavage of SP-C during the final stages of purification.

Two point mutations were introduced (Cys5Phe and Cys6Phe) because bacteria lack the ability to palmitoylate. The replacement with Phe remains close to the native sequence in such as dog and mink surfactant (Lukovic *et al.* 2006).

The SP-C construct was expressed in *E.coli* and extracted using 1% N-lauroylsarcosine (Lukovic *et al.* 2006).

This method demonstrated that 4 mg SP-C protein could be purified from a 1 litre culture.

Functional studies were carried out using a Langmuir trough and results indicated the recombinant SP-C gave superior surface activity. This was seen in the form of data showing a greater expansion in isotherms produced in a Langmuir trough, when compared to native SP-C.

More recently, a fusion protein based on the N-terminal domain from Silk spiders has been used to produce SP-C in high quantities (Kronqvist *et al.* 2017).

An N-terminal spidroin plasmid expressed in *E.coli* was mutated to create a charge reversed mutant in order to make SP-C more pH insensitive, stable, and hypersoluble.

The authors of the above used a method of extracting SP-C that is devoid of detergents and opted to precipitate SP-C using NaCl followed by ethanol extraction and cleavage using proteases.

Their expression levels, before the precipitation was carried out, appeared to be high. However, 50% of the preparation was then lost after precipitation with NaCl and yet more after ethanol extraction and protease cleavage.

The target protein was however pure and displayed a single band at the correct molecular weight on a Coomassie stained gel at 4 kDa on the gel.

Some 93 mg/L of purified SP-C protein was produced in the Kronqvist study (Kronqvist *et al.* 2017). The study also includes *in vivo* animal studies that show the SP-C produced improves lung compliance in pulmonary deficient rabbits.

A comparison of the methods used to isolate and purify SP-C is shown below in Table 1.6.

**Table 1.6 Comparison of the Methods used to Synthesis and Extract Recombinant SP-C**

	<b>(Lukovic <i>et al.</i> 2006)</b>	<b>(Kronqvist <i>et al.</i> 2017)</b>
Host	<i>E.coli</i>	<i>E.coli</i>
Vector	T7 SN/gpa A	pT7HisTrxHisNT
Cloning	Quikchange	Site directed mutagenesis followed by cloning
Transformation	BL21 <i>E.coli</i>	BL21 <i>E.coli</i>
Solubilisation	Detergent followed by organic extraction and lipophilic size exclusion	NaCl washes and CnBr cleavage
Purification	Nickle NTA	NaCl followed by protease cleavage and ethanol extraction
Yield	4mg/L	20-30mg
Secondary structure analysis	CD	CD and NMR
Functional activity	Isotherms generated on produced greater expansions than native SP-C on Langmuir trough.	Animal studies showing SP-C restored lung compliance
Limitations	No palmitoylation, but this was advantageous	None

### 1.3.1 DISCUSSION - *Pichia Pastoris* as a Recombinant Host Cell

The high hydrophobicity of SP-B and SP-C has made their synthesis as recombinant proteins particularly challenging.

***Literature to date shows it has not been possible to synthesise SP-B as a fusion protein, however SP-C has been synthesised.***

An alternative approach as adopted, in this work would offer value for reasonably yielding synthesis of both proteins.

Yeast, in particular *P. pastoris*, has become an increasingly popular host for more challenging recombinant proteins as compared to more commonly-used host cells such as mammalian cell lines, insect lines and *E.coli* (Bill 2014).

Other studies of recombinant protein expression have been carried out in human cell lines such as human embryonic kidney cells (HEKs); Chinese hamster ovary cells (CHOs) (Altamirano 2013); and Insect cells (Kuhn 2003).

Overall, *P. pastoris* has some advantages over these other expression systems, namely; it is able to grow to high cell densities, is similar to microbes, and some very high yields of protein have been achieved.

The increase in the number of structural studies that have used *P. pastoris* as an expression system further supports this (Ahmad *et al.* 2014, Byrne 2015).

The growth media of *P. pastoris* is cheaper than that of insect cells and mammalian cells.

*P. pastoris* is capable of many of the post translational modifications of higher eukaryotes, including disulphide bond formation and the proper protein folding (Macauley-Patrick 2005, Li *et al.* 2007) which may not be possible in prokaryotic host cells such as *E. coli*.

*P. pastoris* is able to use methanol as the sole carbon source and is therefore referred to as a methylotrophic yeast.

Importantly, it is also free of cross contamination from mammalian sources etc.

The pPICZ vectors contain a strong inducible *AOX1* promoter, which controls the *AOX1* and *AOX2* genes; although there is evidence that the promoter is leaky under some conditions (Bawa *et al.* 2014).

*P. pastoris* cells containing recombinant protein vectors are grown on glycerol-containing media which repress the *AOX1* promoter and allow a high biomass to be generated. Adding methanol to the media then causes the *AOX1* promoter to be induced.

Small scale production screens for *P. pastoris* use baffled shake flasks for both colony screening and as the vessel for protein expression.

However, these conditions are unlikely to be optimal for expressing maximal yields of protein.

*P. pastoris* is an aerobic organism, and the oxygen transfer into the growth medium, in shake flasks over the duration of a run, is limited.

Bioreactors can be used for larger scale protein production, as they allow precise control over growth conditions such as pH, temperature and dissolved oxygen.

These parameters are difficult to accurately maintain in shake flasks; and the ability to monitor and tightly control parameters in bioreactors should lead to improved cell growth and yields of protein.

Methanol induction is also difficult to control in shake flasks, importantly, methanol becomes toxic to the cells above 5 g/L (Macauley-Patrick *et al* 2005).

In bioreactors, a more controlled feed can be used rather than simply adding methanol to the media, whilst the feed rate can be adjusted according to the growth of the cells. In addition, the dissolved oxygen can be maintained in bioreactors, at a desired level, by stirring and the introduction of air and/or oxygen-enriched gasses.

Improvements to yields can be achieved by altering the temperature; e.g. lowering the temperature from 30 °C to 20 °C improved the yield of an antibody fragment (Dragosits *et al.* 2011) and several G protein-coupled receptors (André *et al.* 2006).

Optimising the pH can also improve protein stability; a pH of 6 improved the stability of human serum albumin (Kobayashi 2000), while lowering the pH to 3 improved the production of insulin-like growth factor I (Brierley, Davis and Holtz 1994).

Unsurprisingly, the optimal conditions appear to be dependent on the protein being expressed.

### 1.3.2 Phospholipid Composition of *Pichia Pastoris*

The phospholipid composition of *P. pastoris* is made up of: PE, PG, PS, PI, Lysophospholipid (LS), and cardiolipin (full names can be found in the abbreviations, earlier on page 8).

The relative abundances have been identified using mass spectrometry

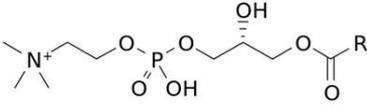
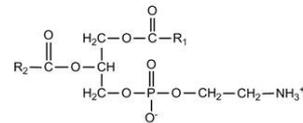
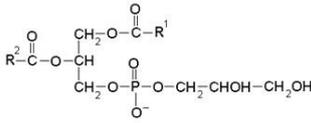
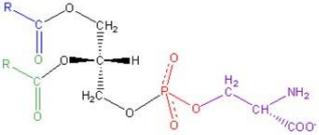
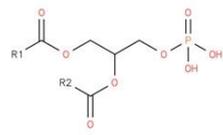
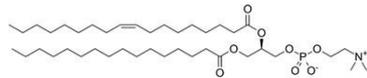
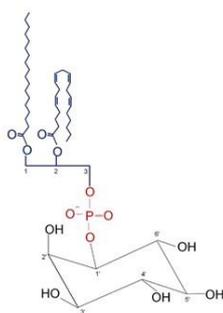
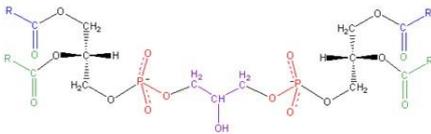
Chapter 5 outlines mass spectrometry analysis of lipids extracted using SMA2000P polymers from SP-B and SP-C (discussed later in 1.4.1).

The mass spectrometry analyses helped to examine preferential protein-lipid interactions in *P. pastoris*, as well as potentially giving an insight into the protein-lipid interactions in the tear film.

The detailed phospholipid composition, including structures, of *P. pastoris* can be found below in Table 1.7, based on compositional findings from (Grillitsch *et al.* 2014).

**Table 1.7 Phospholipid Abundances of *Pichia Pastoris* Identified Using Mass Spectrometry (Grillitsch *et al.* 2014)**

Note: structures of phospholipids are also shown; images taken from Wikipedia.

	Abundance in <i>P.pastoris</i> plasma membrane (% of total phospholipids Grillitsch <i>et al.</i> 2014)
	Lysophospholipid 1.86
	Phosphatidylethanolamine 36.15
	Phosphatidylglycerol -
	Phosphatidylserine 20.42
	Phosphatidic acid 8.23
	Phosphatidylcholine 27.48
	Phosphatidylinositol 4.29
	Cardiolipin 1.06

### 1.4.1 Solubilisation of Proteins from Membranes

Membrane proteins must be extracted from the lipid bilayer in order to isolate and purify them for biophysical investigations, including such as X-ray crystallography, CD and NMR (Seddon, Curnow and Booth 2004).

Detergents with amphiphilic properties are often used for membrane protein solubilisation; the hydrophobic moieties stabilise the surface of the protein, and hydrophilic moieties enable the membrane proteins to remain water-soluble once extracted (Le Maire, Champeil & Moller 2000).

However, solubilisation of membrane proteins using detergents is not ideal as the membrane is composed of many different types of lipid, and detergents cannot fully replicate this complex environment (Oluwole *et al.* 2017).

Detergents may also disrupt intermolecular interactions, which can alter the structure of the protein itself (Booth 2004).

Whilst detergents have a surface tension impact, this was recognised and every effort was made to eliminate residual detergent from the final products tested.

More recently, polymers have been investigated for their ability to solubilise membrane proteins to alleviate some of these problems and indeed they form a significant aspect of this work's investigations.

### 1.4.2 SMA 2000P Solubilisation

Styrene-maleic acid (SMA) co-polymers is an alternative to detergent solubilisation of membrane proteins.

Maleic anhydride and styrene monomers undergo radical polymerization to produce styrene maleic anhydride (Alfrey and Lavin 1945). The styrene-maleic anhydride co-polymer is then hydrolysed to SMA, see below in Figure 1.17.



**Figure 1.17 Structures of Styrene Maleic Anhydride and Styrene Maleic Acid**

The styrene maleic anhydride co-polymer is hydrolysed to form styrene maleic acid (SMA). Image taken from (Lee *et al.* 2016).

The amphipathic nature of the polymer makes them ideal solubilising agents (Tonge and Tighe 2001, Knowles *et al.* 2009).

The responsive hydrophobically-associating polymers allow proteins to be extracted from membranes without causing disruption to their structure.

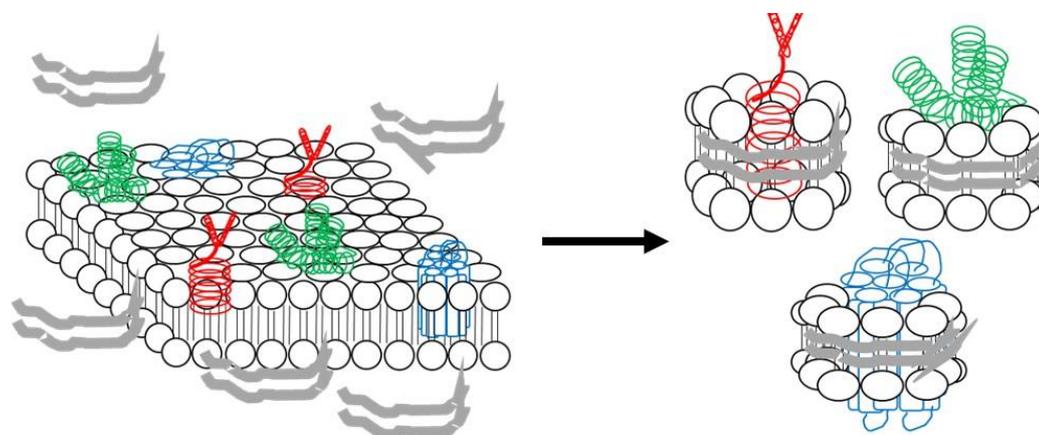
This is achieved by maintaining the native lipid environment, through the formation of disk like nanoparticles.

Solubilisation of membrane proteins is considered to occur in 3 steps:

- 1) Membrane binding
- 2) Insertion of the SMA polymer into the hydrophobic core of the membrane
- 3) Nanodisc formation (Scheidelaar *et al.* 2015).

SMA2000P polymers have been used to successfully solubilise proteins from yeast (Jamshad *et al.* 2015), insect cells (Gulati *et al.* 2014) and human cells (Gulati *et al.* 2014).

Figure 1.18 below illustrates a schematic representation of a SMALP protein.



**Figure 1.18 Protein Solubilisation using SMA2000P**

During solubilisation, the SMA polymer inserts into the membrane and surrounds lipids and proteins forming a 'doughnut' like arrangement.

Encapsulation of the polymer and the phospholipid mimic cause the formation of a discoidal nanostructure 10-12 nm in size. The SMA polymer is represented by the grey structure.

Some advantages of using polymers include; retention of annular lipids which have been shown to be important for structural integrity; increased stability; and there being no need to supplement buffers with the polymer (Pollock *et al.* 2017).

### 1.4.3 DISCUSSION - Detergent Solubilisation

To study the behaviour of membrane proteins, they first need to be extracted from the membrane to a similar environment, importantly ensuring their structure is retained throughout.

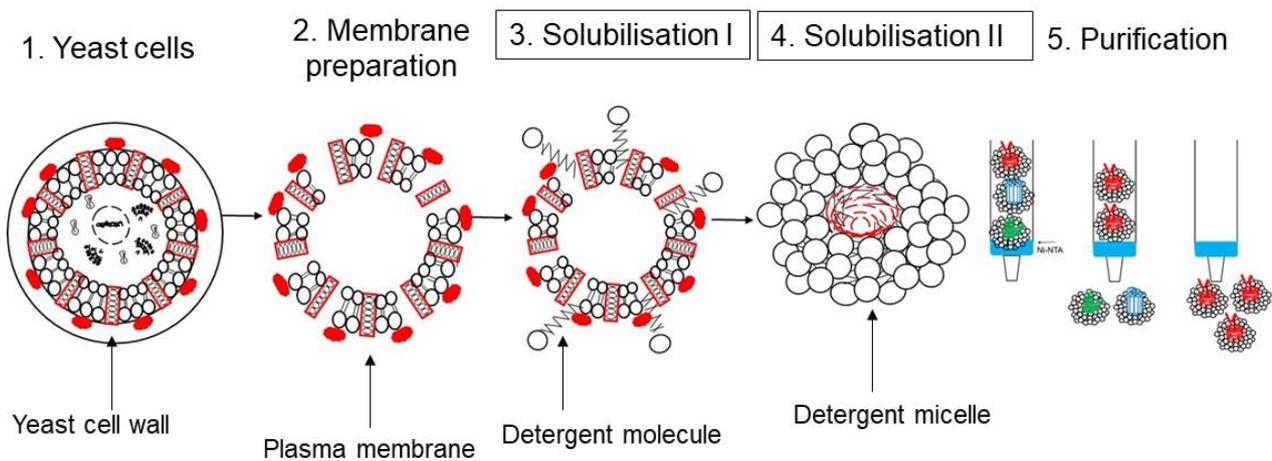
Addition of detergent to the plasma membranes containing the protein of interest then causes lipids around the protein to be replaced by detergent molecules. Excess detergent will encourage the formation of micelles with the protein held in the middle.

The critical micelle concentration (CMC) is the concentration of detergent that will form micelles for each detergent (le Maire, Champeil and Moller 2000), with typically concentrations 10 times over the CMC being recommended.

Detergent solubilisation can be affected by agitation, temperature, length of solubilising time, detergent type and its concentration.

Both detergent molecules and phospholipids are amphipathic; that is they are composed of a hydrophilic head group and a hydrophobic tail.

The formation of micelles mimics the hydrophobicity that membrane proteins require, because the protein is surrounded by hydrophobic detergent tails, see below in Figure 1.19.



**Figure 1.19 Detergent Solubilisation**

Expression in *Pichia pastoris* cells caused SP-B and SP-C proteins to be produced alongside other yeast proteins (1). For illustration purposes, the proteins are shown to be part of the plasma membrane in red. Plasma membrane fractions were isolated (2), and it is to this preparation that the buffers and detergents were added.

Addition of detergent to the plasma membrane causes lipids around the protein to be replaced by detergent molecules and excess detergent will encourage the formation of micelles with the protein held in the middle (4).

Solubilised proteins were then purified to isolate SP-B/SP-C (5).

Due to their amphipathic nature, detergent molecules can disrupt the lipid-lipid and lipid-protein interactions (Seddon *et al.* 2004).

Detergents themselves, can be divided into 3 main groups: non-ionic, ionic and zwitterionic.

Ionic detergents e.g. SDS can have cationic or anionic head groups, and this makes them potentially ideal solubilising agents.

However they can also be denaturing.

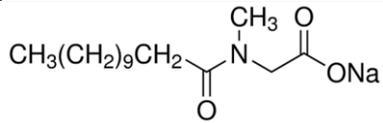
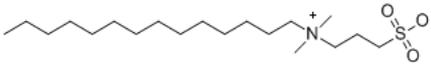
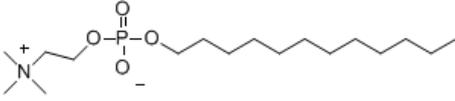
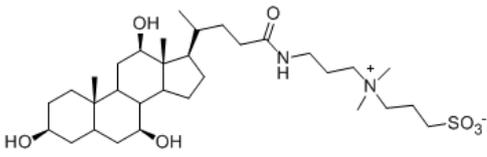
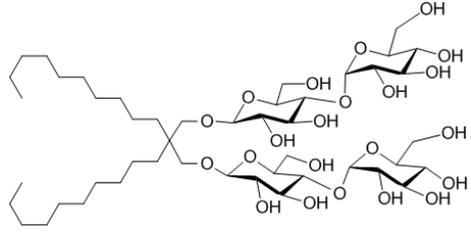
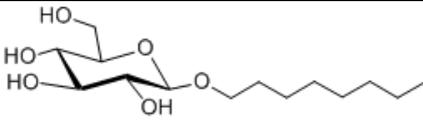
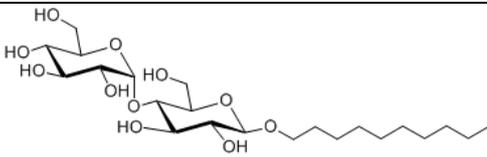
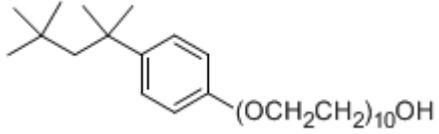
Non-ionic detergents e.g. beta-octyl glucopyranoside have a glycosidic or polyoxyethylene head group which helps disrupt protein-lipid and lipid-lipid interactions. These detergents are mild and non-denaturing and also make good solubilising agents.

Zwitterionic detergents e.g. Triton-X-100 have both positive and negative charges. These can also disrupt protein-protein interactions, making them more denaturing.

Table 1.8 below shows the different detergent groups and their structures.

**Table 1.8 Examples of Detergents used in Purification**

Structures taken from Sigma Aldridge and Anatrace website.

Detergent	Group	Use in purification	Structure
N-lauroylsarcosine sodium salt (NLSS)	Anionic		$\text{CH}_3(\text{CH}_2)_9\text{CH}_2\text{C}(=\text{O})\text{N}(\text{CH}_3)\text{CH}_2\text{C}(=\text{O})\text{ONa}$ 
Anzergent (Anz)	Zwitterionic		
Fos-choline 12 (Fos-12)	Zwitterionic		
3-((3-cholamidopropyl)dimethylammonio)-1-propanesulfonate (CHAPS)	Zwitterionic	1.0%	
Maltose neopentyl glycol-3 (MNG)	Neopentyl glycol class		
Octyl-beta-glucoside (β-OG)	Non ionic	11%	
Decyl maltoside (DM)	Non ionic	13.9%	
Triton X-100	Non ionic	1.9%	

## 1.5 RECAP - Aims and Objectives of the Work of this Thesis

In essence, the aim of this thesis was:

- 1) To synthesise and characterise SP-B and SP-C as recombinant proteins within *P .pastoris*.
- 2) Explore the role of SP-B and SP-C in tear film lipid spreading.

### Thesis Outputs re: Objectives

Synthesis of SP-B and SP-C

Extraction using detergent vs detergent free systems

Purification of both proteins to a high standard

Characterisation of SP-B and SP-C

Functional analysis of SP-B and SP-C using a Langmuir trough

Investigation of protein-lipid interactions of SP-B and SP-C using mass spectrometry analysis

Examination of the mechanism of SP-B and SP-C in the tear film using computational modelling

Each component seeks, both separately and in combination, to advance the wider understanding across these areas.

## Chapter 2

### Materials and Methods

*This chapter details all the materials used, but of especial importance is that all experimental procedures carried out in this work are explained in detail within this section.*

#### 2.1 Materials

##### 2.1.1 Buffers

2.1.1.1 50 X TAE buffer (Thermofisher B49)

2.1.1.2 1% Agarose

1 g agarose dissolved in TAE buffer, total volume 100ml.

2.1.1.3 Yeast Extract Peptone Dextrose Sorbitol Plates (YPDS)

1% (w/v) Yeast extract

2% (w/v) Peptone

2% (w/v) Dextrose

2% (w/v) Agar

100 mM (w/v) Sorbitol

100 µg/ml Zeocin

2.1.1.4 Yeast Extract Peptone Dextrose Medium (YPD)

1% (w/v) Yeast extract

2% (w/v) Peptone

2% (w/v) Dextrose

2% (w/v) Agar

100 µg/ml Zeocin

#### 2.1.1.5 Buffered Glycerol-complex Medium (BMGY) and Buffered Methanol-complex Medium (BMMY)

1 (w/v) % Yeast extract  
2 (w/v) % Peptone  
100 mM Potassium phosphate  
1.34% (w/v) Yeast nitrogen base  
 $4 \times 10^{-5}$  % (w/v) Biotin  
1% (w/v) Glycerol (for BMGY) or 0.5% methanol (for BMMY)  
100 µg/ml Zeocin

#### 2.1.1.6 Minimal Media

2.67 % (v/v) Phosphoric acid (85%)  
0.093 % (w/v) Calcium sulfate  
1.82 % (w/v) Potassium sulfate  
1.49 % (w/v) Magnesium sulfate  
0.41 % (v/v) Potassium hydroxide  
0.4 % (w/v) Glycerol

#### 2.1.1.7 Breaking Buffer (pH 8.0)

50 mM Sodium phosphate dibasic heptahydrate ( $\text{Na}_2\text{HPO}_4$ )  
2 mM Ethylenediaminetetraacetic acid (EDTA)  
100 mM Sodium chloride (NaCl)  
0.5 % (w/v) Glycerol

#### 2.1.1.8 Buffer A (pH 8.0)

20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic  
acid (HEPES)  
50 mM NaCl  
1 % (w/v) Glycerol

#### 2.1.1.9 Protein buffer (pH 8.0)

50 mM HEPES

150 mM NaCl

0.5 % (w/v) Glycerol

#### 2.1.1.10 Laemmli buffer (5X)

60 mM Tris HCl pH 6.8

2% (w/v) SDS

10% (w/v) glycerol

0.01% (w/v) bromophenol blue

5% (w/v)  $\beta$ -mercaptoethanol

#### 2.1.1.11 Binding/Wash buffer pH 8.0

50 mM Tris HCl

500 mM NaCl

20 mM Imidazole

#### 2.1.1.12 Elution buffer pH 8.0

50 mM Tris HCl

500 mM NaCl

50-1000 mM Imidazole

#### 2.1.1.13 Phosphate Buffered Saline (PBS)

5 tablets dissolved in 1L (1282-1680 Fisher Scientific)

2 ml Tween added for PBS-tween (p1379 Sigma Aldrich)

#### 2.1.1.14 Running buffer

900 ml distilled water and 100 ml Tris/glycine/SDS  
(B9-0032 Geneflow)

#### 2.1.1.15 1 x TE Buffer (pH 8.0)

10 mM Tris HCl pH 8.0  
1 mM EDTA

#### 2.1.1.16 Transfer buffer

700 ml Distilled water, 200 ml methanol (M/4056/17  
Fisher Scientific) and 100 ml Tris/glycine  
(B9-0056 Geneflow)

#### 2.1.1.17 Tris HCl (pH 8.0)

50 mM Tris base  
Hydrochloric acid (added to achieve pH 8.0)

### 2.1.2 **Reagents**

2.1.2.1 1 Kb O'GeneRuler DNA ladder (ThermoFisher SM1163)

2.1.2.2 Loading dye (ThermoFisher SM1163)

### 2.1.3 *Restriction Enzymes*

**Table 2.1      Restriction Enzymes used for Cloning**

Enzyme	Source	Consensus sequence	Catalogue number
<i>EcoR1</i>	<i>E. coli</i> RY13	GAATTC	NEB R3101
<i>XbaI</i>	<i>Xanthomonas badrii</i>	TCTAGA	NEB R0145S
<i>PmeI</i>	<i>Pseudomonas mendocina</i>	GTTTAAAC	NEB R0560S
T4 DNA ligase	<i>E. coli</i> C600 pcl857 pPLc28 lig8		NEB M0202S

### 2.1.4 **Competent Cells**

2.1.4.1 XL10 Gold (Agilent 200314)

### 2.1.5 ***Kits and Equipment***

2.1.5.1 PCR kit (Thermofisher F553S)

2.1.5.2 Ligation kit (Thermofisher K1422)

2.1.5.3 Mini prep kit (Thermofisher K0503)

2.1.5.4 Maxi prep kit (Qiagen 12162)

### 2.1.6 **Strains**

2.1.6.1 *E.coli* XL10 Gold competent cells (Agilent)

2.1.6.2 *Pichia pastoris* strain X-33 (Invitrogen)

## 2.1.7 Primers and Vectors

### 2.1.7.1 Vectors

**Table 2.2 Commercial Vectors used in Cloning**

<b>Gene name</b>	<b>Vectors</b>	<b>cDNA size</b>
<i>SFTPB</i>	PUC18	1200bp
<i>SFTPC</i>	PUC18	600bp

**Table 2.3 Recombinant Protein Vectors**

<b>Recombinant vectors</b>	<b>Protein</b>	<b>Expected size of translated protein</b>
pPICZ $\alpha$ A-Pro <i>SFTPB</i>	ProSP-B	42kDa
pPICZ $\alpha$ A-Pro <i>SFTPC</i>	ProSP-C	24kDa
pPICZ $\alpha$ A-Pro <i>SFTPC10</i>	ProSP-C10	24.5kDa
pPICZ $\alpha$ A- <i>SFTPB</i>	SP-B	8kDa
pPICZ $\alpha$ A- <i>SFTPC</i>	SP-C	6kDa

### 2.1.7.2 *pPICZ $\alpha$ Vector Map*

The pPICZ $\alpha$ A vector was used to express recombinant proSP-B, SP-B, proSP-C, proSP-C10 and SP-C.

The main features of the vector include an AOX1 promoter,  $\alpha$ -factor secretion signal, zeocin resistance and a His<sub>6</sub>-tag.

The constituents of the vector are labelled below in Figure 2.1 and a description of each of the role each is listed in Table 2.4.



**Figure 2.1** Invitrogen *Pichia Pastoris* pPICZ $\alpha$ A,B,C Expression Vector

The expression vector includes an  $\alpha$ -factor secretion signal, restriction enzymes and a C-terminus c-myc and His<sub>6</sub> tag, as well as a Zeocin resistance gene.

Image taken from Invitrogen's *Pichia* pPICZ $\alpha$ A,B,C expression manual.

**Table 2.4 Important Features Within the pPICZ $\alpha$ A, B and C Vector and their Role**

<b>Feature in pPICZ<math>\alpha</math>A vector</b>	<b>Description</b>
Multiple Cloning Site (MCS)	ProSP-B, SP-B, ProSP-C, ProSP-C10 and SP-C were inserted in the MCS using <i>EcoR1</i> and <i>Xba1</i>
5' AOX1 promoter	AOX1 promoter induced by methanol. Also involved in targeting plasmid integration to AOX1 locus during transformation
Alpha secretion signal factor	Aids secretion of recombinant proteins
c-myc epitope	Allows detection of proteins using an Anti c-myc antibody
His <sub>(6)</sub> tag	Allows for detection of protein using Anti-His antibody
AOX1 TT	AOX1 transcription termination (TT) region. Polyadenylation signal
TEF1 promoter	Transcription elongation factor 1 gene promoter involved in driving expression of the Zeocin resistance gene in <i>P. pastoris</i>
EM7	Prokaryotic promoter that drives constitutive expression of the Zeocin resistance gene in <i>E. coli</i>
( <i>Sh ble</i> ) Zeocin resistance gene	Helps to identify transformants after transformation in <i>E. coli</i> and <i>P. pastoris</i>
CYC1 TT	Transcription termination region
pUC	Allows replication and maintenance of the plasmid in <i>E. coli</i>

### 2.1.7.3 *Primer Sequences*

**ProSP-B forward primer:** TAACGTGAATTCATGGCTGAGTCACACCTGCTG

**ProSP-B reverse primer:** ACGTTATCTAGAGCAAGGTCGGGGCTGTGGATAC

**ProSP-C forward primer:** TAACTTGAATTCATGGATGTGGGCAGC

**ProSP-C reverse primer:** ACGTTATCTAGAATGATGTAGTAGAGCGGC

**ProSP-C10 forward primer:**  
ATTAACAAGAATTCGCCACCATGGATGTGGGCAGCAAAGAGG

**ProSP-C10 reverse primer 1:**  
GTGATGGTGATGGTGATGGTGATGGTGATGGATGTAGTAGAG

**ProSP-C10 reverse primer 2:** AACTGATTTCTAGAGCGTGATGGTGATGGTGATG

**SP-B forward primer:** TAAATTGAATTCTTCCCCATTCCCTCTCCCCTATT

**SP-B reverse primer:** ATATTATCTAGATGGCAGACCAGCTGGGGCAG

**SP-C forward primer:** TAATATGAATTCTTTGGCATTCCCTGCTGCCCAG

**SP-C reverse primer:** TAATATTCTAGATTGAGACCCATGAGCAGGGCTCCC

**AOX1 forward primer:** 5' GACTGGTTCCAATTGACAAGC 3'

**AOX1 reverse primer:** 5' GCAAATGGCATTCTGACATCC 3'

## 2.2 Methods

### 2.2.1 Molecular Biology

#### 2.2.1.1 *Vector Design*

*SFTPB* and *SFTPC* encode pulmonary surfactant proteins SP-B and SP-C.

Full length versions of the proteins are cleaved to mature active peptides *in vivo*.

Constructs for both full length genes and mature peptides were made for each of these.

The aim of initial experiments was to integrate cDNA for these proteins into a yeast vector.

pPICZ $\alpha$ A was chosen as the expression system because it has a zeocin marker which allows for selection of vectors that have taken up the *SFTPB* and *SFTPC* insert.

A multiple cloning site contained in the sequence, allows for different combinations of restriction enzymes to be used.

Using molecular cloning software, 'Gentle' ([www.gentle.magnusmanske.de](http://www.gentle.magnusmanske.de)) an *in silico* cloning was used to check that the His<sub>6</sub> tag remained in frame with *EcoR1* and *Xba1*.

#### 2.2.1.2 *Primer Design*

Non-coding DNA and restriction enzyme consensus sequences were added to the 5' and 3' ends of both full length and mature sequences.

Adding non-coding DNA allows restriction digestions to be performed more effectively.

Premier Biosoft ([www.premierbiosoft.com](http://www.premierbiosoft.com)) online software was used to check melting temperatures, GC content and likelihood of formation of secondary structures.

Primer sequences for *SFTPB* and *SFTPC* full length and mature can be found in section 2.1.7.3.

#### 2.2.1.3 *Transformation of Surfactant Genes in PUC19 Vector*

480 ng of *SFTPB* (ATCC 65987) and 540 ng *SFTPC* (ATCC 65985) were added to microcentrifuge tubes containing 20  $\mu$ l of competent *Escherichia coli* (*E.coli*) cells (XL10-Gold) and left on ice for 20 minutes.

The tubes were rotated every 2 minutes for 10 seconds and then left for a further 10 minutes on ice. Both tubes were placed in a heat block set at 42°C for 45 seconds and placed back on ice.

1 ml lysogeny broth (LB) was added to the tubes containing SP-B and SP-C and left to incubate for 60 minutes at 37°C.

30 µl of the transformation solution was then spread on LB-ampicillin (50 µg/ml) plates and incubated at 37°C overnight.

*There was a single Transformation of this vector carried out, see Figures 3.2 and 4.2.*

#### 2.2.1.4 *MiniPrep*

Minipreps were carried out using the GeneJet plasmid miniprep kit.

Selected colonies expressing *SFTPB* and *SFTPC* were grown on LB-ampicillin plates.

(25 µg/ml) were inoculated into LB-ampicillin (25 µg/ml) and left overnight at 37°C.

The *E.coli* cells were then centrifuged at 5000 *g* for 5 minutes and the supernatant discarded.

The pellet was resuspended in resuspension buffer and transferred to a microcentrifuge tube.

Lysis solution was added to the microcentrifuge tube and mixed by inverting until the solution became clear.

Next, neutralisation solution was added, the microcentrifuge tube was inverted 4-6 times and centrifuged at 13,000 *g* for 5 minutes. The supernatant was removed using a pipette, transferred to the GeneJet spin column and centrifuged at 13,000 *g* for 1 minute.

The flow through was discarded, wash solution added and the column centrifuged at the same speed as the last step. The eluted DNA was quantified using a Nanodrop spectrophotometer and stored at -20°C.

*Minipreps were routinely carried out (after every cloning experiment) to isolate DNA for subsequent diagnostic digests.*

#### 2.2.1.5 *Nanodrop*

DNA concentrations were determined using a Nanodrop 100 spectrophotometer and analysed with ND2000 software.

### 2.2.1.6 Polymerase Chain Reaction (PCR)

The following components listed below in Table 2.5 were added to a PCR reaction mix and kept on ice throughout. The cycling times are listed below in Table 2.6.

**Table 2.5 Volumes and Final Concentrations of Reagents Added to PCR Reactions**

PCR reagent	Reaction Volume (50 µl)	Final Concentration
ddH <sub>2</sub> O	32.5 µl	-
5x Phusion high fidelity buffer	10 µl	1x
10mM dNTPs	1 µl	200 µM each
Primer 1 (SP-B F, SP-C F)	2.5 µl	0.5 µM
Primer 2 (SP-B R, SP-C R)	2.5 µl	0.5 µM
DNA (SP-B, SP-C)	1 µl	10-15 ng
Phusion DNA polymerase	0.5 µl	0.02 U/µl

The following cycles were programmed into a PCR machine according to manufacturer's (Thermofisher) instructions.

The PCR products were run on an agarose gel (1%), to confirm insert size.

**Table 2.6 PCR Cycling Times**

Cycling Instructions	Temperature	Time
Initial denaturation	98°C	2 minutes
Denaturation	98°C	30 seconds
Annealing	55°C	30 seconds } 30 cycles
Extension	72°C	1 minute
Final Extension	72°C	5 minutes

*PCR was carried out multiple times for each construct, frequency determined by successful confirmation of cDNA insertion.*

### 2.2.1.7 Restriction Digest

PCR products and vector DNA (pPICZαA) were digested using restriction enzymes.

The following components, see below in Table 2.7, were added to a microcentrifuge tube and left at 37°C for 2 hours .

pPICZαA was streaked from a glycerol stock onto LB agar plates with zeocin (25 µg/ml).

A colony was inoculated in 200 µl LB zeocin (25 µg/ml) and left to grow overnight at 37°C.

Vector DNA was extracted using a Genejet miniprep kit (Thermofisher).

**Table 2.7 Restriction Enzyme Reaction Composition**

	Representative volumes
ddH <sub>2</sub> O	28 µl
DNA	1 µg
Cutsmart buffer	10 µl
<i>EcoR</i> 1	1 µl
<i>Xba</i> 1	1 µl
Reaction Volume	50 µl

The digests were run through an agarose gel (1%), and purified using a Genejet gel extraction kit (Sambrook, Fritsch and Maniatis 1989).

#### 2.2.1.8 *Ligation*

T4 DNA ligase was used and the following components added, see below in Table 2.8. An online *in silico* calculator was used to calculate the molar vector to insert ratio of 1:3. This calculated the optimal amount of vector and insert to add in each reaction.

**Table 2.8 Enzymes and Reagents added to Ligation Reactions**

Ligation components	
Vector (3541bp)	50 ng
Insert	26 ng
T4 Ligase buffer	2 $\mu$ l
ddH <sub>2</sub> O	9 $\mu$ l
Ligase	1 $\mu$ l
Reaction volume	Make up to 20 $\mu$ l

Ligations were left for 1 hour at room temperature then inactivated by heating to 65°C.

The ligations were transformed into XL10 gold competent cells.

*Ligations were repeated with different ratios until successful cloning could be confirmed, refer to Figures 3.2 and 4.2.*

#### 2.2.1.9 *Transformation into XL10 Gold*

30  $\mu$ l competent cells were added to pre-chilled tubes.

1  $\mu$ l of  $\beta$ -ME was then added to each of the tubes and left to incubate on ice for 10 minutes, the tubes being swirled gently every two minutes.

5  $\mu$ l of the ligation mix was added to each tube and 50 ng of undigested vector DNA added to a separate control tube.

The tubes were gently rotated and then incubated on ice for 20 minutes.

The transformation mix was then heat pulsed for 45 seconds at 42°C.

The tubes were left on ice for 5 minutes and 0.9 ml of pre-warmed LB was added to each tube and incubated at 37°C for 1 hour in the shaking incubator.

200 µl of the transformation solution and 5 µl of the control vector containing solution were spread on LB zeocin plates (25 µg/ml).

The plates were incubated at 37°C overnight (Sambrook *et al.* 1989).

*Transformations were repeated as necessary for each cloning experiment, until successful insertion could be confirmed, as described above.*

#### 2.2.1.10 *Preparation of Competent X-33 Cells*

*P. pastoris* strain X-33 was streaked onto a YPD plate from a glycerol stock and incubated at 30°C for 4 days.

A single colony was taken and used to inoculate 5ml of YPD in a falcon tube and left shaking at 220rpm in a 30°C incubator overnight.

The optical density ( $A_{600}$ ) was checked the next morning and the culture was resuspended in a volume of YPD that gave a starting  $A_{600}=0.25$ .

This was left to grow again at a temperature of 30°C until an  $A_{600}$  of 1 was reached.

The culture was centrifuged at 4,000 g for 5 minutes in a Beckman Coulter Allegra 25R centrifuge, the supernatant discarded and the pellet resuspended in YPD, 1M HEPES (H3375) and 1M dithiothreitol (DTT BP172-5).

The suspension was then incubated at 30°C and left shaking at 220 rpm for 15 minutes.

The flask was placed on ice and 400 ml of ice cold sterile water added; this was centrifuged at 4,000 g for 5 minutes and the pellets were resuspended in 1 M sorbitol and further centrifuged at 4,000 g for 5 minutes.

The supernatant was discarded and the pellets resuspended in 1 M sorbitol and stored at -80°C.

*The competent cells, whose production is outlined above, were made up in batches. These were stored for no longer than 6 months.*

## **2.2.2 Protein Expression**

### *2.2.2.1 Transformation of Pulmonary Surfactant Genes into Pichia Pastoris Strain X-33*

50 µl of competent cells were transferred to an electroporation curvette (EP202) with 5 µg linearized DNA and left on ice for 20 minutes.

The electroporation curvette containing the transformation mixture was electroporated at 1,800 V and 1 M sorbitol immediately added.

The suspension was transferred to a microcentrifuge tube and incubated at 30°C for 2 hours.

200 µl of the transformed *P. pastoris* was spread on YPDS plates containing 100, 250 or 500µg/ml zeocin and incubated at 30°C for three days (Higgins and Cregg 1998).

*These Transformations were carried out singly, with aliquots spread on YPDS plates containing increasing concentrations of zeocin.*

### 2.2.2.2 Transformation Screen

8 colonies were picked from each plate and grown in BMGY at 30°C overnight.

The OD was measured and cells were resuspended to an  $A_{600}=1$  in BMMY.

The cultures were left shaking for 48 hours at 30°C.

The BMMY cell culture was aliquoted into equal volumes of 50 ml and centrifuged for 10 minutes at 5,530 g (Allegra 25R); the supernatant was discarded and the pellet re-suspended in 500 µl breaking buffer.

1 ml of the suspended solution was transferred to a 1.5 ml breaking tube containing 1 ml glass beads. In addition to this, 1 µl protease cocktail inhibitor (539136 Calbiochem) was added and the tubes placed in a Qiagen Tissue Lyser LT at 50Hz for 10 minutes.

The bottom of each breaking tube was pierced with a hot needle and the tubes were fixed into 15 ml falcon tubes and centrifuged at 5,530 g for 3 minutes (Allegra 25R).

The supernatant was transferred to a 1.5 ml microcentrifuge tube and centrifuged again at 15,000 g for 15 minutes in a Fisher Scientific accuspin Micro R centrifuge.

Finally, the supernatant was transferred into ultracentrifuge tubes and centrifuged at 100,000 g for 60 minutes (Beckman Coulter Optimax TLX ultracentrifuge).

The supernatant was discarded and the pellet re-suspended in 100 µl buffer A and left overnight at 4°C.

This produced the membrane fraction, 25 µg of total protein from each colony was loaded into an SDS-PAGE and a Western blot was then carried out.

*These transformation screens were carried out singly to ascertain protein expression in small shake flasks.*

### 2.2.2.3 *Bioreactors*

A bioreactor (Applikon) was used for recombinant protein production. The media was continuously stirred by impellers, so helping to ensure the culture remained aerated throughout the run.

Yeast extract and peptone were added to 750 ml ddH<sub>2</sub>O, and placed in the bioreactor.

The pH probe was calibrated and inserted through its port into the main bioreactor vessel.

The dissolved oxygen (DO) probe was added after checking the sensor bulb was clean and the level of electrolyte was adequate.

Finally the feed lines and gas cylinder ports were clamped, sealed with foil and autoclaved at 121°C for 20 minutes with a subsequent slow cool cycle.

Once the vessel had cooled, the DO probe was connected to the ADI 1010 controller and left overnight.

A pre-culture of strain X-33 expressing pulmonary surfactant proteins was inoculated in BMGY and left at 30°C overnight.

The equipment set up the following day involved connecting various lines to feed bottles and controller systems.

The pH probe was connected to the ADI 1010 controller and the temperature probe was inserted into the thermowell.

Chiller lines were connected between the bioreactor and the chiller.

The compressor was drained and switched on; and the 60%: 40% ratio oxygen: nitrogen mixture in the cylinder was connected to the gas supply ADI 1026 unit.

The stirrer was placed on top of the bioreactor and feed lines connecting the acid and base were sprayed with ethanol and connected.

YNB, potassium phosphate buffer and glycerol were added and the left for 45 minutes in order to saturate the media with oxygen.

The entire volume was made up to a litre of BMGY in a bioreactor in which the dissolved oxygen was maintained at 30%, the pH at 6.0 and the temperature at 30°C.

The culture was left overnight to increase biomass and it was then fed with 50% glycerol at 14 ml/h for 4 hours from the next morning.

The culture was then starved for 1 hour to ensure the added glycerol was entirely consumed. It was then induced slowly with 20% methanol at 4 ml/h overnight.

The methanol was increased to 14 ml/h the next day and left for 50 hours after the run was started (Routledge and Clare 2012).

*These Bioreactors were run on average weekly to culture cells expressing full and mature constructs.*

#### 2.2.2.4 *Membrane Preparation (Large Scale)*

Culture from the bioreactor was decanted into 1 L centrifuge bottles and centrifuged at 5,000 *g* for 20 minutes.

The pellet was weighed and resuspended in an equal volume of breaking buffer.

EDTA free protease inhibitor tablets were added and left stirring for 20 minutes at 4°C.

The cell suspension was broken open on a C3 homogenizer (Avestin) at a pressure of 60 psi for 20 minutes and centrifuged at 5000 *g* for 20 minutes.

This was to remove heavier cell organelles before increasing to 100,000 *g* for 1 hour to separate the membrane fractions.

The pellets were taken, weighed and re-suspended to a membrane concentration of 80 mg/ml.

A BCA assay (Sigma Aldridge) was carried out to ascertain protein concentration.

#### 2.2.2.5 *BCA Assay*

To determine the concentration of unknown proteins, a standards calibration curve was constructed using 1 mg/ml BSA protein standards in a bicinchoninic acid assay.

For every experiment, 4.9 ml bicinchoninic acid (Sigma Aldridge) was mixed with 100 µl copper (II) sulphate (Sigma Aldridge).

96 well plates were used to assay the proteins and the final volume in each well was 200 µl, made up of 190 µl BCA reagent and 10 µl protein sample.

The plates were then incubated at 37°C for 17 minutes and read at 562 nm on a plate reader.

The generation of a standards curve enabled determination of unknown protein concentrations.

*Triplicate BCA assays were carried on all product protein samples.*

#### 2.2.2.6 *Polymer Preparation*

25 g SMA2000 (Cray Valley) was added to 250 ml NaOH (1 M) and left stirring overnight.

This polymer suspension was transferred to a round bottom flask, placed on a heating mantle with a condenser attached. The solution was refluxed at 130°C for 180 minutes, then left to cool.

The polymer solution was divided between 2 centrifuge tubes and 20 ml of concentrated HCl and 100 ml distilled water was added to each in a fume hood.

Both tubes were placed in a Beckman floor centrifuge and centrifuged at 100,000 g for 10 minutes.

The supernatant was removed and the pellet washed with 150 ml distilled water and centrifuged again; a stage repeated a further 5 times and continued until the pH was at 6.0.

The pellet was dissolved in 0.6 M NaOH and left stirring. Once dissolved, the pH was adjusted to pH 8.0 and the polymer freeze dried (Lee *et al.* 2016).

#### 2.2.2.7 *Polymer Solubilisation*

A final concentration of 2.5% (w/v) freeze dried SMA 2000P polymer powder was added to membranes at a final concentration of 35-40 mg/ml (wet pellet weight).

The solubilisation mix was vortexed and left on a shaker at room temperature for 1 hour.

Finally, it was centrifuged at 100,000 *g* for 20 minutes to separate the solubilised supernatant from the unsolubilised material left in the pellet (Lee *et al.* 2016).

#### 2.2.2.8 *Detergent Solubilisation*

20 mM Tris HCl pH 8, 20 mM NaCl, 20% glycerol, 5% detergent and 2 mg/ml (final protein concentration) membrane preparation were added to a microcentrifuge tube.

The solubilisation was thoroughly mixed by vortexing and left shaking for 2 hours at 22°C.

To harvest the solubilised protein, the mixture was centrifuged at 100,000 *g* for 1 hour and the supernatant collected.

*Small scale detergent solubilisations were each carried out 3 times.*

#### 2.2.2.9 Protein Purification

1 ml Ni-NTA resin was used per 50 ml of solubilised protein sample.

The resin was centrifuged at 5,000 *g* to remove the storage buffer which contained ethanol, and the Ni-NTA resuspended in binding buffer and centrifuged at 5,000 *g* again.

The resin was then added to solubilised protein supernatant and left rocking overnight at 4°C.

Purifications were carried out on a benchtop at 22°C.

The protein supernatant was added to a gravity flow column and any unbound protein collected.

The resin was washed with a Tris buffer containing a low concentration of imidazole.

The elution buffer was then added, 1 ml at a time, and collected in microcentrifuge tubes.

Elution samples were buffer exchanged into protein buffer and then concentrated using an Apollo protein concentrator with a molecular weight cut-off of 20 kDa for proproteins and 10 kDa for mature proteins.

*Full length and mature protein isolations (purifications) were always carried out prior to detergent and polymer solubilisations; their characterisations and their functional activity analyses.*

2.2.2.10 *Western Blot*

Proteins were identified from their molecular weight on an SDS PAGE gel and transferred to a nitrocellulose/PVDF membrane then bound with anti-His<sub>6</sub> antibody. This describes a Western blot.

The various components needed are described below in Table 2.9.

**Table 2.9 Reagents Required for 12% Separating Gel and  
4% Stacking Gel in SDS-PAGE**

The following components were added to a beaker:

<b>12% Separating gel</b>	<b>ml</b>	<b>1</b>
Polyacrylamide (30%)	ml	1.9
Water	ml	2.2
1.5M Tris HCl pH 8.0	ml	1.5
SDS (10%)	µl	60
Ammonium persulphate (20%)	µl	20
TEMED	µl	4.5

<b>4% Stacking gel</b>	<b>ml</b>	<b>1</b>
Polyacrylamide (30%)	ml	0.3
Water	ml	0.6
0.5M Tris HCl pH 8.0	ml	1.5
SDS (10%)	µl	25
Ammonium persulphate (20%)	µl	10
TEMED	µl	2.5

Ammonium persulphate (APS) (215589 Sigma-Aldrich) was freshly made each time and tetramethylethylenediamine (TEMED T9281 Sigma Aldrich) was added last to delay the onset of polymerization.

Laemmli buffer was added to samples, which were subsequently loaded into the wells and run at a voltage of 120 V for 90 minutes.

SDS-PAGE gels were stained with instant blue (Expedeon).

Gels for Western blots were transferred to PVDF 0.2  $\mu$ M membrane (Merck) placed in a sandwich cassette with 4 pieces of filter paper either side. The cassette was placed in a holding tank filled with running buffer and an ice block placed inside. Finally the lid was placed on top and left to transfer at a voltage of 100 V for 50 minutes.

After 50 minutes the membrane was removed and placed in 5% blocking solution, made of milk powder and PBS and stored overnight at 4 °C.

Anti-His<sub>6</sub> monoclonal antibody (Clontech) was added to 5% blocking solution at a dilution of 1/5000, added to the PVDF membrane, and left rocking at 40 rpm for 1 hour.

The membrane was washed in PBS Tween 3 times for 3 minutes.

Secondary antibody (anti mouse IgG HRP linked antibody) was added to 5% blocking solution at a dilution of 1/5000 and poured over the PVDF membrane which was left rocking at room temperature for 1 hour.

To a foil-covered tube 2ml of EZ-ECL (Biological Industries 20-500-500) solution A was added to 2 ml solution B and this mixed solution then poured into the top of the nitrocellulose membrane in a dark room.

The membrane was then able to be visualised using a Genebox.

*Western blots were carried out routinely throughout to identify expressed proteins.*

2.2.2.11 *Tris-Tricine Gel*

There were issues resolving SP-B and SP-C on SDS-PAGE gels.

To help improve the clarity of protein resolution, Tris-Tricine gels were used.

These gels were formed from the crosslinking of acrylamide and bisacrylamide to form smaller pores.

These improved the protein resolution.

**Table 2.10 Reagents Required to Make Tris-Tricine Separating and Stacking Gels**

The following components were added to a falcon tube

<b>12% Separating gel</b>	<b>ml</b>	<b>1</b>
3X Gel buffer	ml	3.3
Acrylamide/bisacrylamide solution	ml	2
ddH <sub>2</sub> O	ml	3.64
Glycerol (100%)	ml	1
Ammonium persulphate (10%)	μl	20
TEMED	μl	4.5

<b>4% Stacking gel</b>	<b>ml</b>	<b>1</b>
3X Gel buffer	ml	1.2
Acrylamide/bisacrylamide solution	ml	0.4
ddH <sub>2</sub> O	ml	3.34
Ammonium persulphate (10%)	μl	40
TEMED	μl	6

#### 2.2.2.12 *Reconstitution of SP-B and SP-C in Synthetic Lipids*

SP-B and SP-C are associated with lipids *in vivo*.

To improve relevance, detergent extracted proteins were purified and mixed with synthetic phospholipids. Adsorbent beads were used in order to remove any residual detergent from the extracted proteins.

40 µl of 10 mM POPC stock (Avanti 850457) was placed in a 1 ml microfuge tube and a 2:1 ratio chloroform and methanol was added.

The lipids were dried down to a film using a stream of nitrogen gas and then re-suspended in 95 µl 50 mM sodium phosphate buffer (pH 8).

The lipids were then passed through the extruder 21 times through 100 nm filters (Whatman) using Hamilton syringes, added to 350 µl purified protein and left rotating for 15 minutes at 21°C.

40 mg Biobeads (Biorad) were added to the protein-lipid suspension and left rotating at room temperature for 1 hour.

Biobeads were removed and exchanged for fresh Biobeads and left for another 1 hour before being repeated once more.

Reconstituted suspension was separated using a high speed spin at 100,000 *g*.

The pellet was re-suspended in 50 mM sodium phosphate buffer pH 8.0.

A BCA assay was carried out on both the supernatant and pelleted material.

*Despite initial experimental indications, ProSP-C10 reconstitution was attempted 3 times in total without sufficient success. Attempts to reconstitute SP-C were carried out twice but met with a similar conclusion.*

#### 2.2.2.13 *Sucrose gradients*

200-400  $\mu$ l of the pelleted material was placed at the bottom of the sucrose gradient in a centrifuge tube.

On top of this, 2.5ml of sodium phosphate buffer pH 8.0 was added and then 2.5ml of the following concentrations of sucrose: 60%, 20%, 10% and 5%. A final layer of buffer was added and these stepped gradients were left to spin at 26,000 *g* for 12-14 hours.

After the spin was completed, each layer was removed in 1.25 ml increments into microcentrifuge tubes and labelled with 0a, 0b, 5a, 5b, 10a, 10b, 20a, 20b, 30a and 30b.

Trichloroacetic acid (TCA 0.15%) and sodium deoxycholate DOC 72%) were added to each fraction to precipitate the protein and then vortexed for 10 seconds.

The fractions were then centrifuged at 13,000 *g* for 10 minutes. The supernatant was removed and the pellet re-suspended in 50  $\mu$ l TCA resuspension buffer.

The samples were then run on a SDS-PAGE/ Tris-Tricine gel and stained with Coomassie dye.

*Sucrose gradients were carried out, subsequent to reconstitution experiments, to attempt to ascertain if reconstitution had been successful. This was pursued after each attempted reconstitution - thrice for proSP-C10, and twice for SP-C.*

#### 2.2.2.14 *Determination of Protein Concentration*

0.1mg/ml BSA was made up in the following amounts: 0.125  $\mu$ g, 0.25  $\mu$ g, 0.5  $\mu$ g, 0.75  $\mu$ g, 1.0  $\mu$ g and 1.25  $\mu$ g and loaded into an SDS-PAGE/ Tris-Tricine gel.

Against the standards, 5-10  $\mu$ l of the protein being measured was also loaded.

The gel was stained with Coomassie dye and an image taken using a Genebox.

The intensity of each band was analysed by densitometry using the software Image J.

BSA samples were used to make a standards calibration curve. This was then used to measure the concentration of the purified SP-B/SP-C proteins. Method based on protocol in (Rothnie *et al.* 2004).

*Protein concentration assays were carried out on every occasion to ascertain both protein concentration and yield for characterisation stages such as CD and end purification yields.*

#### 2.2.2.15      *Densitometry*

SDS-PAGE gels and Western blots were analysed using Image J software.

A box was drawn around the first lane of each gel to encapsulate the contents without overlapping into neighbouring lanes.

Boxes of the same size were then placed over the other lanes on the same gel.

Numerical values were generated using Image J to represent the band density within the boxes for each lane. These values were then exported into Excel for further analysis.

The pellet and supernatant values were added together and this value represented the total protein in the solubilisation process (100 %).

The values for the supernatants from different detergent extractions were divided by the corresponding value for the total protein in each condition; the percentage of protein that had been solubilised under each condition was then calculated.

The average of each % solubilised and % unsolubilised was calculated; the standard deviation and standard error of the mean were then also determined.

### 2.2.2.16 *Langmuir Trough*

The Langmuir trough (KSV NIMA, UK) is used to observe the surface activity of SP-B and SP-C proteins. It is used to determine whether or not they have functional surface activity.

Molecules are deposited on the surface (sub-phase) of the Langmuir trough and the behaviour of the molecules is observed under maximum compression. Compression of the sub-phase surface, reduces the surface area and causes molecules to adopt a different conformation. Changes in surface tension and surface pressure are measured by a Wilhelmy plate.

Initially, the trough is very thoroughly cleaned to remove any contamination. Powder-free nitrile gloves are additionally used (Fisherbrand, Fisher Scientific, UK) and Kimtech precision wipes (Code 75512, Kimberley Clark Professional, UK).

HPLC grade water is then used to fill the trough forming the sub-phase.

This liquid is maintained at 298K and 25°C throughout.

Molecules are injected using a Hamilton syringe onto the surface of the HPLC water sub-phase, in between the computer controlled barriers.

The injected material is then left to disperse for 10 minutes.

Computer controlled barriers, having been set to a speed of 50cm<sup>2</sup>/min, are initiated after that time, the surface molecules having by then dispersed across the surface.

The Wilhelmy plate (23 mm X 10 mm X 0.5mm) is connected to a balance. This balance measures the surface pressure data outputs as the trough surface area is increased or reduced.

This information is data logged, recording surface area against surface pressure.

Pressure data against varying surface area (100-400cm<sup>2</sup>) is plotted with the pressure data. Pressure data on the X axis and the surface pressure on the Y axis. This data then yields an indication of levels of surface activity, if any.

See Figure 1.8 for an illustration of the Wilhelmy plate.

SP-B and SP-C were added individually to the trough in 25 µl aliquots. These were rinsed with HPLC grade methanol between additions.

#### 2.2.2.17 *Circular Dichroism*

All CD experiments were carried out by Charles Moore-Kelly at Birmingham University using a JASCO J-1500 Spectropolarimeter. (Data pitch: 0.5 nm, bandwidth: 1 nm, accumulations).

Surfactant proteins at a concentration of 0.1 mg/ml were transferred to a 1 mm path length cuvette.

CD ellipticity data between wavelengths 195 nm-260 nm was collected.

Proteins were suspended in a 10 mM Sodium phosphate buffer (pH 8), this formed a blank, which was deducted from the spectra.

Thermal melts were carried out at 5 °C intervals.

#### 2.2.2.18 *Statistical Analysis*

PRISM software was used to analyse the densitometry data (from 2.2.2.15 above) to enable statistical analysis of solubilisation efficiency.

### **2.3 Mass spectrometry**

#### 2.3.1.1 *Sample Preparation - Lipid Extraction*

Lipids were extracted from SMALP-encapsulated samples using methyl-tert-butyl ether (Matyash *et al.* 2008).

Lipids from 50 µl of sample were extracted in polypropylene tube (2 ml, microcentrifuge).

The sample was mixed with 375 µl of methanol and vortexed for 10 seconds.

1250 µl methyl-tert-butyl ether was then added and vortexed again for 10 seconds.

Lipids were extracted for 60 minutes at 4 °C on a rotary shaker.

375 µl of MS grade water was added, followed by a 10 seconds vortex, and a further 10 minute extraction on rotary shaker at 4 °C.

The sample was centrifuged at 1,000 g for 10 minutes, and upper phase then collected and dried under vacuum (Vacuum concentrator, Eppendorf, Germany).

Samples were stored at -20 °C prior analysis.

#### 2.3.1.2 *Mass Spectrometry: Associated Analytical Processes*

Lipid extracts with a volume of 150 µl were dissolved in electrospray (ESI) solution containing 5 mM ammonium formate in a mixture of methanol and chloroform (2:1 v/v) before analysis by mass spectrometry (MS).

Using a syringe pump (Harvard Apparatus, USA) operating with a flow rate of 500 nl/min and PicoTip ESI emitters (New Objective, Germany), the sample was directly infused in a 5600 TripleTof mass spectrometer (AB Sciex, UK). This was operating in a positive ion mode with an ionisation voltage of 2500 V.

The mass spectrometer was calibrated prior to use (<10 ppm sensitivity) on both the MS and tandem mass spectrometry (MS/MS) levels.

MS and MS/MS spectra were acquired in an information dependent mode (IDA) for 15 minutes.

One IDA cycle consisted of a MS survey scan ( $m/z$  range from 400 to 1250) followed by consecutive CID fragmentations (10 V collision energy) of the six most abundant ions in MS survey scan.

Acquired ions were temporarily excluded from MS/MS acquisition for 300 seconds.

#### 2.3.1.3 *Data Analysis*

Data analysis was carried out using Peak View software (version 2.2, AB sciex, UK).

Ion list with precursor  $m/z$  values, averaged over the acquisition time, was exported to the Lipid maps search.

Phosphatidylcholines (PC), phosphatidyletanolamines (PE) and phosphatidylserines (PS) were searched against the database as  $[M+H]^+$  ions.

Phosphatidylinositols (PI), phosphatidic acids (PA) and phosphatidylglycerols (PG) were searched against the database as  $[M+NH_4]^+$  and  $[M+H-H_2O]^+$  ions.

Diacyl-(DAG) and triacyl-glycerols (TAG) ions were searched in all positive ions forms.

In addition, 'open' lipid search against database was performed to include any lipids that may have remained unidentified in 'targeted' searches.

All positive identifications were manually inspected for the presence of lipid-specific fragment ions or neutral losses.

#### 2.3.1.4 *Mass Spectrometry confirming Protein Identification*

Proteins were sent to Birmingham University's mass spectrometry department for independent identification.

Protein samples were trypsin digested and analysed using liquid chromatography in tandem with LTQ Orbitrap Elite ETD mass spectrometer (ThermoFisher Scientific).

Data analysis was performed using Sequest (Proteome discover - ThermoFisher).

The Sequest algorithm calculates all molecular formulae that theoretically fit a given mass, and this is compared with derived peptide sequences.

A probability score is assigned, dependent on the match between the theoretical spectra and the actual spectra.

## 2.4 **Computational Modelling**

All computational modelling, including the contextual inputs from myself as described in Chapter 6, was carried out by Dr John Simms (Aston University/Now Coventry).

*Ab initio* models of SP-B and SP-C were generated using the membrane module of Rosetta using standard methods.

Rosetta uses short sequences of the peptide (SP-B) and uses a protein database to compare proteins with the same sequence that have a known structure.

10,000 models were generated using Rosetta.

The top 100 scoring models were then exposed to Replica Exchange Monte Carlo (REMC) (Sugita and Okamoto 1999).

Temperature simulation boxes, examining sample SP-B proteins at different temperatures, are then examined to assess against thermodynamic issues e.g. overcoming energy barriers.

This is based on empirical data drawn from a wide data base of membrane proteins.

The final SP-B model, presented later in Chapter 6, was ultimately refined using an in house scoring method (manuscript in preparation; Simms and Poyner).

This scoring method has been validated by a model protein (RAMP) that was created using this method. Its structure was then later confirmed by X-ray crystallography data (Simms 2006).

REMC simulations were used to examine the conformational changes in SP-B within the confines of an implicit membrane, 90Å in depth.

The terminal ends of SP-C are charged and so insertion into a hydrophobic lipid environment would be energetically unfavourable.

However, unlike a bilayer, the tear film is too large for the peptide to traverse.

In order to generate a suitable starting orientation for SP-C, an objected procedure was used to position the peptide.

The resulting position placed the peptide perpendicular to the membrane (the Z axis normal) at a position in which the peptide was partially inserted in the lipid headgroups.

REMC was employed to overcome any energy barriers between conformations in the absence of lipids, but in the presence of a hydrophobic region with lateral pressure and friction.

The advantage of this technique is that all conformations can be sampled rather than a subset when lipids are introduced.

The *ab initio* folding simulation started from an extended polypeptide chain and is “condensed” into a folded structure with a defined 3D structure.

The primary sequences for SP-B and SP-C were obtained from uniprot (accession numbers: SP-B: P07988 SP-C: P11686) and used to generate the *ab initio* model.

HIPPO was used for the implicit simulations and optimal parameters for liquid simulation force field (OPLS) (Ulmschneider and Ulmschneider 2009).

Implicit membranes were 90 Å to reflect the size of the tear film lipid layer.

Gromacs was used for the explicit models; Martini coarse grain force field used (Monticelli *et al.* 2008).

Coarse grain simulations take the main groups of a molecule and simulate these atoms as beads (Van Der Spoel *et al.* 2005).

The whole concept and deductive power of such modelling is fully described later in Chapter 6 of this work.

## Chapter 3

### **Production and Purification of Recombinant ProSP-B and SP-B**

To explore the role of SP-B in lipid spreading within the tear film lipid layer, SP-B needed to be produced as a recombinant protein.

The approach adopted in this work allowed the cellular machinery of the yeast cell to be utilised to produce SP-B, alongside other yeast proteins.

ProSP-B, the full length protein, and SP-B the mature active peptide, have both been produced within this work.

This chapter details the sequence of experiments designed and executed to pursue this; from the cloning of both genes; expression of the recombinant proteins; subsequent extraction using detergents and polymers; and consequent studies to examine surface tension lowering functionality.

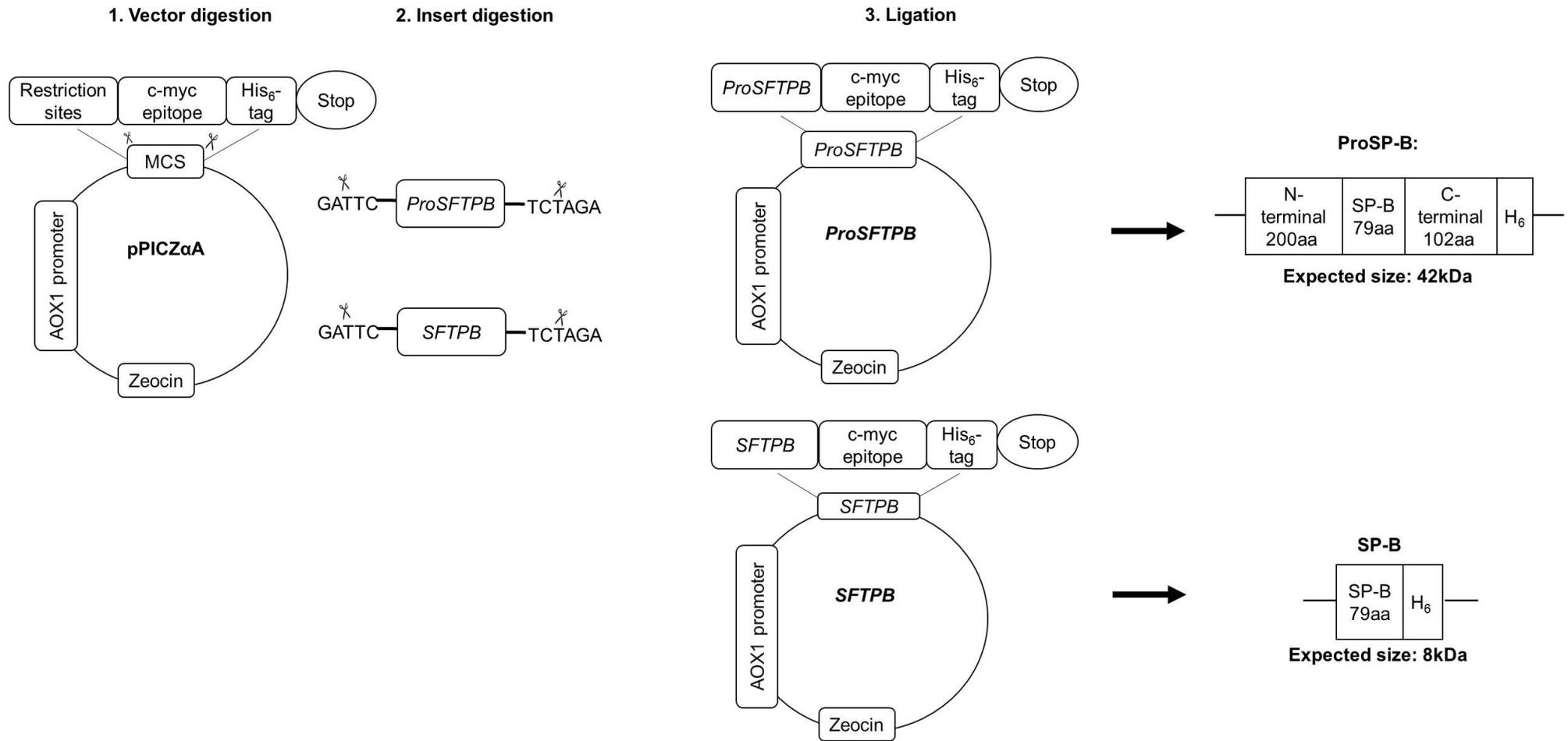
#### **3.1 Design and Construction of ProSP-B and SP-B Expression Constructs**

This section describes the cloning of *ProSFTPB*, encoding full length proSP-B and *SFTPB*, and encoding the mature active SP-B protein.

ProSP-B is flanked by an N-terminus and a C-terminus on either side of the mature polypeptide sequence, whilst SP-B is composed of the mature domain only. See the primary sequence in Figure 1.11 for further details on the amino acid composition and protein domains.

A cloning simulation was first undertaken using Gentle, an online molecular cloning tool. Gentle allowed *in silico* digestions and ligations to be simulated.

This helped to confirm that the *ProSFTPB* and *SFTPB* sequence were in frame with the pPICZ $\alpha$ A vector His<sub>6</sub>-tag. The procedures adopted to form these *in silico* transformed *ProSFTPB* and *SFTPB* vectors are illustrated in Figure 3.1.



**Figure 3.1** *ProSFTPb* and *SFTPb* Vector Map Design

DNA sequences for yeast plasmid pPICZαA and *ProSFTPb* were digested (1 & 2) and ligated (3). Gentle molecular cloning software was used to check if the cloning of *SFTPb* was in frame with the pPICZαA vector His<sub>6</sub>-tag. The protein transcripts are also illustrated. *ProSFTPb* encoding proSP-B is expected to be at 42 kDa; *SFTPb* encoding SP-B is expected to be 8 kD.

### 3.2 *ProSFTPB* Synthesis

*ProSFTPB* cDNA was amplified using PCR; (Figure 3.2 panel 1).

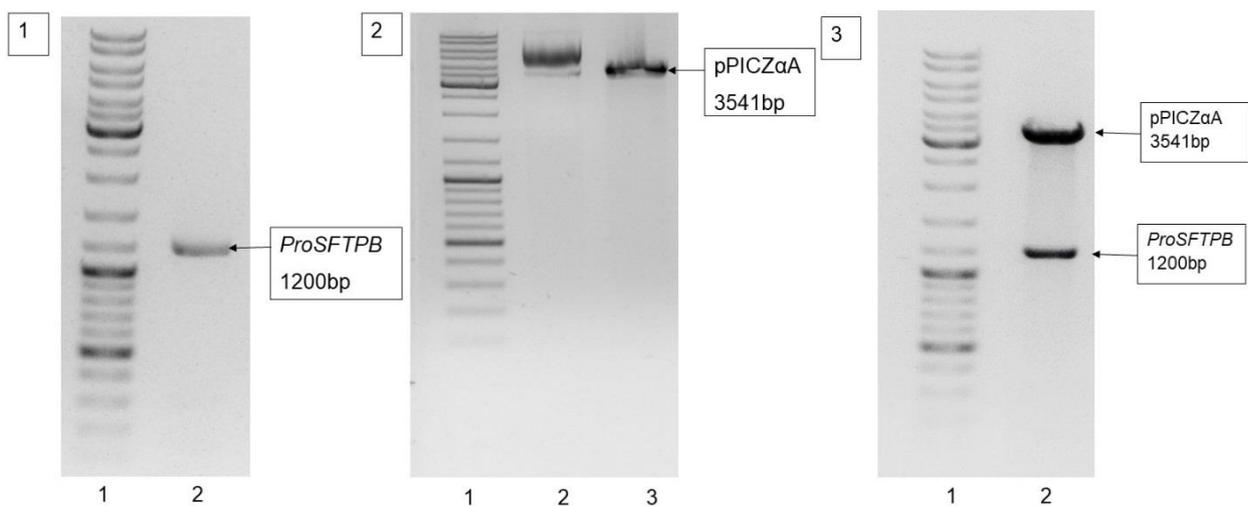
The PCR products were purified and digested with *EcoR1* and *Xba1*.

pPICZ $\alpha$ A vector DNA was also digested with *EcoR1* and *Xba1*; (Figure 3.2 panel 2).

Ligations were transformed into XL10 Gold competent *E. coli* cells.

Transformed *SFTPB* DNA was digested with *EcoR1* and *Xba1*; (Figure 3.2 panel 3).

Digestion of the transformed *ProSFTPB* vector resulted in two bands. A higher band at 3,541 bp represented the vector backbone, and a lower band of 1,200 bp represented the *ProSFTPB* insert, thus encoding the full-length sequence of proSP-B.



**Figure 3.2 *ProSFTPB* Synthesis**

*ProSFTPB* DNA was amplified using PCR (1).

The PCR fragments (1) and pPICZ $\alpha$ A vector were digested with *EcoR1* and *Xba1* (2), ligated and transformed into competent *E. coli* XL10 Gold.

The transformed vectors were digested again (3) with *EcoR1* and *Xba1* and revealed *ProSFTPB* (1,200bp) inserts and vector backbone (3,451bp).



The differing molecular weights are a reflection of the effect of each extraction process on the structure of SP-B, this was further complicated by the propensity of SP-B to form dimers and higher order structures.

The formation of these structures and potential inter-relationship with the extreme hydrophobicity of the protein are also rather likely to have contributed to this phenomenon.

The overall influence of the extraction process and the extreme hydrophobicity manifested during protein separation on SDS-PAGE was seen in the form of a failure of SP-B to migrate to the anticipated “correct” size/molecular weight consistently.

Later, fuller discussions of the likely reasons for this relate to various aspects of the different experimental processes employed in pursuit of optimising both quality and yields.

These processes seek to accommodate the protein’s hydrophobicity whilst optimising the separation and purification stages.

These are the constructs and their expected sizes:

ProSP-B (full) 42 kDa

SP-B (mature) 18 kDa

### **3.3 Isolation of High-Yielding ProSP-B-Producing Colonies**

Transformed *ProSFTP*B was integrated within the yeast genome.

Yeast colonies were preliminarily screened for protein expression, with the aim of isolating a high-yielding proSP-B colony.

Transformed colonies were spread onto YPDS plates containing 100 µg/ml, 250 µg/ml and 500 µg/ml zeocin and left for 3 days in a 30°C incubator.

The 500 µg/ml plate was screened preliminarily to test for protein expression. Eight colonies were randomly selected from the YPDS zeocin plate, cultured in BMGY and then induced in BMMY at an  $A_{600}$  of 1.

Isolated membrane fractions were then visualised using a Western blot (Figure 3.4 panel A).

The 225 kDa band seen in lanes 1, 2, 3, 4 and 8 was deduced to be aggregated proSP-B; as the protein had not moved far from the well and was so much larger in size than the 42 kDa expected size.

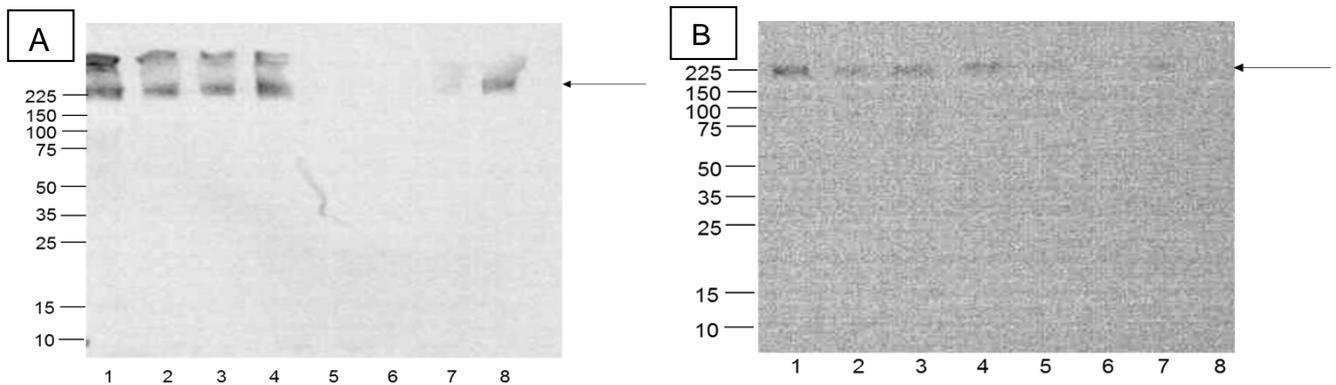
The experiment was repeated with urea sample buffer being added to attempt to disrupt non-covalent bonds and prevent the apparent aggregation from occurring. The proSP-B samples were again loaded onto an SDS-PAGE gel and transferred to a Western blot, see below in Figure 3.4 panel-B.

The proteins bands were however at the same point as previously seen.

Betamercaptoethanol disrupts disulphide bonds, and Urea is a denaturant, each being used routinely to unfold proteins in SDS-PAGE.

However, the high hydrophobicity of proSP-B may well have contributed to the failure here of SDS to unfold the proteins uniformly. A further additional contributing factor is the use of a detergent (SDS), which is itself surface active, to unfold a surfactant protein.

This could have well also have led to the aggregation seen.



**Figure 3.4 ProSP-B Protein Expression Screen**

A) Transformed *ProSFTP*B vector DNA was linearized and transformed into X-33. Protein, yield was assessed using a Western blot.

A 225 kDa band can be seen in lanes 1, 2, 3, 4 and 8. B) Repeated Western blot of proSP-B samples from panel A analysed in urea sample buffer. The proSP-B protein ran to the same 225 kDa size as seen in panel A.

25  $\mu$ g total membrane protein was loaded into each well.

The arrows marks the point of suspected aggregation.

### 3.4 Identification of a High-Yielding Colony for Further Production Scale-Up

Despite the possible aggregation, colony 1 see below in Figure 3.5A, was cultured in a bioreactor to determine if more precise control over the culturing conditions themselves, could promote more efficient folding and thus yield a more definitive expression of proSP-B.

The initial transformation screen had been carried out in flasks. Whilst these conditions are valuable for initial trialling, they are distant from optimum conditions, due to there being no ongoing process control, particularly over either pH or dissolved oxygen (DO). These control issues may have resulted in/exacerbated proSP-B aggregation.

Bioreactors allow significantly better control over the growth conditions of the culture, providing the ability to monitor it in real time.

This allows appropriate changes to the fermentation process control as required.

Parameters such as pH, temperature and dissolved oxygen (DO) can be tightly controlled; with the tailoring of these growth conditions generally helping to maximise the amount of proSP-B produced.

Colony 1 was chosen because it had the most intense band, and so was expected to result in higher protein expression compared to the other colonies.

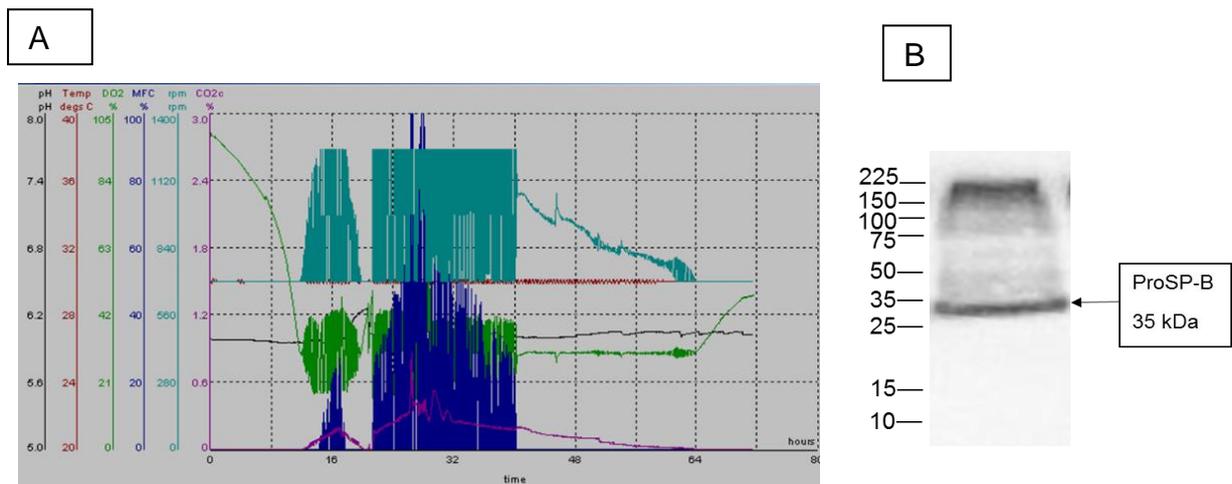
A representative trace from the bioreactor, see below in Figure 3.5 panel-A.

A Western blot of the corresponding membrane fraction from the proSP-B-producing culture is shown below in Figure 3.5 panel-B.

The green trace reflects the growth of the culture; the oscillations that can be seen around 30%, indicated the culture was growing well.

The membrane fraction was isolated from the culture and loaded onto an SDS-PAGE gel and transferred to a Western blot.

The results show that isolated membrane fractions from cultures grown in bioreactors displayed a very clear band at 35 kDa.



**Figure 3.5 Scale Up of a High-Yielding ProSP-B Colony in a 2L Bioreactor**

**A)** A bioreactor trace representing growth of the culture over 60 h.

The growth of the proSP-B-expressing culture was monitored by the green dissolved oxygen (DO) trace. PH was set to pH 6.0 throughout the run. The temperature (red) was maintained at 30°C for the duration of the production run.

**B)** The isolated membrane fraction was loaded onto an SDS-PAGE gel and a Western blot carried out. Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used for detection.

A band at circa 35 kDa can be seen.

Later experiments confirmed the 35 kDa band seen is proSP-B, see later in Tables 3.1 and 3.2.

Although the molecular weight of 35 kDa was lower than the expected level of 42kDa, the clear isolation of a strong band, see above in Figure 3.5 panel-B.

***This demonstrates that, for the first time, proSP-B has been successfully expressed in Pichia pastoris.***

### 3.5 Solubilisation of ProSP-B with Organic Solvents

Chloroform ( $\text{CHCl}_3$ ) and methanol (MeOH) extraction is the most cited method of isolating SP-B (Johansson, Curstedt and Jornvallt 1991); and so this method was initially investigated.

The method separates hydrophobic proteins into an organic layer, and hydrophilic proteins into an aqueous layer.

A 20  $\mu\text{l}$  sample was removed from each layer and loaded onto an SDS-PAGE gel, then transferred to a Western blot, see below in Figure 3.6.

A faint 35 kDa band and a 55 kDa band can be seen in the aqueous, organic layers and in the pellet, indicating that not all the proSP-B had separated into the organic layer.

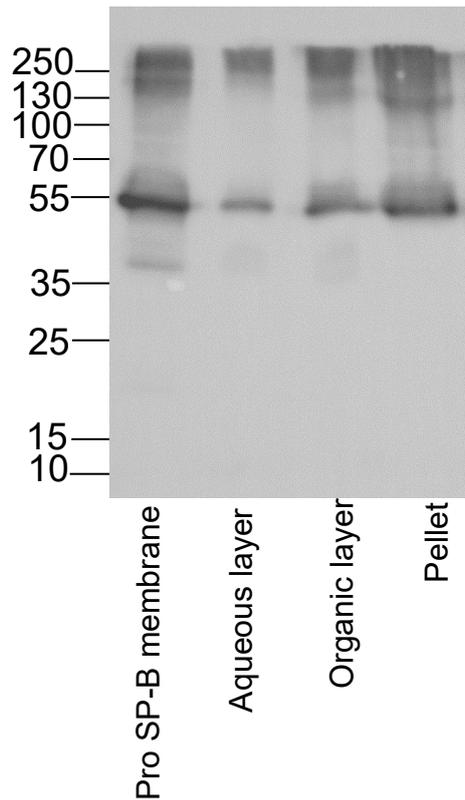
This 35 kDa band is lower again compared with an anticipated size of 42 kDa but compares with the previous molecular weight of 35 kDa, again see above in Figure 3.5 -panel B.

Aggregated proSP-B can also be seen at the top of the blot, indicating that some proSP-B may have again aggregated during the solubilisation process; or even earlier in the production process.

The production of recombinant proteins places stress on the yeast cellular machinery, something which may have resulted in the formation of aggregated protein during the protein expression.

The change in molecular weight from the 35 kDa size to 55kDa, may well also be closely attributed to the hydrophobicity of proSP-B when separating from membrane lipids.

Previously proSP-B was associated with membrane lipids but extraction from the membrane may cause hydrophobic regions of the protein to repeatedly unfold in a non-linear manner in the SDS-PAGE.



**Figure 3.6 Extraction of ProSP-B using Chloroform and Methanol**

ProSP-B was solubilised using chloroform and methanol.

The method separates hydrophobic proteins between an organic layer and an aqueous layer.

The different fractions were run on an SDS-PAGE gel and transferred to a Western blot.

A 35 kDa and 55 kDa protein band can be seen in the aqueous layer, organic layer and the pellet.

Anti-His<sub>6</sub> monoclonal and anti-mouse IgG, HRP-linked secondary antibodies were used.

Solubilisation of proSP-B using organic solvents indicated that proSP-B could be isolated from the membrane, however during this process other membrane proteins are also co-isolated.

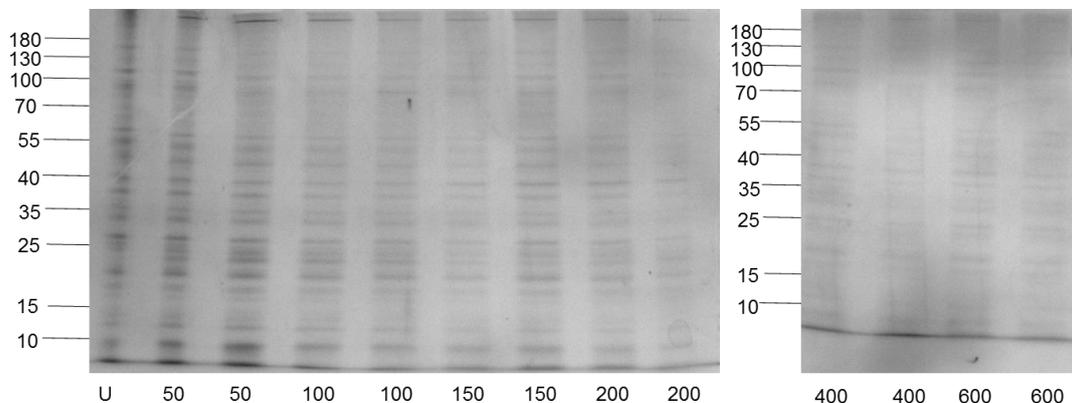
In order to separate proSP-B from other membrane proteins, it needed to be isolated via purification.

A large scale solubilisation of proSP-B using chloroform and methanol was therefore subsequently carried out with the aim of getting pure fractions of proSP-B.

The organic layer was mixed with Ni-NTA and purified using an imidazole gradient.

The purification of proSP-B then resulted in the co-purification of many other proteins, making it difficult to absolutely positively identify proSP-B.

Consequently, this extraction method was not pursued further, see below in Figure 3.7.



**Figure 3.7 Purification of Chloroform and Methanol Extracted ProSP-B**

ProSP-B was solubilised using chloroform and methanol, the organic layer being separated and mixed with Ni-NTA using an imidazole gradient.

Purified proSP-B proteins were run on an SDS-PAGE and stained with Coomassie dye.

A cohort of other proteins was co-purified with proSP-B.

ProSP-B was identified at 55kDa previously. Due to this co-purification of other membrane proteins, it is difficult to identify using this purification process.

### **3.6 Polymer Solubilisation of ProSP-B**

Chloroform and methanol extraction of proSP-B was found to be unsuitable as a precursor to purification, as shown in Figure 3.7 above.

An alternative method of extraction was therefore explored.

Styrene maleic acid (SMA) is an alternating co-polymer of styrene and maleic acid. SMA is known to encapsulate proteins and lipids in 10-12nm nanodiscs (Knowles 2009).

ProSP-B-containing membranes were solubilised with SMA 2000P, 25010 and 30010.

Both the supernatant and pellet were loaded onto an SDS-PAGE gel and transferred to a Western blot. The supernatant contained solubilised protein; denoted by (S), and un-solubilised protein denoted by (P).

There is no proSP-B present in any of the supernatant fractions. This indicated proSP-B had not been solubilised, see below in Figure 3.8 - panel A.

The experiment was repeated with the addition of a sonication step.

Sonication provided a more mechanical disruption of the membrane, the membrane lipids and proteins are temporarily 'pulsed' and this causes the hydrophobic regions to seek the hydrophobic areas of the SMA thus becoming encapsulated.

The solubilised fractions were loaded into an SDS-PAGE and transferred to a Western blot (Figure 3.8 panel-B) and now indicated that proSP-B could be effectively solubilised using all the SMA variants.

This was clearly identified by the presence of a 35 kDa band in all the supernatant fractions, consistent with the size of the band seen above in Figure 3.5 panel-B.

Further discussion follows below.

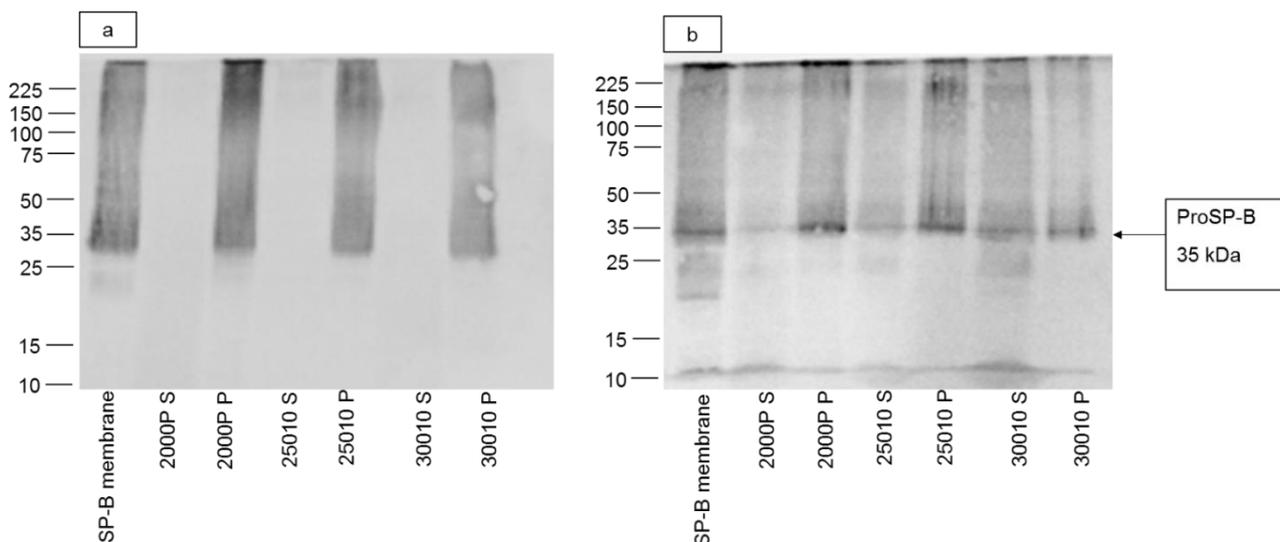
The success of the solubilisation after sonication is further explained by an effective physical disruption of the membrane, causing the hydrophobic proSP-B to seek the hydrophobic environment in the SMALP.

The clear result was that a distinct 35 kDa proSP-B protein could be seen in the solubilised fractions. This was again in line with the 35 kDa size seen from the membrane fraction isolate.

Some aggregated proSP-B protein can also be seen at the top of the blot.

Solubilised proSP-B was purified using Ni-NTA and yet again the same 35 kDa band was observed with a high degree of purity, see below in Figure 3.9.

***This was the first time extraction of proSP-B using the SMA polymers has been demonstrated.***

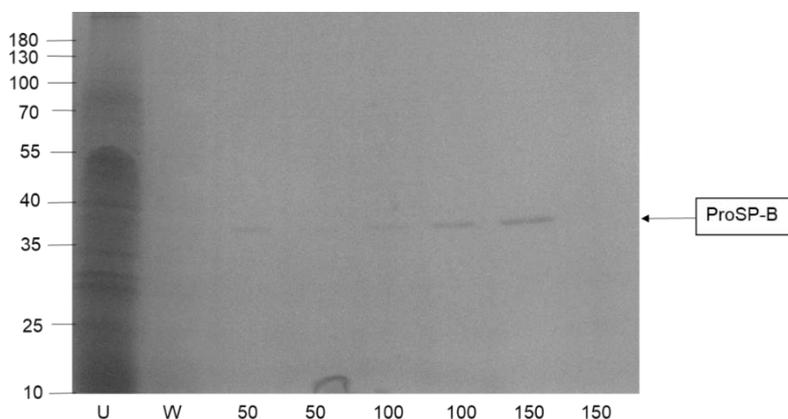


**Figure 3.8 ProSP-B SMA Solubilisation Screen**

A solubilisation screen, using different SMA polymers, was set up; one set of proSP-B samples being sonicated (B), whilst the other set was not (A).

The sonicated proSP-B samples have each been solubilised by the SMA and migrated to a size of 35kDa - indicated by the arrow.

The samples were separated and loaded into an SDS-PAGE gel and a Western blot carried out. Anti-His<sub>6</sub> monoclonal antibody and anti-mouse IgG-HRP-linked antibodies were used.



**Figure 3.9 ProSP-B Purification via Solubilisation with SMA2000P**

ProSP-B was solubilised and purified using SMA2000P.

An imidazole gradient ranging from 50 mM – 200 mM was used to find the optimal concentration that proSP-B eluted at. Purified proSP-B proteins were run in an SDS-PAGE gel and stained with Coomassie dye.

A single band, under the 40 kDa molecular weight was observed faintly in the 50 mM and 100 mM elutions. The strongest band was present in the 150 mM elutions.

SMA solubilisation experiments were run in parallel with functional physical measurement (surface tension) experiments on proSP-C10 that had also been extracted with SMA, see Chapter 4; section 4.11.

The SMA polymer, as with detergents, had of course needed to be subsequently removed because of potential contributory surface activity. This would have been likely to affect the subsequent protein functionality experiments using a Langmuir trough.

A decision to move to detergent solubilisation with a view to reconstitute proSP-B was therefore made.

This process involves detergent solubilisation initially, with the detergent subsequently removed and replaced with lipids.

### **3.7 Detergent Solubilisation of ProSP-B**

The following detergents were all used to solubilise proSP-B.

Detergents N-lauroylsarcosine sodium salt (NLSS), Anzergent (Anz), Fos-choline 12 (Fos-12), 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS), maltose neopentyl glycol-3 (MNG), octyl-beta-glucoside ( $\beta$ -OG), decyl maltoside (DM), and Triton X-100.

This screen represented a wide range of potentially applicable detergents, and was aimed at determining an optimal detergent for proSP-B extraction.

Both solubilised and un-solubilised fractions were run on an SDS-PAGE gel and a Western blot was carried out thereafter, see below in Figure 3.10.

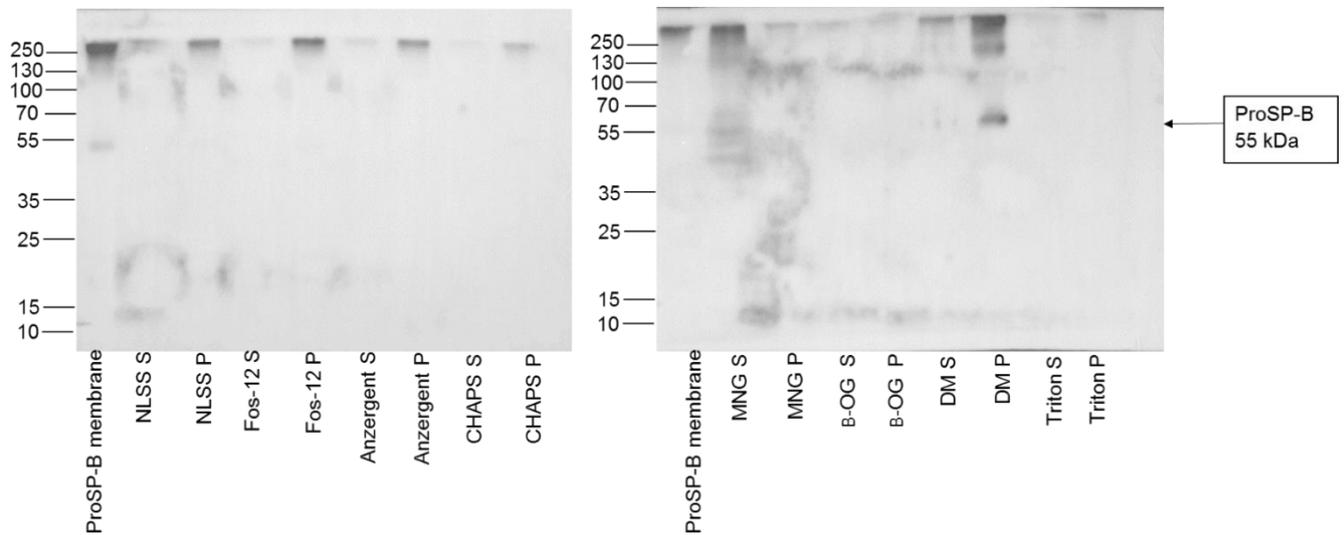
Much of the protein appeared to be aggregated; was predominantly in the pellet fractions; as well as being unable to move any further than the well in all the lanes.

The exception was a band at 55 kDa in the DM P lane; this protein size had been previously observed in the chloroform and methanol extraction work, see above in Figure 3.6.

This detergent solubilisation screen of proSP-B was *repeated three times*, and the same protein aggregation was seen each time.

Further detergent solubilisation was therefore not pursued because here the proSP-B protein was only present in the un-solubilised fractions.

The difficulty in solubilising proSP-B with detergents, as mentioned earlier, may have resulted from the particularly hydrophobic nature of the protein and the inability of these detergents to solubilise such hydrophobic surfactant proteins.



**Figure 3.10 Detergent Solubilisation of ProSP-B**

ProSP-B was solubilised with 5% NLSS, Fos-12, Anzergent, CHAPS, MNG,  $\beta$ -OG, LDAO, DM and Triton.

Solubilised (S) supernatants and un-solubilised fractions (P) were loaded into an SDS-PAGE, and a Western blot then carried out.

ProSP-B protein aggregated in the presence of all detergents used, despite multiple attempts to solubilise it.

Both solubilised and un-solubilised fractions don't appear to run far from the loading well.

The 55kDa protein that is arrowed appears to represent; aggregated, un-solubilised proSP-B.

Anti-His<sub>6</sub> monoclonal antibody and anti-mouse IgG-HRP-linked antibodies were used.

### 3.8 Solubilisation of ProSP-B with Guanidine Hydrochloride

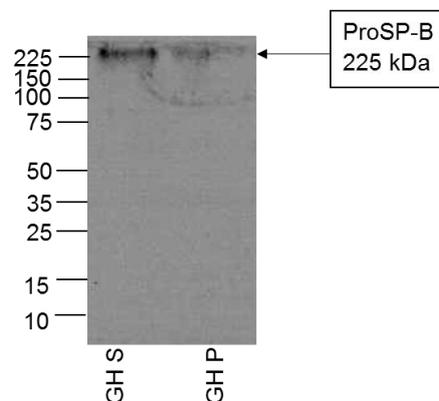
Due to the difficulty in extracting proSP-B using detergent, Guanidine Hydrochloride, (GH) a known denaturant was then used to solubilise the proSP-B membrane pellet. This is based on extraction of SP-B from *E.coli* in a published patent (Weaver 1996).

The supernatant and pellet were loaded onto an SDS-PAGE gel and transferred to a Western blot (Figure 3.11);

ProSP-B is seen at the top of the gel, and is present in both the solubilised supernatant fraction and also the pellet.

Previous results indicated a protein size of 35 kDa for proSP-B, see above in Figure 3.5 - panel B. It appears that proSP-B in this extraction seems to have aggregated.

This aggregation was also again likely to be caused by the unfolding of proSP-B by GH causing the hydrophobic “domains” to fuse together.



**Figure 3.11 Solubilisation of ProSP-B using GH**

GH was used to solubilise proSP-B.

The supernatant (S) and pellet (P) were separated, loaded onto an SDS-PAGE gel and transferred to a Western blot.

ProSP-B aggregated at the top of the gel and is indicated by a labelled arrow.

Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used for detection.

### 3.9 Purification of ProSP-B using GH

Despite the aggregation seen in the proSP-B sample, see in Figure 3.11; GH-solubilised proSP-B was purified using Ni-NTA with the intention of then identifying proSP-B via either using an SP-B specific antibody or mass spectrometry.

Purified fractions were visualised using a Western blot.

These revealed proSP-B had aggregated at the top of the gel.

See below in Figure 3.12 panel-A.

GH causes the unfolding of proteins and the aggregation seen may be the result of the preference for the hydrophobic regions to be “buried” so causing the aggregation seen.

Proteins solubilised using GH are then thought to refold once the GH has been removed. Although the purification was thought to be enough for proSP-B to refold, there may still have been an excess of GH.

This issue is addressed below.

Solubilisation of proSP-B in GH was repeated, but in order to address the aggregation, the GH was removed by dialysis prior to purification.

Dialysis helped to prevent proSP-B aggregation at the top of the gel and as a result, the purification yielded protein bands around 75 kDa. Though still larger than the expected size of 42kDa and the previously seen 35 kDa, the higher 75 kDa molecular weight band may well reflect a dimer conformation, see below in Figure 3.12 panel - B.

The authors of the patent, (Weaver 1996) cited the use of  $\beta$ -OG after solubilisation with GH.

This was also investigated as a method to obtain a protein of the 35 kDa size previously seen.

ProSP-B was solubilised with GH and then dialysed against Tris HCl (pH 8) to remove the GH that was thought to be the cause of the aggregation seen previously.

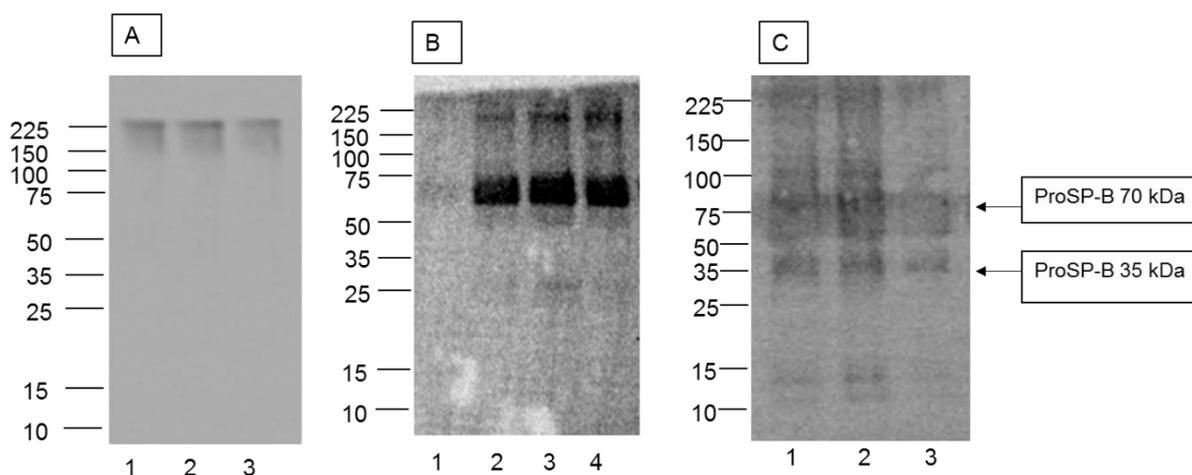
The proSP-B supernatant was then treated with 1%  $\beta$ -OG and purified using Ni-NTA.

ProSP-B purification fractions were visualised using a Western blot.

The 35 kDa band that had previously been seen was indeed visible, this time together with a higher band at 75 kDa as seen above and Figure 3.12 panel - C below.

As these sizes of proSP-B in each of the purifications differed, both the 75 kDa and 35 kDa bands were also sent for independent identification using protein mass spectrometry.

Most Importantly, BOTH molecular weight sizes had a positive ID for proSP-B, which is shown in Tables 3.1 and 3.2 indicating proSP-B may be in a dimeric complex at the higher 75 kDa molecular weight.



**Figure 3.12 Purification of ProSP-B using GH**

ProSP-B was solubilised with GH and purified using Ni-NTA.

The fractions were separated and run on an SDS-PAGE gel and a Western blot carried out.

ProSP-B aggregated at the top of the gel (Panel A).

To prevent aggregation, the GH was dialysed away before purification. This prevented proSP-B aggregating and it moved down the gel to a size of circa 70 kDa (Panel B).

Panel C shows purification of proSP-B that had been treated with  $\beta$ -OG after solubilisation with GH producing a 35 kDa proSP-B band. ProSP-B bands are arrowed and labelled.

ProSP-B can be seen at a size of 35 kDa and 70 kDa (as confirmed by mass spectrometry in Tables 3.1 and 3.2).

The two bands thought to be proSP-B were excised from a Coomassie stained gel, above in lane 2 Figure 3.12 Panel-C, and sent for mass spectrometry analysis at Birmingham University.

The gel plug was trypsin digested and samples were run on a tandem LC-MS.

Data analysis was performed using Sequest (Proteome Discoverer, Thermo Fisher).

Identified peptides were given a score based on the expected number of peptides found, relative to the number of peptides identified from the spectra. The higher the score the better the peptide matches between the spectra and the database.

Band 1 was thus identified as proSP-B; this corresponded to the 70kDa excised sample.

The peptides identified from the mass spectrometry analysis matched peptides from the primary sequence of proSP-B confirming the correct protein was expressed, see in Table 3.1.

The score assigned to band 1 (proSP-B) was 757.26, see below in Table 3.2.

The next identified peptide was Keratin, which had a score of 131.55, much lower than the proSP-B score, and this difference confirmed that the proSP-B peptides identified had the highest probability score, refer below to Table 3.2.

Note; Keratin from human skin is commonly found as a contaminant in mass spectrometry samples of this type (Hodge *et al.* 2013).

Band 2 corresponded to the 35kDa proSP-B sample, to which the mass spectrometry analysis assigned a score of 106.50 (Table 3.3). This was the highest score in the list and indicated there was a high abundance of proSP-B peptides found in the spectra that matched the peptide database. These peptides matched peptides in the primary sequence of SP-B and are listed in Table 3.1.

***The mass spectrometry analysis indicated that the 35 kDa and 70 kDa bands are proSP-B and confirms the successful synthesis and expression of proSP-B.***

**Table 3.1 Positive Identification of Recombinant ProSP-B**

Identified proSP-B fragments from mass spectrometry analysis of band 1 (70 kDa) and band 2 (35 kDa) have sequence complementarity to peptides in the primary sequence of proSP-B. In the table, the fragments have been organised according to the domains that they were found to match (N terminus, C terminus or mature). The highlighted yellow region represents the mature amino sequence of SP-B.

MAESHLLQWLLLLLPTLCGPGTAAWTTSSLACAQGPEFWCQSLEQALQCRAIGHCLQEIVWGHV GADDLCQECEDIVHILNKMAKEAIFQDTMRKFLEQECNVLPLKLLMPQCNQVLDDYFPLVIDYFQN QTDSNGICMHLGLCKSRQPEPEQEPGMSDPLPKPLRDPLPDLKLVLPVLPALQARPGPHTQ DLSEQQ <b>FPIPLPYCWL</b> <b>CRALIKRIQAMIPKGALAVAVAQVCRVPLVAGGICQCLAERYSVILLDT</b> <b>LLGRMLPQLVCRLVLRCSM</b> DDSAGPRSPTGEWLPRDSECHLCMSVTTQAGNSSEQAIPQAMLQ ACVGSWLDREKCKQFVEQHTPQLLTLVPRGWAHTTCQALGVCGTMSSPLQCIHSPDL		
N terminus	Mature	C terminus
ALGHcLQEIVWGHVGADDLcQEcEDIV HILNK	VVPLVAGGlcQcLAERYs VILLDTLLGR	cKQFVEQHTPQLL TLVPR
SRQPEPEQEPGmSDPLPKPLRDPLP DPLLDK	GALAVAVAQVcR	cSmDDSAGPRsPT GEWLPR
KFLEQEcNVLPLK	IQAMIPKGALAVAVAQVc R	qFVEQHTPQLLTL VPR
DSEcHLcmSVTTQAGNSSEQAIPQAm LQAcVGSWLDREK	YSVILLDTLLGRmLPQLV cR	cKQFVEQHtPQLLT LVPR
QFVEQHTPQLLTLVPR	VVPLVAGGlcqcLAERYs VILLDTLLGR	SPTGEWLPR
QPEPEQEPGMSDPLPKPLRDPLPDP LLDK		cSMDDSAGPRsPT GEWLPR
LLmPQcNQVLDDYFPLVIDYFQnQTD SnGlcmHLGLcK		
EAlFqDTMR		

**Table 3.2 Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-B (70kDa)**

Bands thought to be proSP-B were excised from a Coomassie stained gel and sent for mass spectrometry analysis at Birmingham University.

The gel plug was trypsin digested and samples were run on a tandem LC-MS.

**ProSP-B had the highest score of 757.26. GN: gene.**

Description	Score
1. Pulmonary surfactant-associated protein B GN=SFTPB [PSPB_HUMAN]	757.26
2. Keratin, type I cytoskeletal 10 GN=KRT10 [K1C10_HUMAN]	131.55
3. Keratin 1 GN=KRT1 [H6VRF8_HUMAN]	113.34
4. Keratin, type II cytoskeletal 2 epidermal GN=KRT2 [K22E_HUMAN]	107.01
5. Keratin, type I cytoskeletal 9 GN=KRT9 [K1C9_HUMAN]	90.23

**Table 3.3 Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-B (35kDa)**

**ProSP-B had the highest score of 106.50.**

1. Pulmonary surfactant-associated protein B GN=SFTPB [PSPB_HUMAN]	106.50
2. ATP synthase subunit beta, mitochondrial GN=ATP2 [ATPB_KLULA]	71.05
3. ATP synthase subunit alpha, mitochondrial GN=ATP1 [ATPA_KLULA]	57.17
4. MICOS complex subunit MIC60 GN=MIC60 [MIC60_PICPG]	32.26
5. Elongation factor 1-alpha GN=TEF1 [EF1A_PICGU]	28.07
6. Serine hydroxymethyltransferase, cytosolic GN=SHM2 [GLYC_CANAL]	21.68

The expression and purification of proSP-B has also established a method which may be used for crystallisation studies into the investigation of the 3D structure of proSP-B.

Previously, this has proved to be extremely difficult due to the high hydrophobicity of the protein.

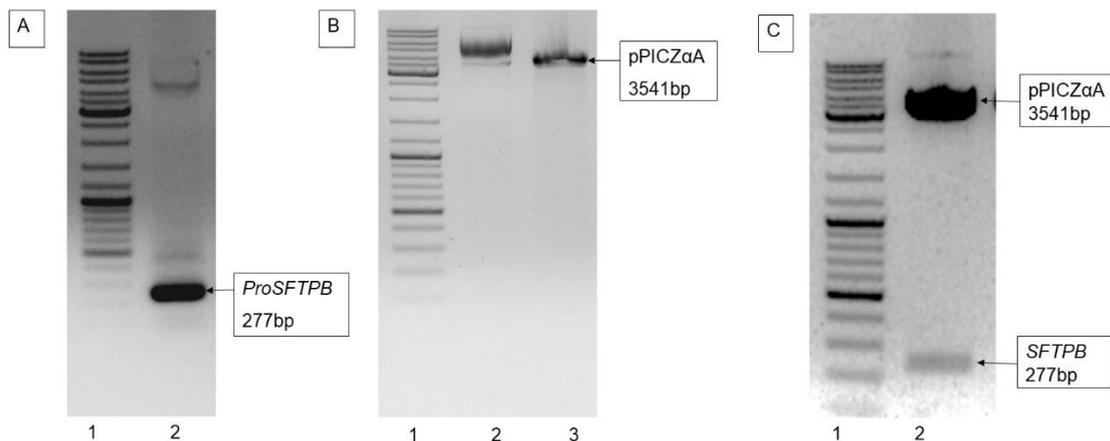
Further experiments were carried out on the mature form, SP-B see below, in section 3.10.

### 3.10 The Recombinant Production of SP-B: SFTPB Vector Construction

The cDNA encoding *SFTPB* was integrated into pPICZ $\alpha$ A vector DNA using the same methodology as carried out previously for proSP-B.

Purified PCR products and vector DNA are shown in Figure 3.13 panels A and B.

Digestion of the transformed *SFTPB* vector yielded the *SFTPB* insert (277 bp) and vector backbone (3,451bp) are seen in Figure 3.13 panel - C.



**Figure 3.13 Diagnostic Digest of the Transformed *SFTPB* Vector**

*SFTPB* DNA was amplified using PCR, see panel A.

The PCR fragments and pPICZ $\alpha$ A vector were digested with *EcoR*I and *Xba*I, see panel B, ligated and transformed into super competent *E. coli*.

A colony screen was carried out and the transformed vectors were extracted and digested again with *EcoR*I and *Xba*I to reveal the *SFTPB* (277bp) insert and vector backbone (3,541bp), see panel C.

### 3.11 Isolation of a High-Yielding SP-B Colony

Transformations into *P. pastoris* X-33 were carried out, and eight colonies were selected and cultured in BMGY and BMMY.

The membrane fractions were imaged using a Western blot, see below in Figure 3.14.

A 15 kDa band could be seen from colonies grown on the 100 µg/ml and 200 µg/ml YPDS plate. A 25 kDa band and a 15kDa band could be seen from colonies grown on the 500 µg/ml plate.

The 15 kDa molecular weight is in line with the expected 18 kDa dimeric size of SP-B, the 25 kDa may reflect a higher oligomeric structure.

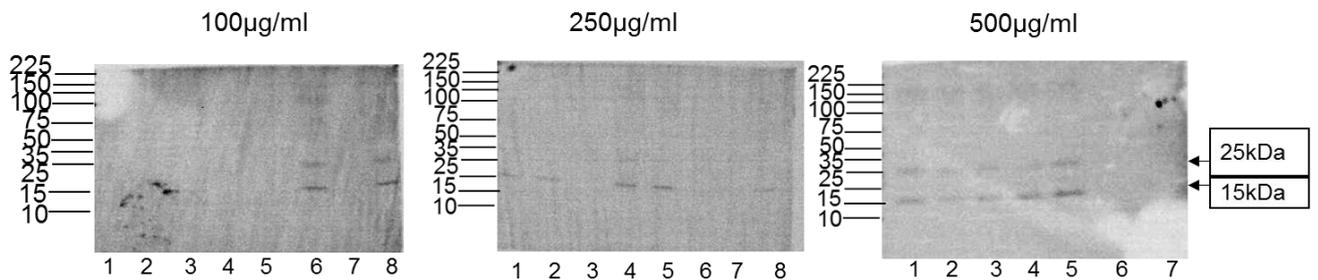
8 kDa is the expected monomeric size of SP-B, but despite using reducing agents in the sample buffer, a molecular weight size of 8 kDa was never observed.

The 15 kDa size most likely reflects the dimer conformation.

As the zeocin concentration in each plate increased, the band intensity also increased.

Colony 5 from the 500 µg/ml plate was chosen for further scale-up experiments because it had the brightest band intensity.

This indicated it was likely to be a highly-SP-B expressing colony.



**Figure 3.14 SP-B Protein Expression Screen**

The *SFTPB* vector was transformed into X-33. The transformations were spread onto YPDS plates containing 100, 250 and 500 µg/ml zeocin.

Protein bands at 25 kDa and 15 kDa could be seen in each screen.

The 15 kDa band is likely to reflect the dimer conformation of SP-B, whilst the 25 kDa band may be representative of an oligomer structure.

The protein yield from each colony was assessed using a Western blot. Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used.

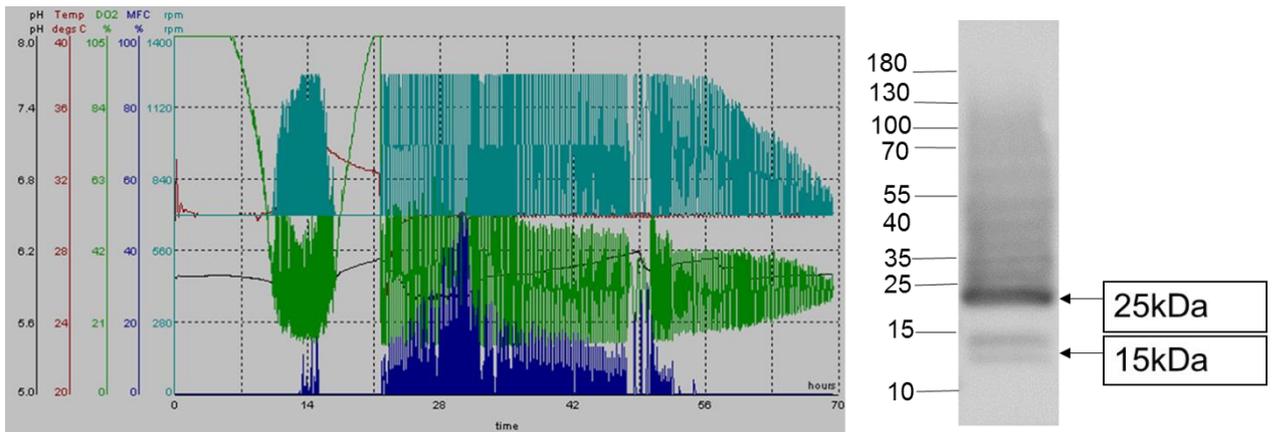
SP-B is indicated by the arrow at 25 and 15 kDa.

### 3.12 Scale up of Highly-Expressing Colony 5

Colony 5 from the 500 µg/ml plate, see above in Figure 3.14, was grown in a bioreactor.

A representative trace and a Western blot of SP-B membrane preparation from the culture is shown below in Figure 3.15.

***The growth trace and membrane isolate reflects the successful expression of SP-B in a yeast system for the first time.***



**Figure 3.15 Large Scale Fermentation of a High SP-B Expressing Yeast Colony**

The growth of the culture was monitored, over 70 hours, with the dissolved oxygen (DO) trace (green).

A crude membrane preparation from the bioreactor run was isolated from the culture and a sample was loaded onto an SDS-PAGE gel to assess protein yield using a Western blot.

Large scale culture favoured the higher 25 kDa SP-B molecular weight. Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used for detection.

SP-B is indicated by the labelled arrows at 15 kDa and 25 kDa.

### 3.13 Solubilisation and Purification of SP-B with SMA 2000P

Having established that SP-B could be produced recombinantly, optimising its extraction was then investigated.

SMA 2000P was used in the first instance, in order to produce a sample that had its protein-lipid interactions intact. These are known as SP-B-SMALP; SMALP is SMA lipoparticle.

SP-B was solubilised with SMA2000P, the solubilised fraction being mixed with Ni-NTA and purified using an imidazole gradient ranging from 100- 600 mM imidazole.

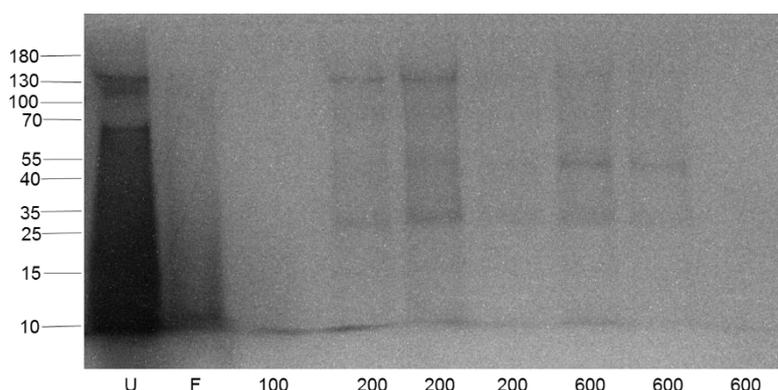
SP-B purification yielded protein band sizes of 25, 40 and 130 kDa. Some SP-B eluted at 200 mM imidazole, but the bulk of the protein was eluted at 600 mM imidazole. Other contaminating proteins were also eluted, see below in Figure 3.16.

As recombinant SP-B is trafficked to the plasma membrane, membrane fractions containing SP-B were isolated.

Figure 3.14 and Figure 3.15 represent membrane fractions containing SP-B.

In Figure 3.16, SP-B was extracted from the membrane using a polymer.

The extraction of SP-B with the SMA2000P polymer and membrane lipids might reflect isolation of the higher oligomeric structures within the SMALP.



**Figure 3.16 Purification of SP-B after Solubilisation with SMA2000P**

SP-B was solubilised using SMA2000P and an imidazole gradient was carried out to determine the optimal concentration of imidazole required to elute SP-B.

Protein bands of sizes 25 kDa, 40 kDa and 130 kDa have been eluted at 200 mM imidazole, whilst 25 kDa and 40 kDa bands have been eluted at 600 mM imidazole. [U=unbound protein, F=flow through (wash)]

Protein bands at molecular weights 25, 40 and 130 kDa were analysed using mass spectrometry.

Band 1 corresponded to the excised 40 kDa band, the mass spectrometry results are shown below in Table 3.4.

Band 1 was identified as cDNA FLJ57475 (highly similar to pulmonary surfactant B). It was the highest scoring peptide at 62.96 as shown below in Table 3.4.

Note: The peptide matches found in the sequence of SP-B, suggest that band 1 is in fact likely to be SP-B. As we now know, the FLJ57475 identification is in fact 100% identical to the amino acid sequence for SP-B (accession number: SP-B: P07988).

Thus it is more than likely there has been a failure of the mass spectrometry facility to update its comparator database.

Recognition as SP-B was subsequently indicated by a separate mass spectrometry expert (Justin MacDonald, Calgary University).

Band 2 corresponded to the excised 25 kDa band. Like band 1, band 2 was also identified as cDNA FLJ57475 (highly similar to pulmonary surfactant B), see Table 3.6 below.

This was identified as the highest scoring peptide at 38.21 (Table 3.4).

**Table 3.4 Identified SP-B Fragments from Mass Spectrometry Analysis**

FPIPLPYCWLCRALIKRIQAMIPK GALAVAVAQVCR VVPLVAGGICQCLAERYSVILLDTLLGRMLPQ LVCRLVLRCSM	
<b>25 kDa SP-B band</b>	<b>40 kDa SP-B band</b>
GALAVAVAQVcR	VVPLVAGGICQcLAER
VVPLVAGGICQcLAER	GALAVAVAQVcR
YSVILLDTLLGR	YSVILLDTLLGR
RIQAmIPK	RIQAMIPK

**Table 3.5 Identified Peptide Fragments from Mass Spectrometry Analysis of SP-B (40kDa)**

Description	Score
1. cDNA FLJ57475, highly similar to Pulmonary surfactant-associated protein B [B4E1F5_HUMAN]	62.96
2. Alcohol dehydrogenase 2 GN=ADH2 [ADH2_KLUMA]	38.02
3. 40S ribosomal protein S1 GN=RPS1 [RS3A_PICPG]	34.54
4. Keratin 1 OS= <i>Homo sapiens</i> GN=KRT1 [H6VRF8_HUMAN]	33.43
5. rRNA biogenesis protein RRP36 GN=RRP36 [RRP36_VANPO]	25.96

Bands thought to be SP-B were excised from a Coomassie stained gel and sent for mass spectrometry analysis at Birmingham University.

The gel plug was trypsin digested and samples were run on a tandem LC-MS.

SP-B had the highest score of 62.96, the next peptide identified was Keratin at 38.02.

**Table 3.6 Identified Peptide Fragments from Mass Spectrometry Analysis of SP-B (25kDa)**

SP-B had the highest score of 38.21.

The peptides identified from the mass spectrometry analysis matched peptides in the sequence of SP-B, again supporting the protein identification

Description	Score
cDNA FLJ57475, highly similar to Pulmonary surfactant-associated protein B [B4E1F5_HUMAN]	38.21
Keratin 1 OS= <i>Homo sapiens</i> GN=KRT1 PE=3 SV=1 - [H6VRF8_HUMAN]	29.71
40S ribosomal protein S5-A GN=rps5a [RS5A_SCHPO]	19.28
40S ribosomal protein S1 GN=RPS1 PE=3 SV=1 - [RS3A_PICPG]	13.93
Keratin, type II cytoskeletal 2 epidermal GN=KRT2 [K22E_HUMAN]	13.63

The peptides identified from the mass spectrometry analysis matched peptides in the sequence of SP-B, and this helped to support SP-B identification.

### 3.14 SP-B-SMALP Secondary Structure Analysis

The purified SP-B-SMALP was analysed using Circular Dichroism (CD) spectroscopy, to examine the secondary structure of recombinant SP-B and check that the protein was indeed folded.

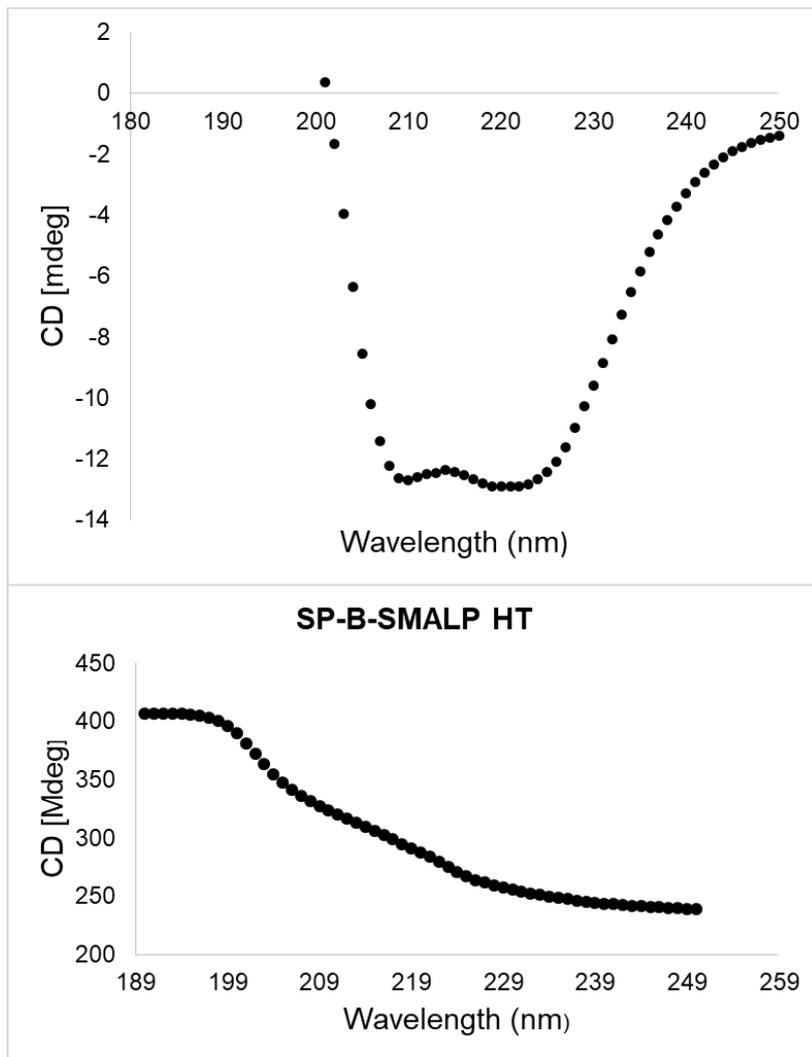
The spectra recorded is shown below in Figure 3.17.

The High Tension (HT) recording in the bottom panel shows that below 200nm the sample absorbed strongly, probably due to the SMA.

This means that CD spectra could not be measured below this wavelength.

However the CD spectra shows strong negative peaks at 208 nm and 222 nm which are indicative of  $\alpha$ -helical structure. This also indicates that the purified SP-B is fully-folded as well as being a predominantly alpha helical protein.

This agrees with its known structure (Andersson *et al.* 1995, Pérez-Gil and Cruz. 1993, Vandebussche *et al.* 1992, Perez-Gil *et al.* 1993).



**Figure 3.17 CD Trace of SP-B-SMALP and Accompanying HT**

SP-B solubilised and purified using SMA 2000P, was analysed using CD.

In the figure above, the secondary structure of SP-B reflects an alpha helical protein conformation. As mentioned in the text above this reflects it's already known structure and so helps confirmation.

Protein concentration 0.1 mg/ml.

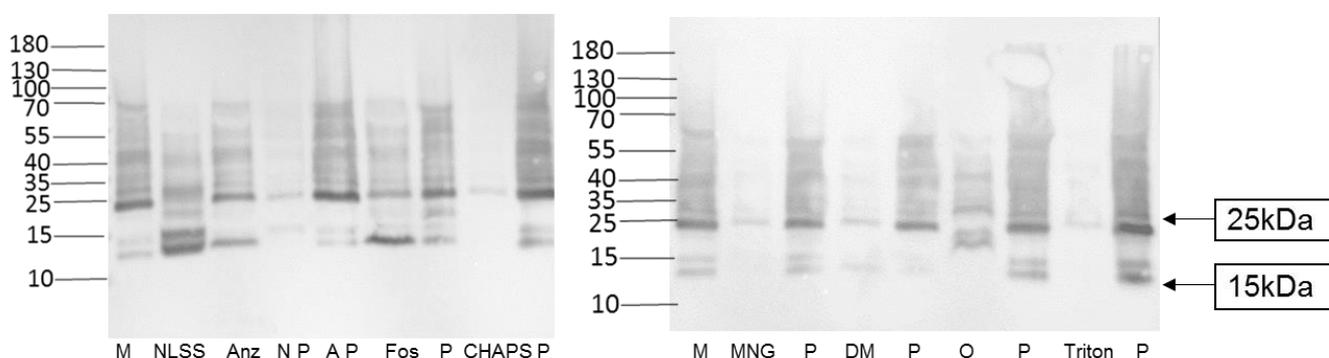
This CD was carried out by Charles Moore-Kelly at Birmingham University. JASCO J-1500 spectropolarimeter used. CD carried out in a 1 mm path length cuvette.

### 3.15 SP-B Detergent Solubilisation Screen

Detergent solubilisation was also separately investigated, because of the potential contributory activity of SMA in subsequent surface activity experiments.

Protein yield was assessed using a Western blot.

The amount of protein in the Anz supernatant was higher than from the other detergents, see below in Figure 3.18.



**Figure 3.18 SP-B Detergent Solubilisation Screen**

SP-B was solubilised with detergents and visualised on a Western blot.

Lane M: membrane prep, NLSS supernatant, Anz supernatant, NLSS pellet, Anz pellet, Fos-12 supernatant and pellet, CHAPS supernatant and pellet, MNG supernatant and pellet, DM supernatant and pellet,  $\beta$ -OG supernatant and pellet, Triton supernatant and pellet.

Bands at 25 kDa and 15 kDa are marked with labelled arrows and indicated to be SP-B.

Anti-His<sub>6</sub> monoclonal antibody and anti-mouse IgG-HRP-linked secondary antibody was used for detection.

This is in line with the 25 kDa and 15 kDa molecular weight observed previously in Figures 3.14, and 3.15. The 40 kDa molecular weight seen above in Figure 3.16 is not present.

The differing molecular weight sizes in this case may be due to the different extraction methods used.

As the SMALP encapsulates both SP-B proteins and lipids without disrupting protein-lipid interactions, it is possible that more than one protein molecule is being captured by the SMALP.

The molecular weight should be closer to 50 kDa if this were the case, but the challenge of not getting uniform unfolding using SDS has already been highlighted previously and may reflect the lower size.

A statistical ANOVA test, followed by a Dunnet's test, was undertaken to help determine the optimal detergent for extraction yield.

The solubilisation of SP-B was repeated and the Western blot image of each gel was pasted into Image J.

Densitometry was carried out on both gels to assess the percentage solubilisation of each detergent solubilisation carried out.

An ANOVA and Dunnet's test were carried out to determine which solubilisation had the greatest extraction.

Solubilisations that were statistically significant included NLSS, Fos-12, Anz and B-OG, see below in Figure 3.19.

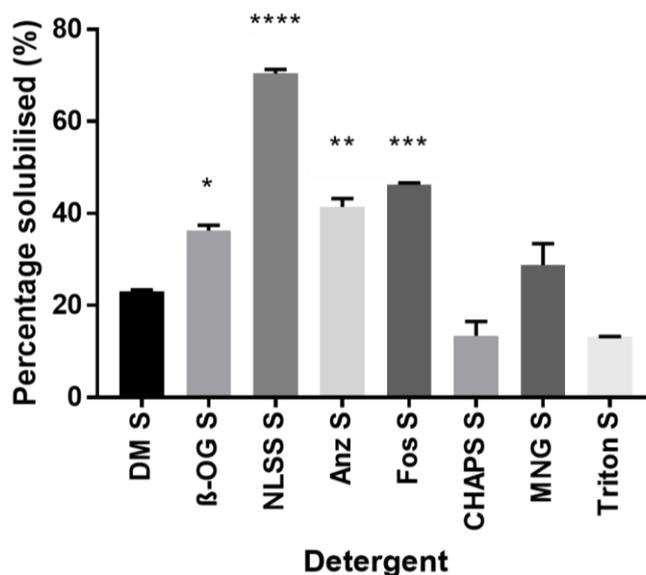


Figure 3.19 Statistical Analysis of SP-B Detergent Solubilisation

An ANOVA and Dunnett's test determined NLSS, Fos-12, Anz and B-OG detergents were statistically significant, these results and are marked with a \*. \*\*\*\*  $P < 0.0001$ , \*\*\*  $P < 0.001$ , \*  $P < 0.05$ .

Table 3.7: Statistical Significance of Different Detergents vs DM using the Dunnetts Test

Dunnett's multiple comparisons test	Significant?	Summary	Adjusted P Value
DM S vs. β-OG S	Yes	*	0.0118
DM S vs. NLSS S	Yes	****	<0.0001
DM S vs. Anz S	Yes	**	0.0016
DM S vs. Fos S	Yes	***	0.0003
DM S vs. CHAPS S	No	ns	0.0573
DM S vs. MNG S	No	ns	0.3526
DM S vs. Triton S	No	ns	0.0522

As a result of the statistical analysis shown above in Figure 3.19, the Anz solubilisation was taken forward into large scale solubilisations as it showed a purer band compared to NLSS, Fos-12 and B-OG.

NLSS was also previously found to cause the nickel from Ni-NTA to leach into the purification elutions.

### **3.16 Purification using Anzergent**

Anzergent was used to solubilise SP-B, above in Figure 3.18.

This solubilisation was then taken forward to large scale purification. A stepped imidazole gradient was designed to determine that optimal concentration which resulted in elution of pure SP-B.

At 50 mM imidazole, a 70 kDa protein band started being eluted and can be seen in fractions up to 200 mM imidazole.

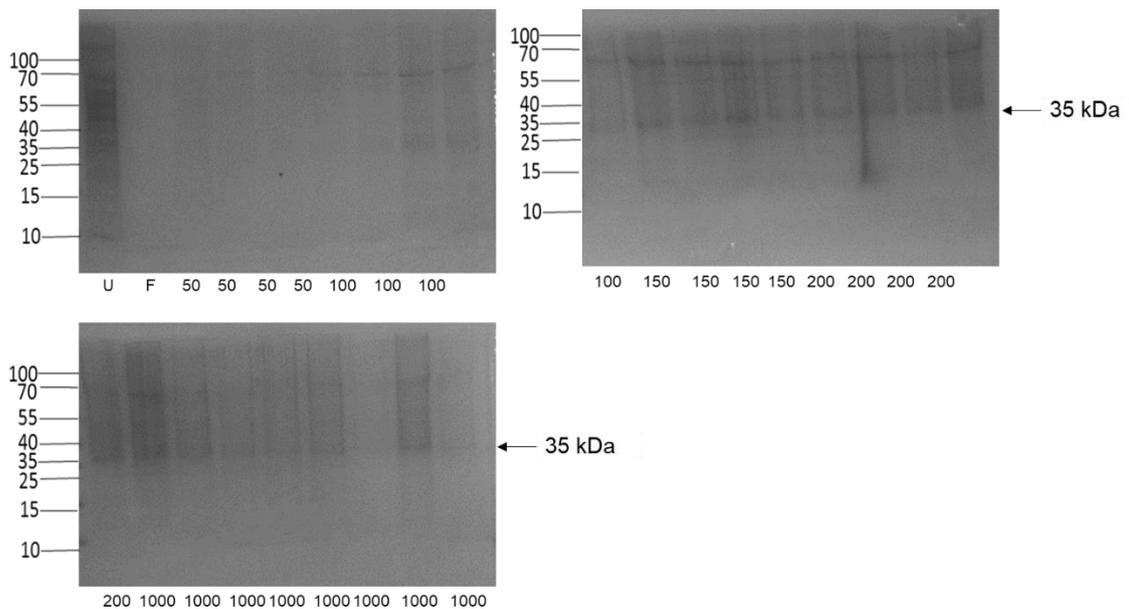
A 35 kDa band also started being eluted at 100 mM imidazole and continued up to 1000 mM imidazole.

A molecular mass of 35 kDa is the same size as previously observed in the SMA 2000P solubilisation, see above in Figure 3.16.

The 35 kDa size is higher than the 25 kDa size that was seen at the small screen purification, see above in Figure 3.18.

The acrylamide used in this gel was made using a different acrylamide/bisacrylamide ratio. This may have possibly resulted in smaller pores being made, and the SP-B protein size may have run to a higher molecular weight as a result.

The yield of purified SP-B attained from 1 L of culture, calculated using densitometry, was 0.4 mg/L.



**Figure 3.20 SP-B Purification using Anzergent**

An imidazole gradient ranging from 50 mM-1000 mM imidazole was used to determine the concentration of imidazole with which to elute.

The purification resulted in protein bands at 35 kDa (marked with an arrow).

Eluted proteins were loaded in an SDS-PAGE and stained using Coomassie dye.

### 3.17 Effect of Recombinant SP-B on Surface Tension

SP-B was solubilised using Anzergent, purified and concentrated.

The sample was analysed using a Langmuir trough.

Langmuir troughs are used to look at the behaviour of molecules as exhibited at the surface of its bulk fluid (sub-phase).

Once purified SP-B molecules were deposited on the surface of the Langmuir sub-phase, computer controlled barriers increased and decreased the total sub-phase surface area. Movement of these barriers provides the expansion and compression phases via which the effect of the SP-B on surface tension is established.

Surface pressure (outward force at the surface) and surface tension (downward force exerted by the molecules) are measured using the Wilhelmy plate.

See sections 1.1.5 and 1.1.6 in Chapter 1 (Introduction) for an explanation of the Langmuir trough and surface tension derivation and determination.

Initial surface pressure data was plotted against surface area as an 'compressive/expansive Isotherm'.

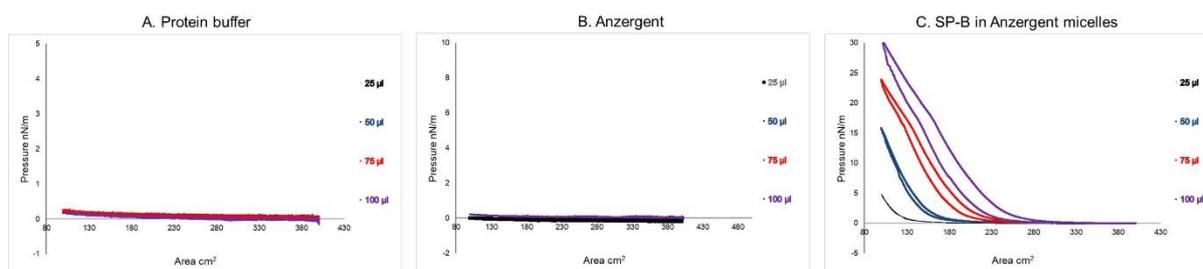
The surface pressure of each 25  $\mu$ l volume increment of both protein buffer, then of the Anzergent (blank) registered effectively 0 mN/m throughout, see below in Figure 3.21.

This indicated that there was no influence of surface pressure from the extracting agent or the buffer.

As a result the isotherm, usually an envelope shape, is flat, see below in Figure 3.21 panels A and B.

Subsequent addition of SP-B in Anzergent micelles to the Langmuir sub-phase resulted in increased pressure with each incremental volume. This effect arises from a higher number of SP-B molecules packing at the surface, so causing a higher surface pressure on the barriers.

This is a good indicator that the recombinant SP-B expressed is surface active, refer below to Figure 3.21 panel-C.



**Figure 3.21 Surface Activity Measurement of SP-B using a Langmuir Trough**

The effect of SP-B on surface tension was measured in a Langmuir trough.

SP-B was suspended in a protein buffer (panel A), the same protein buffer is shown to exhibit no surface activity.

Anzergent micelles also showed no surface activity.

SP-B micelles, added in increasing volumes, were however shown to interact at the surface causing the inflection seen in panel B as the volume is increased.

SP-B at the highest volume showed the greatest inflection (panel C).

Data points from each change in surface area and pressure, at 25 µl increments of protein, added to the available surface area were collected and analysed, see Figure 3.21 above.

Surface pressure was plotted to its maximum level against each incremental protein volume (25 µl).

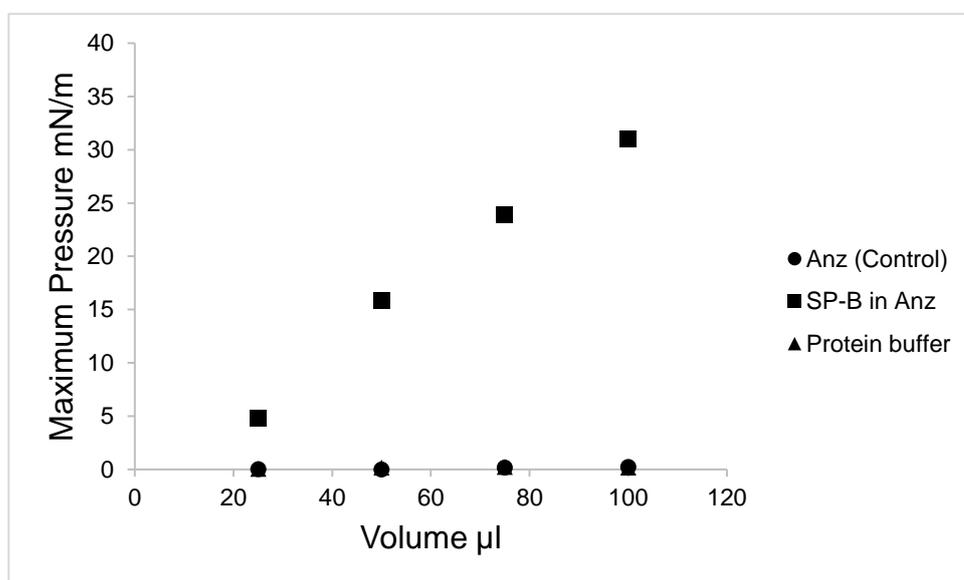
The Anzergent control was composed of buffer and Anzergent (195 mM) only and confirmed the negligible amount of surface activity seen above in Figure 3.21 panels A and B.

Specifically, this is represented as circles on Figure 3.22 below, and as an experimental “blank”, had a surface pressure of 0.0308 mN/m at 25  $\mu$ l, 0 mN/m at 50  $\mu$ l, 0.172 mN/m at 75  $\mu$ l and 0.23 mN/m at 100  $\mu$ l.

SP-B in micelles (16  $\mu$ M), represented by the squares in the curve, seen below in Figure 3.22, had a starting surface pressure of 5 mN/m at 25  $\mu$ l, this increased to 15 mN/m at 50  $\mu$ l, then increased to 24 mN/m at 75  $\mu$ l and increased again at the final volume of 100  $\mu$ l to 30 mN/m.

As aforementioned, the increasing number of SP-B molecules at the surface causes an increased pressure to be exerted on the barriers, reflecting the increase in surface pressure.

This was a good indicator that any surface tension effects observed from SP-B in detergent micelles was indeed essentially from the activity of SP-B itself.



**Figure 3.22** Maximum Surface Pressure of SP-B in Anzergent Micelles Plotted Against Volume

SP-B was extracted using detergent micelles and injected onto the surface of the Langmuir trough.

The maximum surface pressure of the protein buffer (triangles) and Anzergent control (circles) is a surrogate 0 mN/m throughout.

The maximum surface pressure exhibited by SP-B (squares) is low, 5 mN/m at 25  $\mu$ l; but as the volume increased, an increase in maximum surface pressure values is also seen.

## DISCUSSION

It is useful to consider the differences in the behaviour of detergent molecules and SP-B in detergent molecules at the molecular level.

Figure 3.23 (panel 1A) see below, represents the Langmuir surface when the molecules are spread out and before the barrier movements are initiated.

As the barriers decrease the surface area during compression, SP-B molecules were forced to eventually interact, see in Figure 3.23 panel - 1B.

As the number of SP-B molecules were increased, the pressure exerted on the barriers increased at each volume, see Figure 3.23 panel - 1C.

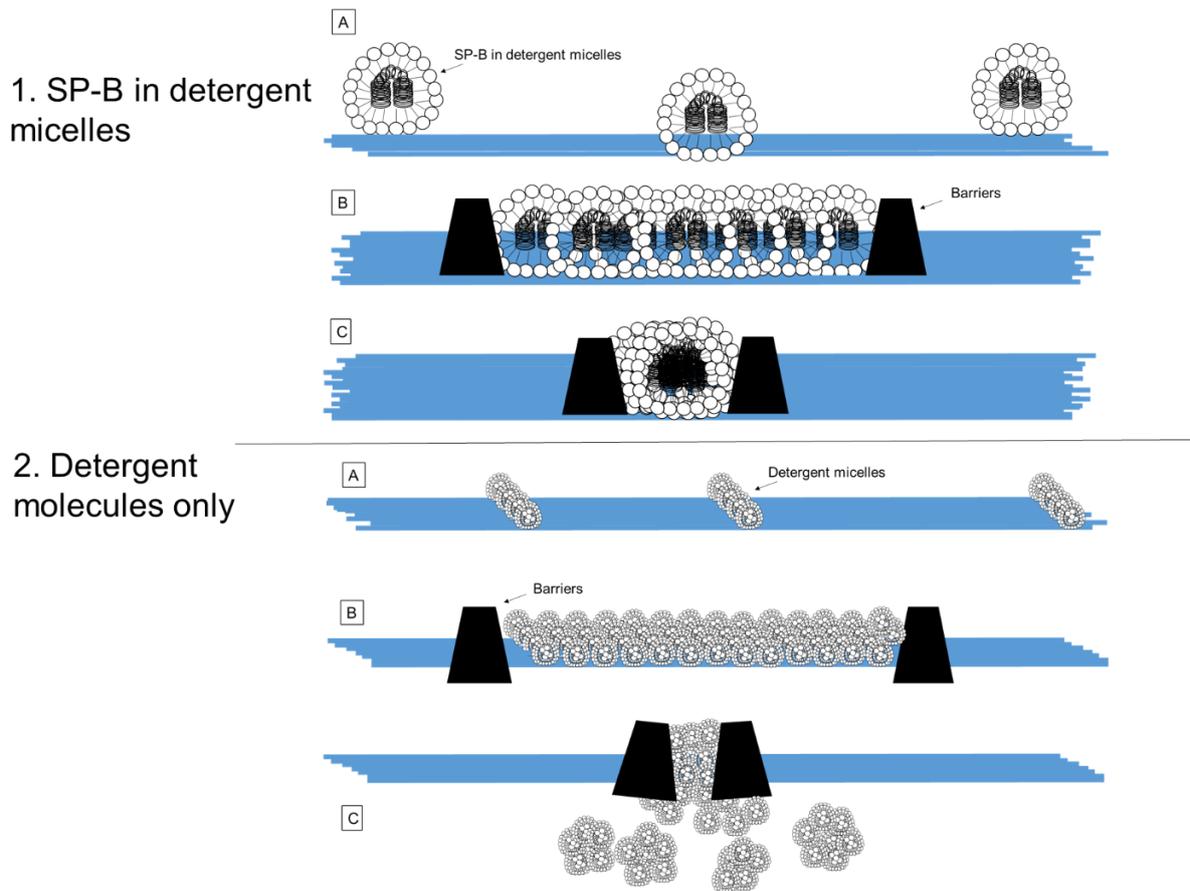
In contrast, the detergent molecules are forced together on compression (Figure 3.23 panel 2B) and are forced under the surface of the water (Figure 3.23 panel 2C).

The detergent molecules themselves are much smaller than SP-B and therefore exerted essentially zero pressure on the barriers.

***This work represents the first surface activity data from recombinant SP-B.***

DISCUSSION continues below, aided by Figure 3.23.

An illustration of the behaviour of SP-B and detergent molecules is shown below in Figure 3.23.



**Figure 3.23 Behaviour of SP-B and Detergents in the Langmuir Trough**

1A) The computer controlled barriers are the furthest away from each other allowing SP-B molecules in detergents to be spread out. B) As the barriers compress, the surface area available to the molecules decreases and the SP-B molecules are forced closer together. C) At the point of maximum pressure the SP-B molecules remain at the surface and exhibit a surface pressure on the barriers.

2A) The behaviour of detergent molecules in A and B are the same as above. C) The detergent molecules interact and are forced under at the point of maximum pressure.

The experimental focus was on SP-B because it is known to be the mature, active form of the protein.

Another step investigated was whether SP-B exhibited surface behaviour in the absence of detergents

In this approach, SP-B was precipitated using an ultracentrifugation device.

Figure 3.24 below demonstrates graphically the maximum pressure vs volume for precipitated SP-B.

The sodium phosphate buffer SP-B, in (triangles) had negligible surface activity.

The surface pressure exerted by SP-B (44  $\mu\text{M}$ ) protein (squares) increased from 0 mN/m at 25  $\mu\text{l}$  to 10 mN/m at 50  $\mu\text{l}$ , then to 15 mN/m at 75  $\mu\text{l}$ , to 18 mN/m at 100  $\mu\text{l}$  and 22 mN/m at 125  $\mu\text{l}$ .

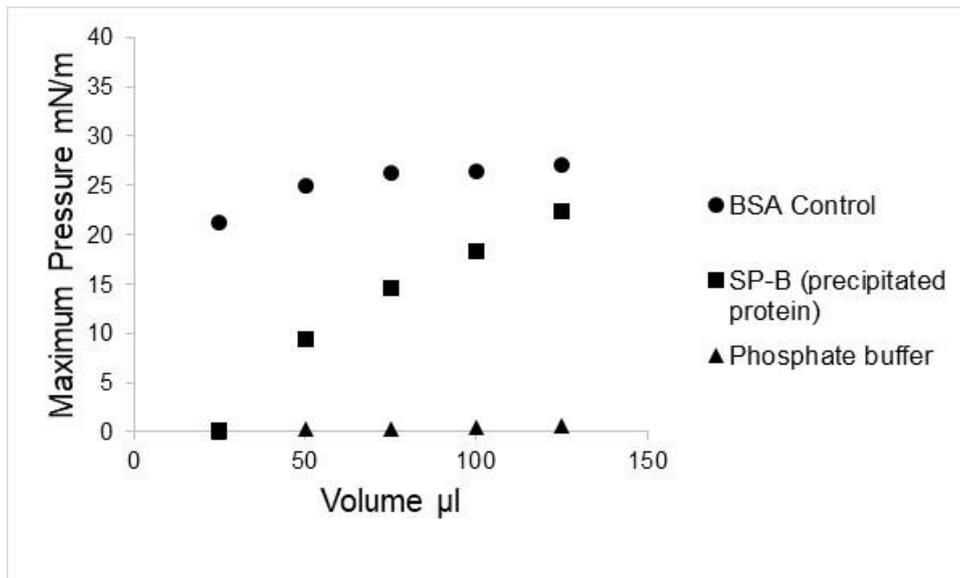
A Bovine Serum Albumin (BSA) was used as a control, this being a protein that was known, by itself, to be devoid of surface activity.

BSA was made to a concentration of 44  $\mu\text{M}$ , and this exhibited a higher surface pressure starting at 20 mN/m (25  $\mu\text{l}$ ), increasing steadily to 27 mN/m (125  $\mu\text{l}$ ).

Note; BSA is a much larger molecule than SP-B, and as a result exhibited a high surface pressure at the initial volume; once the number of molecules is increased, the maximum pressure then quickly plateaued.

The plateau is caused by the rearrangement of BSA molecules such that they are expelled from the surface. Their moving under the molecules at the surface appears similar to the behaviour of molecules in Figure 3.23 panel – 2B.

In contrast, SP-B continues to stay at the surface, indicated by the increasing maximum surface pressure values as the number of molecules increases.



**Figure 3.24 Maximum Surface Pressure of Precipitated SP-B**

SP-B was added in 25 µl increments to the surface of the Langmuir trough.

The maximum surface pressure of SP-B is lower than the surface pressure exhibited by the BSA control.

BSA is represented by circles; precipitated SP-B is represented by squares, and phosphate buffer is represented by triangles.

The SP-B and BSA concentrations were 44 µM.

The maximum surface pressure of precipitated SP-B, above in Figure 3.24, was much lower (0 mN/m) at 25 µl than the BSA (21 mN/m).

The behaviour of the SP-B molecules at the surface, exhibiting higher and higher surface pressures are indicative of the lower surface tension values. Again, this is caused by the 'packing' of SP-B molecules at the surface disrupting surface tension.

As the volume of SP-B is increased, the surface pressure also increased from 0 mN/m to 22 mN/m. Note; The BSA begins at a surface pressure of 20 mN/m because it's a larger molecule. Larger molecules exhibit a higher surface pressure because they take up more surface space in the trough.

***These experiments on their functionality do indeed indicate that recombinant SP-B has surface tension activity.***

In a further approach, SP-B extraction was performed with detergents as previously described, but now the detergent molecules surrounding SP-B were replaced with POPC lipids.

The detergent was adsorbed to biobeads during the reconstitution process.

Again, all efforts were taken to fully remove all the detergent molecules during this reconstitution process.

SP-B in Anzergent micelles had a greater surface pressure than the Anzergent control seen above in Figure 3.22, but the real test of the surface activity of SP-B comes from its ability to also organise associated lipids.

Thus to observe if similar maximum pressure values could be seen when SP-B was surrounded by lipids, a decision to move to reconstitution of SP-B was pursued.

To achieve this, an additional step after purification was added, namely; the detergent molecules around SP-B would be replaced with POPC lipids with the detergents subsequently removed.

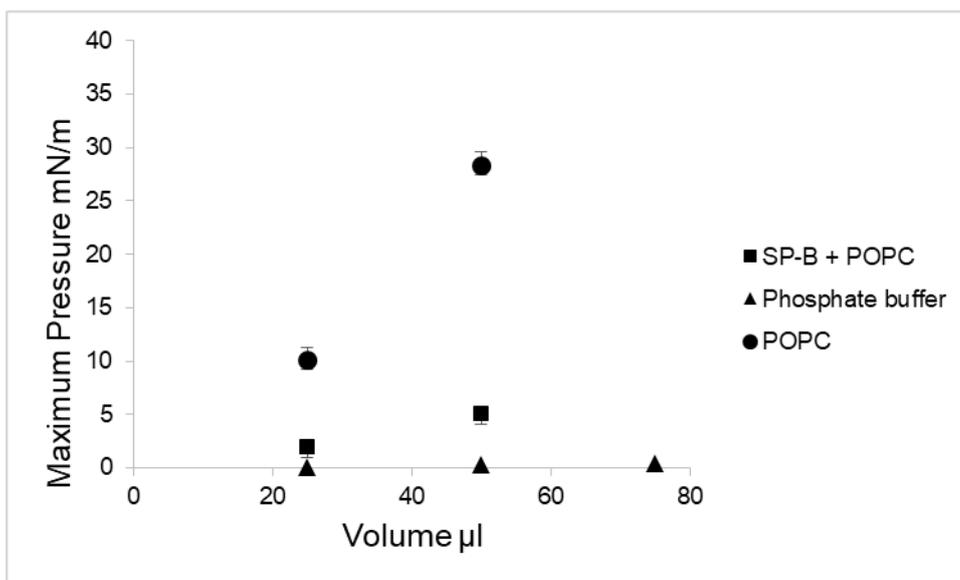
Whilst SP-B is associated with lipids *in vivo*, this experiment formed an *in vitro* model of the physiological state of SP-B.

POPC (2.2 mg/ml) in sodium phosphate, formed the control, and is represented by circles, see below in Figure 3.25.

SP-B + POPC is represented by the squares and phosphate buffer is represented by the triangles.

POPC lipids had a surface pressure of 10 mN/m at 25  $\mu$ l, this increased to 30 mN/m at 50  $\mu$ l. SP-B + POPC showed a much lower surface pressure of 2 mN/m at 25  $\mu$ l and 5 mN/m at 50  $\mu$ l.

No surface activity was again seen from the phosphate buffer



**Figure 3.25 Maximum Pressure of SP-B Reconstituted in POPC Lipids**

SP-B associated with POPC was added in 25 µl increments to the Langmuir trough sub-phase.

SP-B (7 µM) + POPC showed a much lower surface pressure than POPC lipids alone, giving the indication that SP-B was organising the lipids between the surface and bulk sub-phase.

No surface activity was seen from the phosphate buffer.

POPC is represented by circles, SP-B + POPC is represented by squares and phosphate buffer is represented by triangles.

SP-B was reconstituted into POPC lipids at a concentration of 2.2 mg/ml.

***These results confirm the expressed SP-B has the ability to organise lipids.***

In Figure 3.22 and Figure 3.23, see above, SP-B in detergent micelles exhibited an increasing maximum surface pressure.

In contrast, the surface pressure was lowered when SP-B was added with POPC lipids, see above in Figure 3.25.

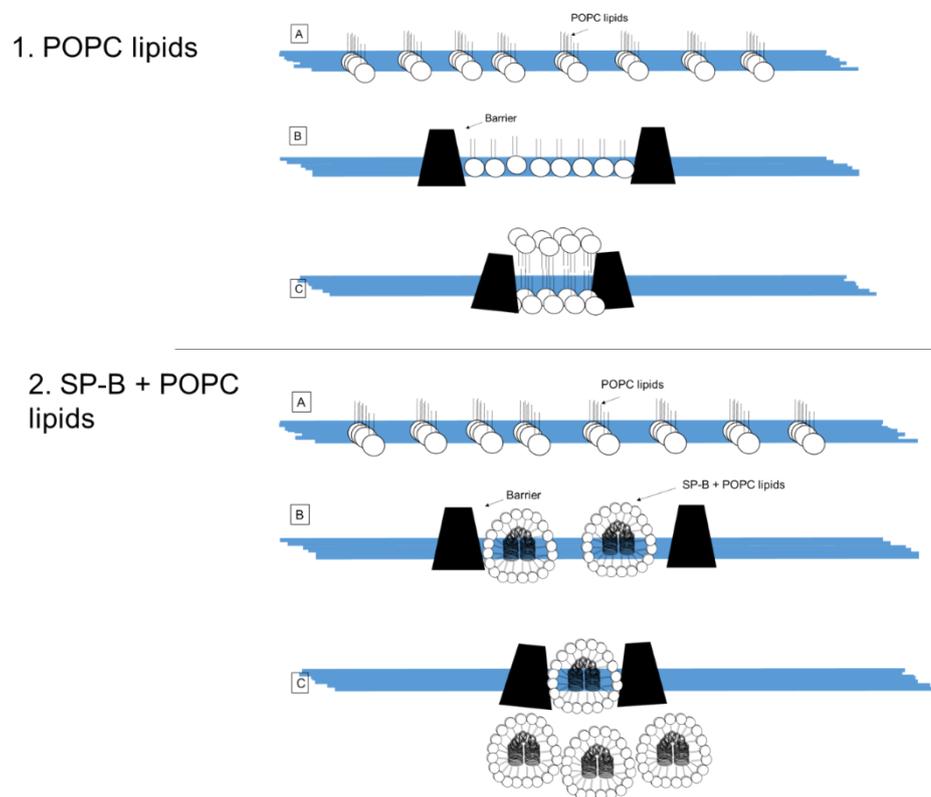
SP-B when added to POPC lipids caused a decrease in maximum surface pressure.

One of the orientations that POPC molecules may have taken is that the fatty acid chains may have formed a bilayer type of structure under compression and so exhibiting a small amount of overall surface pressure, see below in Figure 3.26 panel 1C.

When SP-B and POPC lipids were added together, the surface pressure values decrease at each increased volume compared to the POPC molecules alone.

This may be explained by the 'fusion' of POPC lipids around the surface of SP-B, affected by the ability of SP-B to organise lipids.

The overall effect would be that SP-B surrounded by POPC entered the 'bulk' liquid causing the surface pressure to decrease (Figure 3.26 panel 2C).



**Figure 3.26 Examination of Surface Activity of SP-B in POPC Lipids**

The behaviour of POPC lipids alone on the surface of the trough cause a surface pressure to be exhibited (panels 1B & 1C).

The addition of SP-B and POPC to the surface of the Langmuir causes the lipids to organise around SP-B (panel 2B) so that when the barriers are fully compressed the SP-B and POPC micelles are forced under the surface of the water (panel 2C).

## SP-B Summary

Recombinant SP-B reduced surface tension. This was a result of the surface area taken up by the molecules disrupting the hydrogen bonding between water molecules at the surface. Higher surface pressures were indicative of lower surface tensions.

SP-B on its own, surrounded by detergent micelles and lipids indicated SP-B exhibited an effect on surface pressure and surface tension.

To this end, the most promising results were those of SP-B with POPC lipids.

This indicates that the SP-B produced is indeed functionally active, being also a good indicator that recombinant SP-B had an effect on surface tension.

### **3.18 Conclusions - Recombinant Synthesis of ProSP-B and SP-B**

The genes encoding proSP-B and SP-B were successfully integrated into the *Pichia pastoris* genome.

Extraction of the resultant recombinant proteins was investigated using detergents, SMA 2000P and GH.

The expressed proteins were identified using mass spectrometry; proSP-B (Tables 3.1 and 3.2) and SP-B (Tables 3.3 and 3.4).

The secondary structure of SMALP-extracted and purified SP-B was also analysed by Circular Dichroism (CD) (Charles-Moore-Kelly) (Figure 3.17).

This indicated SP-B is alpha helical, in line with that expected from the literature (Olmeda *et al.* 2015).

Functional physical surface activity experiments using a Langmuir trough, indicated that SP-B had an effect on surface tension. Further experiments in this area would be valuable to more widely define SP-B's surface-active properties.

## Chapter 4

### **Production and Purification of Recombinant ProSP-C and SP-C**

Pulmonary surfactant C is involved in lipid spreading in the lung. The role of SP-C in the tear film is thought to have a similar role to its role in the lung, but its molecular mechanism is less well understood.

The aims for this part of my work were to synthesise proSP-C and SP-C; and to examine their contribution to lipid spreading in the tear film lipid layer.

To synthesise proSP-C and SP-C, the strategy was to integrate proSP-C and SP-C cDNA into a yeast vector. Expressed recombinant proteins were isolated and purified.

SP-C is an alpha helical protein cleaved from a proprotein, proSP-C.

Both the full length proprotein proSP-C and the mature, physiologically active protein SP-C were produced.

There are no crystal structures for SP-C, and this is due to the extremely hydrophobic regions of the protein that have made it difficult to express as a recombinant protein.

The full length construct was synthesised and expressed to overcome the hydrophobicity of the mature regions, so permitting possible future structural studies.

This chapter also demonstrates the cloning strategy used to integrate full length and mature SP-C within the genome of *Pichia pastoris* yeast.

Both detergents and SMA 2000P polymers were used to extract proSP-C and SP-C from the membrane.

Purified proteins were then carried forward into surface activity experiments for functional analyses using a Langmuir trough.

#### 4.1 Design and Construction of ProSP-C and SP-C Expression Constructs

SP-C is synthesised as a pro-protein, proSP-C, which is cleaved to produce the smaller active peptide (SP-C) *in vivo* (Weaver 1998).

Given the challenges posed by hydrophobicity of these proteins, especially SP-C, the full length protein construct ProSP-C could well help to counter the hydrophobicity of the mature regions.

This is achieved by the overall inclusion of more hydrophilic domains with the effect of 'shielding' less hydrophobic domains.

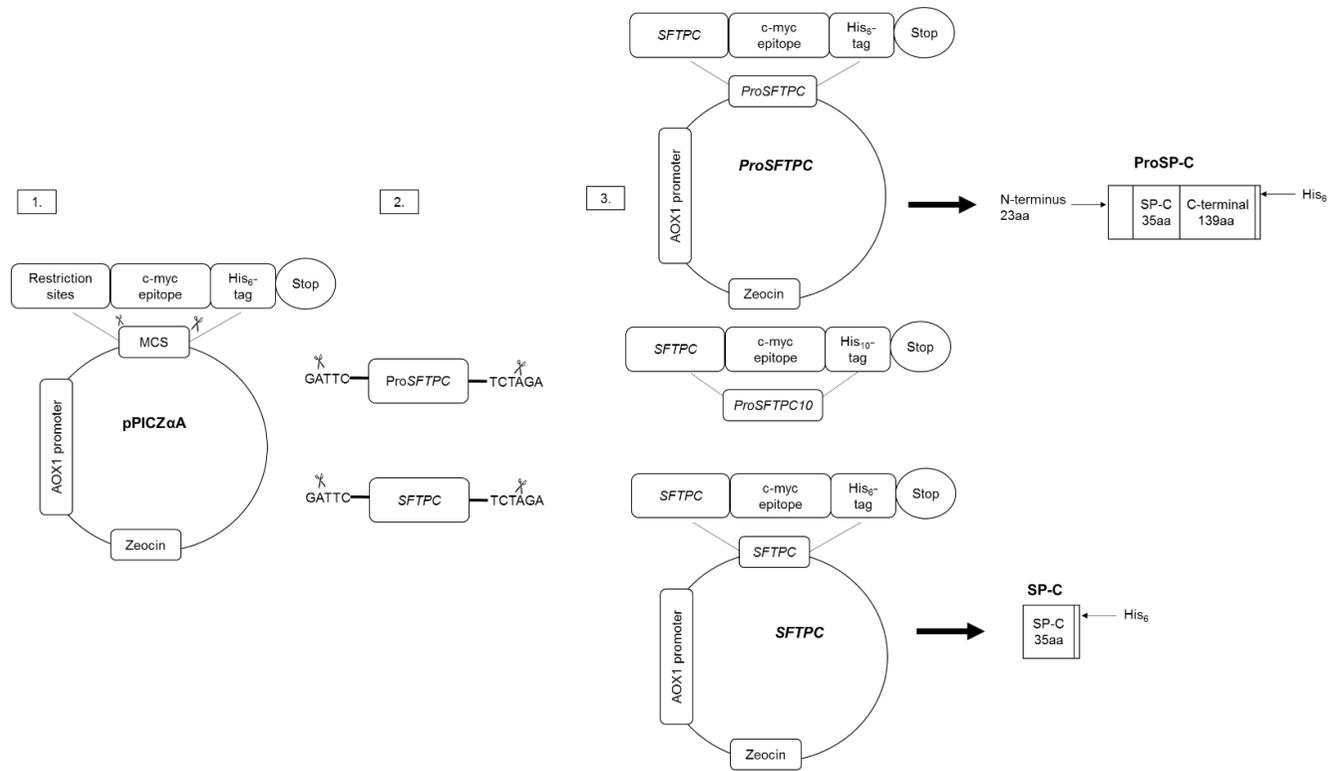
*ProSFTPC* encodes the full length SP-C protein whilst *SFTPC* encodes the shorter, mature active peptide.

"Gentle®" cloning software was used to verify that insertion of the cDNA of SP-C into the yeast vector would keep the His<sub>6</sub>-tag in frame. The vector constructs, and the steps taken to ensure that *ProSFTPC* and *SFTPC* were in frame with the vector His<sub>6</sub>-tag, are illustrated, see below in Figure 4.1.

The predicted sizes of proSP-C and SP-C were calculated using Expasy online software and were found to be 24 kDa and 6 kDa, respectively.

Transcription and translation of *ProSFTPC* leads to the production of recombinant proSP-C which is flanked by a C-terminus and an N-terminus.

These are not present in SP-C, see below in Figure 4.1.



**Figure 4.1** *ProSFTPC* and *SFTPC* Vector Map Design

Sequences for yeast plasmid pPICZαA, and both *ProSFTPC* and *SFTPC*, were digested with *EcoR*1 and *Xba*1 (1 & 2).

They were then ligated (3) using Gentle molecular cloning software to check that the cloning of constructs were in frame with the pPICZαA vector His<sub>6</sub>-tag.

Protein products are also shown: *ProSFTPC* encodes the 24 kDa full length proSP-C protein that has C- and N-terminus flanking regions; *SFTPC* encodes the 6 kDa mature protein.

The vector His<sub>6</sub>-tag was positioned at the C-terminus of the proSP-C and SP-C proteins.

## 4.2 Isolation of a High-Yielding *ProSP-C* Expressing Colony

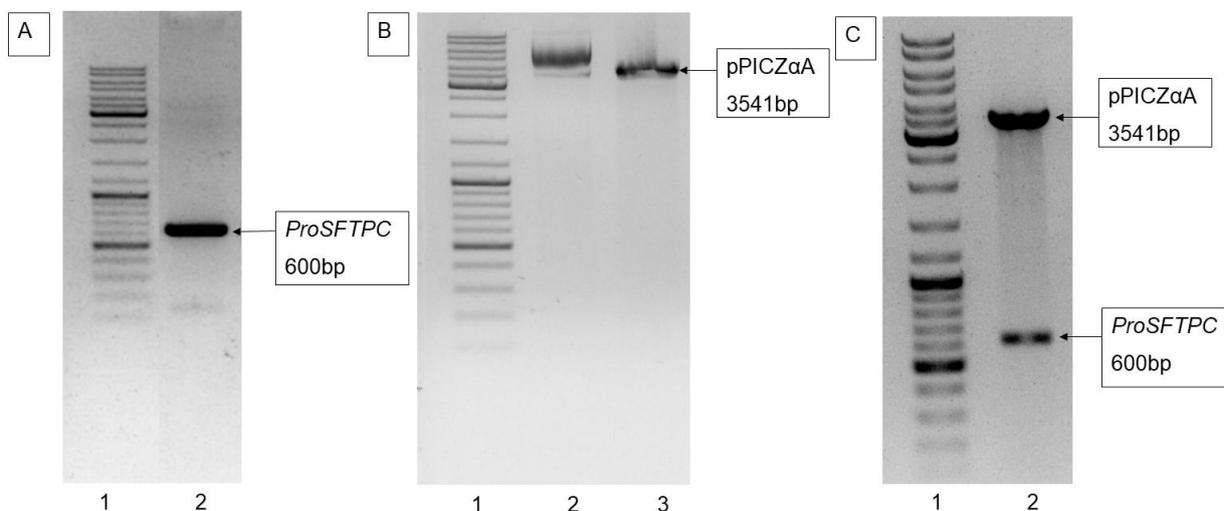
Both *ProSFTPC* and *SFTPC* DNA sequences were amplified in a PCR reaction and digested with *EcoR1* and *Xba1*, see below in Figure 4.2 panel 1.

pPICZαA vector DNA was also digested with *EcoR1* and *Xba1*, run on agarose gels and excised; see below in Figure 4.2 panel B;

*ProSFTPC* and pPICZαA fragments were then ligated together and transformed into XL10-Gold competent cells.

Colonies expressing the transformed *ProSFTPC* vector were cultured in LB and digested.

Digestion of the transformed *ProSFTPC* vector resulted in two bands, see below in lane 2 of Figure 4.2 panel C. The higher band at 3,541 bp is the pPICZαA vector backbone and the lower band at 600 bp is the *ProSFTPC* insert.



**Figure 4.2 *ProSFTPC* Synthesis**

*ProSFTPC* encoding proSP-C was amplified using PCR (panel A).

The PCR fragments and pPICZαA vector were digested with *EcoR1* and *Xba1* (panel B), ligated and transformed into *E.coli* XL10-Gold.

Transformed vectors were digested with *EcoR1* and *Xba1* to reveal a 600 bp *ProSFTPC* insert (panel C, lane 2) and vector backbone (3541 bp, lane 2), lane 1 contains the molecular weight ladder.

The *ProSFTPC* vector was sent for external specialist sequence analysis to “Eurofins” along with the *alcohol oxidase 1 (AOX1)* sequencing primers.

The resultant sequence was translated into its corresponding amino acid sequence using ExPASy online software, see below in Figure 4.3.

The translated amino acid sequence was aligned with the amino acid sequence from the NCBI database, ***with which it had perfect sequence alignment.***

Recom ProSP-C	MRFPSIFTAV LFAASSALAA PVNTTTEDET AQIPAEAVIG YSNLEGFDFV AVLPFSNSTN NGLLFINTTI ASIAAKEEGV SLEKREAEAE FFGIPCCPVH
ProSP-C NCBI	.....
Identity	*****
Recom ProSP-C	LKRLLIVVVV VVLIVVVIVG ALLMGLNLEQ KLISEEDLNS AVDHHHHHH
ProSP-C NCBI	.....
Identity	*****

**Figure 4.3 ProSP-C Sequence Alignment**

The *ProSFTPC* sequence, independently analysed by “Eurofins”, was translated into its corresponding recombinant amino acid sequence (Recom ProSP-C).

This sequence was aligned against the proSP-C NCBI sequence.

The amino acid sequence of recombinant proSP-C is identical (as denoted by \*) to the amino acid sequence of that in the NCBI databank.

*E. coli* transformed with the *ProSFTPC* vector, was cultured, with the DNA extracted for transformation into the yeast *P. pastoris*.

A subsequent expression screen helped to isolate a high yielding colony so as to then ensure maximum recombinant protein production.

The transformation solution was spread onto YPDS plates containing increasing concentrations of zeocin: with 0 µg/ml; 100 µg/ml; 250 µg/ml and 500 µg/ml zeocin plates being used. These concentrations were in the range recommended in the *Pichia pastoris* expression manual (Sunga, Tolstorukov and Cregg 2008).

Colonies were observed on YPDS plates as follows: 250 µg/ml zeocin - (approx. 100 colonies were observed) and 500 µg/ml zeocin - (approx. 50 colonies were observed).

The plates with 100 µg/ml zeocin had greater numbers of colonies compared to plates with higher concentrations of zeocin and were correspondingly more difficult to accurately count.

Some 8 colonies from each plate were grown in BMGY and then induced in BMMY.

The membrane fractions were isolated and loaded into the wells of an SDS-PAGE.

A Western blot was then carried out on the membranes, whence protein expression levels from each of the colonies can be seen below in Figure 4.4.

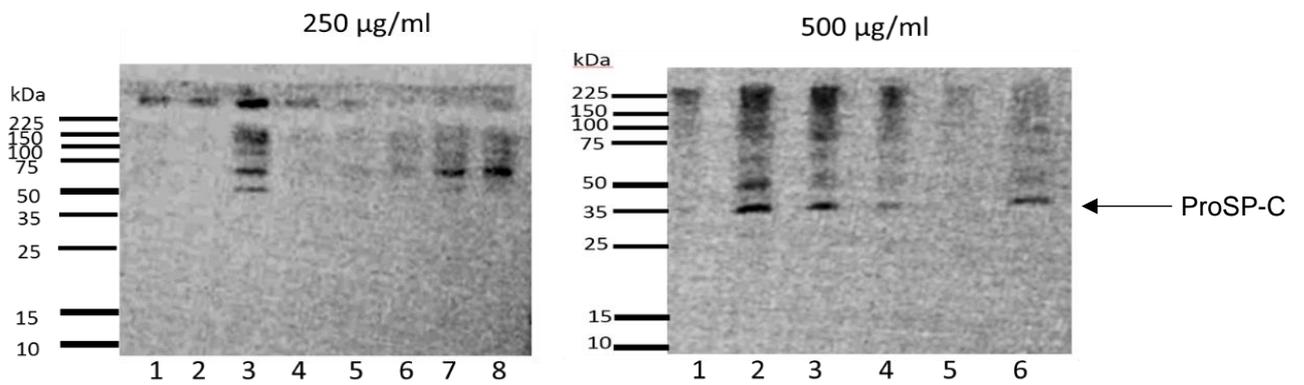
Colony 3 from the 250 µg/ml plate had membrane protein bands at 150 kDa, 100kDa, 75 kDa and 50 kDa.

Colonies 6, 7 and 8 gave protein bands around the 75 kDa size only.

Colonies 2 and 3 from the 500 µg/ml plate had bands at the 35 kDa and 50 kDa size, the same banding pattern can be seen in colonies 3, 4 and 6.

Colony 2 from the 500 µg/ml plate was then chosen for further scale up in a bioreactor because it had the greatest band intensity. The band intensity correlates to protein expression, and this suggested that the best yield of recombinant proSP-C would be produced by this colony compared to the others.

The size of recombinant proSP-C was found to be 35 kDa, which is greater than the 24 kDa predicted size.



**Figure 4.4 ProSP-C Expression Screen**

Colonies from the expression screen were assessed for protein yield.

Proteins in the range 50-225 kDa were observed in the 250 µg/ml plate, these may well reflect higher oligomeric structures.

Proteins from the 500 µg/ml plate were dominated by 35 kDa and 50 kDa protein bands.

The 35 kDa band that can be seen in the 500 µg/ml zeocin plate colony screen was considered to be proSP-C see below in Table 4.1.

Here, a Western blot using an anti-His<sub>6</sub> monoclonal antibody and an anti-mouse HRP-linked secondary antibody was also used for detection.

25 µg total protein was loaded into each well.

### 4.3 Context of Molecular Weight Prediction - Differences and Outcomes

Amongst the many differences between the expected molecular weight (size) and predicted size, one may indicate that the protein has been glycosylated (Teh, Fong, and Mohamed 2011).

Another particular reason for the higher than expected size is that the protein may not be fully unfolded into a linear form by the SDS detergent; due to the formation of 'hairpin loops' (Rath *et al.* 2009).

Proteins that are completely unfolded should travel further than a protein that is partially unfolded because of the "migration" disruption which the hairpin loop(s) pose.

In addition, the surface active nature of proSP-C and any unfolding using SDS detergent may have affected the complete unfolding of the proteins.

Furthermore, the different combinations of extractants and purifications involved in the production of purified proteins is also likely to affect their 'stereo-biology'.

This is why the sequencing conformation, along with other mass spectrometry confirmatory tests, throughout this work are of importance because they confirm the identity of the recombinant proSP-C generated.

#### 4.4 Scale-Up of ProSP-C Expressing Cultures

Colony 2 from the 500 µg/ml plate, see above in Figure 4.4, was also identified as a high-yielding proSP-C expresser. This was based on the fact that it had the highest band intensity.

Scale up with this high-yielding clone was therefore carried out to maximise anticipated proSP-C protein production.

The following details the steps taken to scale up the amount of proSP-C produced using standard conditions routinely deployed

Standard conditions used routinely in the lab for yeast expression in bioreactors were implemented throughout the run.

The dissolved oxygen concentration was set to 30%, pH to 6.0 and temperature to 30°C, whilst “Bioexpert” software ensured the above pre-set values were maintained.

The growth of the proSP-C-expressing culture, see below in Figure 4.5 panel - A is represented by the green dissolved oxygen (DO) trace.

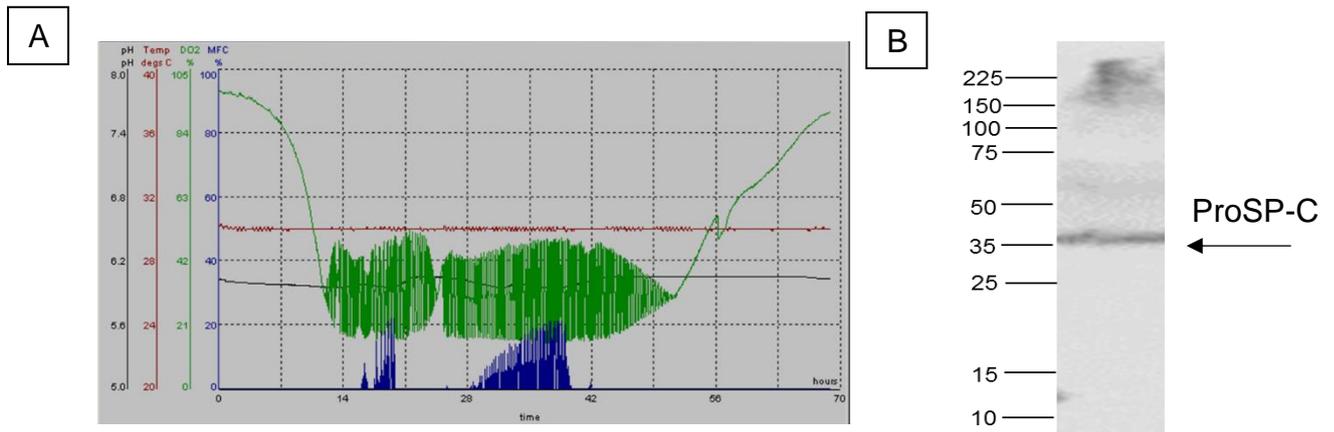
Oscillations around the 30% point indicate the culture is growing well and consuming the oxygen in the medium.

pH is shown by the line in black, and the trace is seen at pH 6.0 consistently throughout the run.

Temperature, shown by the trace in red, indicates the temperature was maintained at 30°C for the duration of the run.

A bioreactor trace representing the growth of the yeast culture expressing proSP-C (Figure 4.5 panel A) and the corresponding proSP-C membrane fraction, extracted from the culture, are shown below in Figure 4.5 panel - B.

A membrane protein band at 35 kDa can be seen; this being larger than the expected size of 24 kDa for proSP-C, but was consistent with the size of the protein purified from the shake-flask culture seen previously in Figure 4.4 above.



**Figure 4.5 Scale-Up of High-Yielding ProSP-C Colony in a 2 Litre Bioreactor**

[A] A representative bioreactor trace representing growth of a culture expressing proSP-C over 70 hours.

Conditions are set to specific values and the “Bioexpert” software adjusts conditions accordingly, ensuring the value is kept within a specific set range.

Growth is monitored by the green dissolved oxygen (DO) trace. pH, shown by the line in black and it is at pH 6.0 consistently throughout the run. The trace in red shows the temperature is kept at 30°C for the duration of the run.

[B] The cell pellet was broken open and the membrane fraction isolated and loaded onto an SDS-PAGE gel.

A Western blot was carried out; an anti-His<sub>6</sub> monoclonal antibody and an anti-mouse IgG-HRP-linked secondary antibody were used to detect the protein.

A single band at 35 kDa can be seen. 60 µg total protein (assessed by BCA assay) was loaded.

***These results indicate that, for the first time in *Pichia pastoris*, proSP-C has been successfully scaled up.***

Extraction methods were next investigated.

## 4.5 Polymer Solubilisation of ProSP-C

ProSP-C, when produced in *P. pastoris*, could be visualised in the membrane fractions using a Western blot, see above in Figure 4.5 panel B. However it is also possible some may have remained in the soluble fractions.

ProSP-C thus needed to be extracted from the membrane and separated from the native *P. pastoris* proteins to isolate it from other membrane proteins.

Styrene maleic acid (SMA) polymers have been recently used to extract other proteins from membranes.

Upon solubilisation, SMA wraps around lipids that include proSP-C and other proteins, encapsulating them into a nanodisc, called an SMA lipoparticle or a 'SMALP'.

Extraction of proSP-C using SMA hadn't been previously investigated and it was hoped the advantages of membrane stability conferred by the SMA could lead to a better understanding of the protein function.

This difference is because detergent solubilisation causes annular lipids to be replaced with detergent molecules; whereas the SMA polymer enables these lipids to be encapsulated alongside the protein.

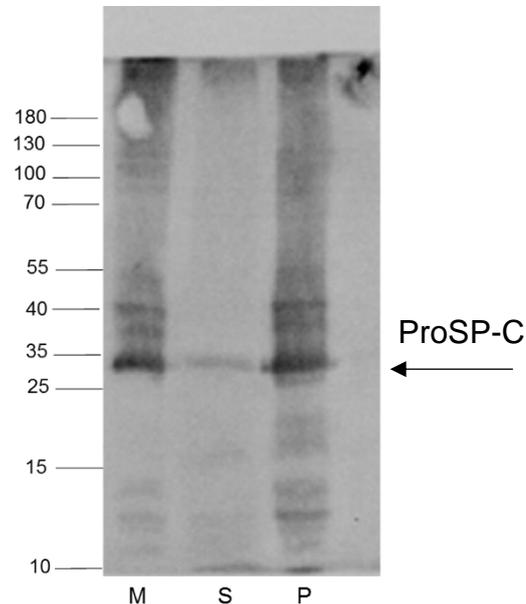
The above difference is why there is a resultant greater stability of protein structure.

There is always a careful balance to be struck between using any extractant to remove proteins from the membrane whilst also ensuring that the remaining protein structure is kept intact.

In the next section, the ability of SMA to effectively extract proSP-C was investigated.

Small-scale solubilisations of proSP-C were carried out, and both the solubilised (supernatant) and unsolubilised fractions (pellet) were analysed using a Western blot.

A faint 35 kDa proSP-C band can be seen in the supernatant (S), see in Figure 4.6 below. This indicated successful solubilisation of a small fraction of proSP-C. The percentage of solubilised proSP-C was found to be 29% when analysed by Image J.



**Figure 4.6 SMALP Extraction of ProSP-C**

ProSP-C was solubilised using SMA 2000P polymer.

The solubilised fraction was separated and loaded into an SDS-PAGE gel and antibody stained in a Western blot.

A 35 kDa band can be seen in the supernatant fraction demonstrating extraction of the proSP-C protein, this protein size correlated with the membrane protein sizes observed previously.

An anti-His<sub>6</sub> monoclonal antibody and an anti-mouse IgG, HRP-linked secondary antibody were used. M: crude membrane protein, S: solubilised supernatant and P: pellet. (15 µg) protein were loaded per lane.

***ProSP-C has been extracted for the first time using the SMA 2000P polymer.***

#### 4.6 Purification of ProSP-C-SMALP

This section describes how solubilised membrane proteins were purified in order to isolate proSP-C from all the other proteins.

Nickel-NTA agarose (Ni-NTA) was used to purify proSP-C; the Ni<sup>2+</sup> on Ni-NTA agarose beads chelated to the His<sub>6</sub>-tag on proSP-C.

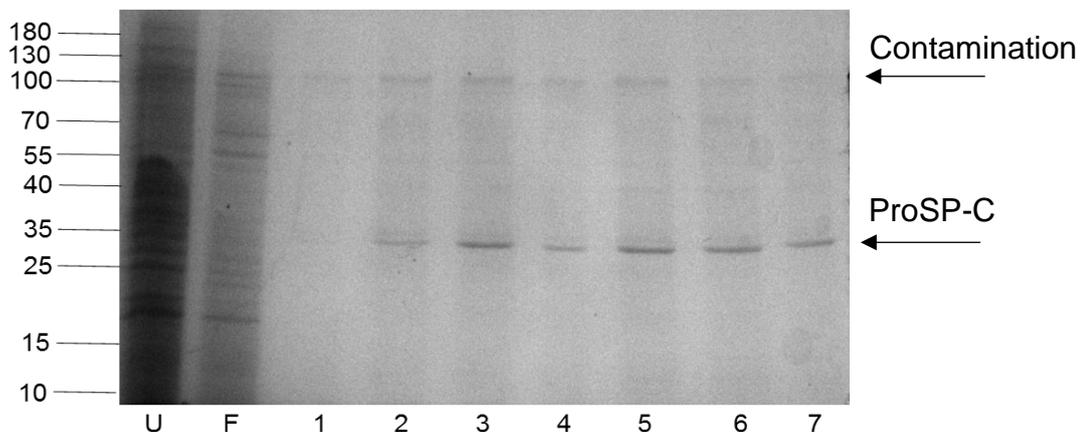
Buffers containing low concentrations of imidazole (20 mM) were used to wash the Ni-NTA to elute weakly-binding contaminating proteins. This was followed by buffers containing higher concentrations of imidazole (e.g. 250 mM), which were used to elute proSP-C.

Collected purification fractions were loaded onto an SDS-PAGE gel and stained with Coomassie, see below in Figure 4.7; U= unbound protein, W= wash and Lanes 1-7 representing elution fractions.

The majority of the proSP-C protein can be seen in the elution fractions.

A 35 kDa band was seen across the lanes and a 100 kDa band can also be seen.

The 100 kDa band was a suspected contaminant because only the 35 kDa band could be seen in the solubilised fraction of proSP-C on the associated Western blot, above in Figure 4.6.



**Figure 4.7 Purification of ProSP-C in SMALP**

ProSP-C was solubilised using SMA2000P and mixed with Ni-NTA and purified. Washes were carried out using 20 mM imidazole and eluted using 250 mM imidazole.

A 100 kDa and 35 kDa band can be seen in the elution fractions. Eluted proteins were stained with Coomassie dye, the 35 kDa size is consistent with previous sizes seen in the expression, while the 100 kDa protein is a contaminant. U: unbound protein, W: wash; Lanes 1-7: elution fractions.

ProSP-C was identified with anti-His<sub>6</sub> antibody.

This is specific for his-tagged proteins, but such binding may not be completely specific because natural his residues are also found within proteins.

The anti-His<sub>6</sub> antibody might be binding to other histidine residues in the protein.

Given previous experience with lack of signal of an antibody for SP-B, identification of purified protein in the case of proSP-C was independently carried out using mass spectrometry.

The 35 kDa band was excised from a Coomassie-stained gel and submitted for mass spectrometry analysis.

The gel plug was trypsin digested and analysed using a tandem LC-MS.

Peptides identified are given a score based on the expected number of peptides found, relative to the number of peptides identified from the database spectra. The higher the score the better the peptide matches between the spectra and the reference database.

Peptides were searched against yeast and human databases.

Band 7 in Table 4.2 below indicated that proSP-C was present with a coverage score of 13.95 and whilst this indicated that proSP-C had a low match between expected and identified peptides from the spectra, fragments of the peptides of proSP-C were more clearly present, see below in Table 4.1.

***Encompassing all the findings above from Figure 4.4 onwards, it was concluded recombinant proSP-C was being expressed.***

**Table 4.1 Identified ProSP-C Peptide Fragments from Mass Spectrometry Analysis**

Identified peptides from mass spectrometry analysis are numbered and highlighted in the amino acid sequence of proSP-C.

ProSP-C sequence: MDVGSKE <b>VLMESPPDYSAAPR</b> GR <b>FGIPCCPVHLK</b> RLLIVVVVVVLIVVVIVGALLMGLHMSQ <b>KHTE</b> <b>MVLEMSIGAPEAQQR</b> LALSEHLVTTATFSIGSTGLVVYDYQQLLIAYKPAPGTCCYIMKIAPE <b>SIPSL</b> EALNRKVHNFQMECSLQAKPAVPTSKLGQAEGRDAGSAPSGGDP <b>AF</b> LGMAVNTLCG EVPLY <b>YI</b>
E <b>VLMESPPDYSAAPR</b>
HTE <b>mVLEmSIGAPEAQQR</b>
E <b>VlmESPPDYSAAPR</b>
<b>FGIPccPVHLK</b>

**Table 4.2 Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-C (35 kDa)**

The fragments identified by mass spectrometry analysis of proSP-C.

Sequencing was carried out at University of Birmingham.

ProSP-C was present with a coverage score of 13.95, indicating a lower number of peptide matches to the overall number of peptides. GN: gene

Description	Score
1. C-type lectin domain family 14 member A GN=CLEC14A [CLC14_HUMAN]	55.58
2. cDNA FLJ50830, highly similar to Serum albumin [B4DPR2_HUMAN]	34.90
3. Alcohol dehydrogenase 2 GN=ADH2 [ADH2_KLUMA]	32.58
4. Alcohol dehydrogenase 1 [ADH1_PICST]	27.78
5. Glyceraldehyde-3-phosphate dehydrogenase GN=GPD [G3P_PICPA]	16.74
6. Actin GN=ACT1 [ACT_PICPG]	14.85
7. Uncharacterized protein GN=SFTPC [E5RG93_HUMAN]	13.95

## 4.7 ProSP-C Detergent Solubilisation

It was important to isolate a pure proSP-C sample for later functional studies; to ensure that any surface activity detected was from a pure proSP-C sample.

Solubilisation and purification of proSP-C with SMA revealed the elution of a contaminating protein, see above in Figure 4.7.

This experiment was repeated twice using an AKTA purification system to remove the contaminant.

However, no proteins bands were detected in any of the elution fractions.

Further, the yield from SMA solubilisation was low, and unable to be detected in any elution fractions from AKTA purification.

In an attempt to improve both the yield and purity of proSP-C, detergent solubilisation and purification were investigated.

This section describes the extraction of proSP-C using conventional detergents, which have a similar amphiphilic structure to phospholipids.

These detergents help extract proteins from the membrane via exchange of lipids for detergents through the formation of micelles.

The first detergent screens of proSP-C were carried out at 22°C.

A final concentration of 5% detergent was used because it had resulted in successful solubilisations in previous experiments.

Prior to producing SP-B and SP-C, in order to develop experimental techniques the author of this thesis had carried out detergent screens using a different membrane protein.

The protein in question, Aquaporin 1, was used because it was appropriate and readily available.

During this particular exercise, the author experimented with a 1% detergent concentration (which failed to solubilise the Aquaporin 1), before then raising the detergent concentration to 5%; at which the solubilisations were successful.

The detergents selected were chosen because they were a mix of ionic and zwitterionic detergents.

These detergents were: N-lauroylsarcosine sodium salt (NLSS), Anzergent (Anz), Foscholine-12 (Fos), 3-((3-cholamidopropyl) dimethylammonio)-1-propanesulfonate (CHAPS), Lauryl maltose neopentyl glycol (L-MNG), n-octyl- $\beta$ -D-glucopyranoside ( $\beta$ -OG), n-dodecyl  $\beta$ -D-maltoside (DDM) and Triton-X-100.

Solubilised proSP-C was recovered in the supernatant via centrifugation, whilst unsolubilised protein was found in the pellets.

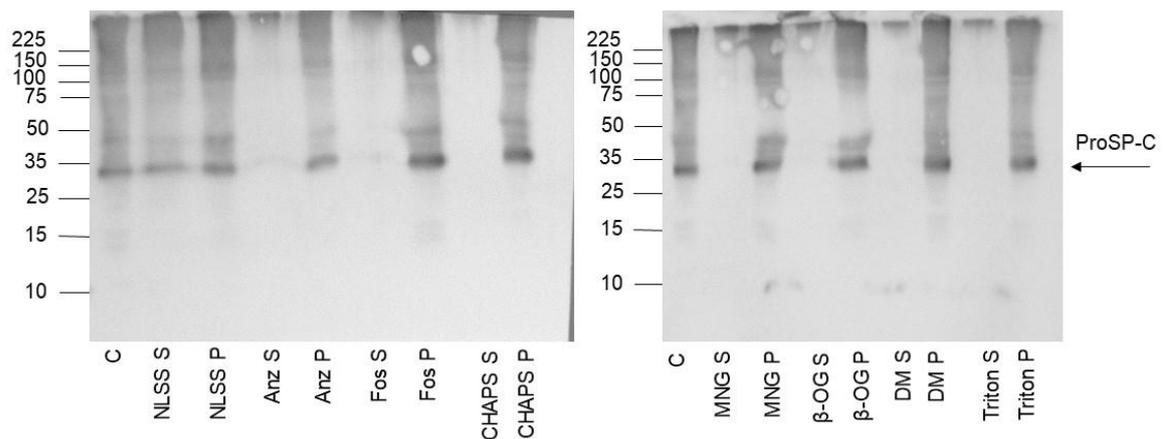
These products were run on an SDS-PAGE and a Western blot, see below in Figure 4.8.

Solubilised protein was recovered in the NLSS supernatant only.

Separate supernatants from the other detergent samples appeared to have aggregated in the loading wells of the SDS PAGE gel.

Very faint bands were also seen in the Fos and Anz supernatants.

For most of the detergent samples, a 35 kDa band can be seen throughout the pellet fractions.



**Figure 4.8 ProSP-C Detergent Screen**

The detergents used in the screen were NLSS, Anz, Fos, CHAPS, MNG,  $\beta$ -OG, DM and Triton. Successful solubilisations can be seen in the lanes (S) and were visualised using a Western blot.

The Western blot shows the amount of proSP-C present in both the solubilised fraction (S) and the unsolubilised fraction (P).

ProSP-C is present in the supernatant of NLSS, but has aggregated, in the presence of the other detergents and appeared at the bottom of the loading well.

120  $\mu$ g total protein was loaded per lane.

#### 4.8 Purification of Detergent-Solubilised ProSP-C

The detergent solubilisation screen indicated that NLSS was the best detergent for extracting proSP-C.

NLSS was therefore carried forward into large scale solubilisations and purifications.

The next step involved separation of proSP-C from all the other solubilised membrane proteins, with further purification using Ni-NTA or cobalt ( $\text{Co}^{2+}$ ) ions being examined.

The first purification used immobilized Ni-NTA in a pre-packed His trap column hooked up to an AKTA FPLC system.

This was directly compared to a purification using Ni-NTA agarose beads.

ProSP-C supernatant which had been solubilised with NLSS, was mixed with Ni-NTA agarose beads, washed and then eluted using a stepped gradient of samples that ranged from 50 mM to 500 mM imidazole.

A stepped imidazole gradient enables contaminating proteins to be eluted at the lower concentrations, thus allowing the protein of interest to be eluted at a higher concentration.

Low concentrations of proSP-C (35 kDa) were eluted at 100 mM imidazole, see below in Figure 4.9.

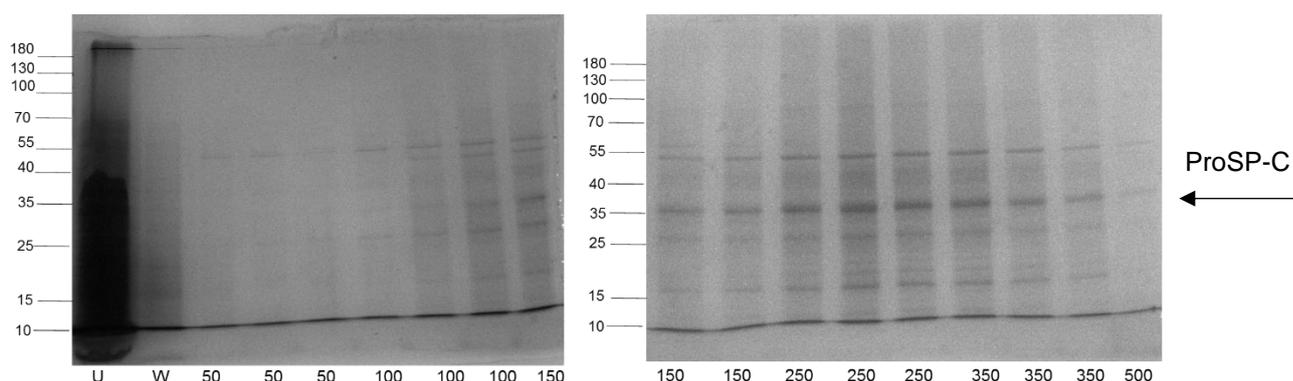
As the concentration of imidazole was increased, the intensity of the proSP-C band also increased up to 500 mM imidazole.

ProSP-C elution appears strongest between 250-350 mM imidazole.

Other contaminating bands at 55 kDa, 25 kDa and 15 kDa were also eluted as the concentration of imidazole was increased.

Imidazole gradients usually help to identify the optimal concentration of imidazole to elute at, but that isn't observed here.

Thus a decision was made to seek to purify proSP-C using an AKTA FPLC to try and improve the purification quality.



**Figure 4.9 ProSP-C Purification Visualised on a Coomassie Stained Gel**

ProSP-C was solubilised and purified using NLSS.

A stepped imidazole concentration in the range 50-500 mM was used to elute proSP-C protein. The 35 kDa band that represents proSP-C, indicated by the arrow, starts being eluted faintly at 100 mM and continues being eluted to 500 mM imidazole.

Other contaminating bands are also eluted at 55 kDa, 25 kDa, and 15 kDa throughout the gradient. U=unbound protein, W=wash.

The attempted separation of elution fractions from solubilisation, and specific purification of proSP-C, resulted in other contaminating proteins being eluted, see above in Figure 4.9.

To improve the purification purity, the experiment was now repeated also using a Ni-NTA His-trap column, and purified using an AKTApurifier 10 system.

His-tagged proSP-C interacted with immobilized Ni-NTA as it flowed through the column.

The column was connected to an AKTApurifier 10 system, with the purification carried out at a flow rate of 1 ml/min.

A wash step was carried out, as previously described, and the imidazole concentration that each fraction elutes at was refined.

The gradient was run from 20 mM to 500 mM imidazole.

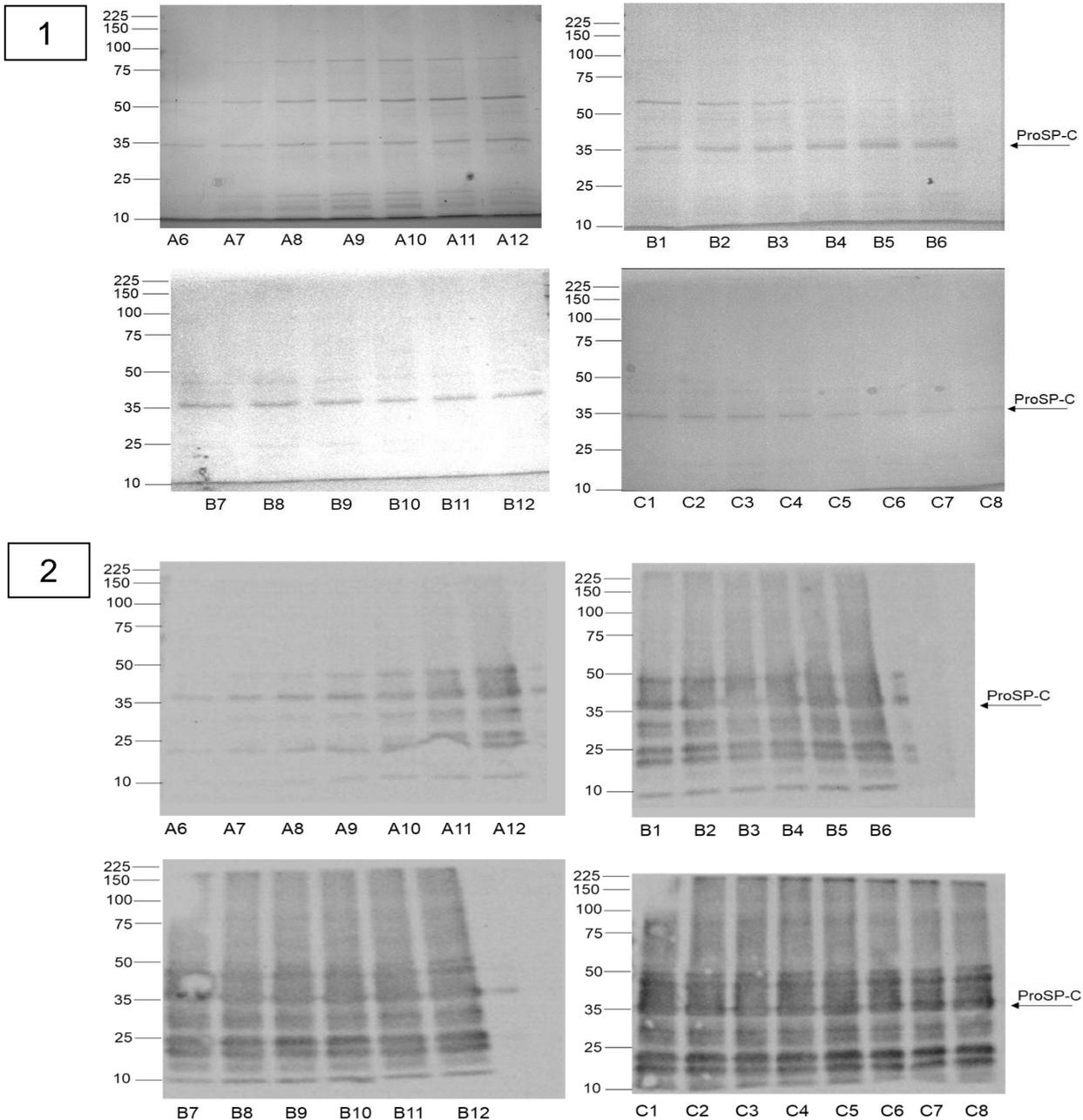
Collected elution fractions were visualised using Coomassie stained SDS-PAGE gels, see below in Figure 4.10 panel -1 and Western blots in Figure 4.10 panel - 2.

Fractions A6-12 correlate to an imidazole concentration ranging from 20 to 100 mM. Fractions B1-B6 represent 150-250 mM imidazole, B7-B12 represent 300-400 mM imidazole and C1-C8 represent 400-500 mM imidazole.

The 35 kDa proSP-C band increased in intensity as the imidazole concentration was increased from fractions A6-C8, again seen below in Figure 4.10 panel -1.

A contaminating protein band at 50 kDa could however be seen throughout fractions A6-B12 whilst a 15 kDa band was also seen throughout fractions A6-C3.

The corresponding Western blots, seen below in Figure 4.10, reveal a large range of contaminating proteins that have been eluted with proSP-C as the concentration of imidazole was increased.



**Figure 4.10 Purification of ProSP-C using an AKTApurifier 10 System**

ProSP-C eluted at 20 mM imidazole, and continued to be eluted as the imidazole concentration was increased.

Fractions A6-A12 represent imidazole concentrations of 20-100 mM. Fractions B1-B6 represent imidazole concentrations of 15-250 mM. B7-B12 represent imidazole concentrations of 300-400mM and C1-C8 represent imidazole concentrations of 400-500 mM. Panel 1: A6-C8 are all Coomassie stained SDS PAGE gels, panel 2: A6-C8 represent the corresponding Western blots.

The Western blots reveal a multitude of contaminating proteins that have been eluted as the concentration of imidazole was increased. ProSP-C bands are indicated by the labelled arrows.

Solubilisation and purification of proSP-C indicated it was being eluted at a low concentration of imidazole, see above Figure 4.10 panel -1 A6-A12.

This may have been due to weak binding of the His<sub>6</sub> tag to the Ni-NTA resin.

The use of cobalt (Co<sup>2+</sup>)-NTA was then investigated because the purity of some proteins eluted from this type of metal ion has been found to be high (Björner, Lundqvist and Hedlund 2012).

The purification was carried out using a peristaltic pump in order to pump proSP-C and buffers through the column at a flow rate of 0.5 ml/min.

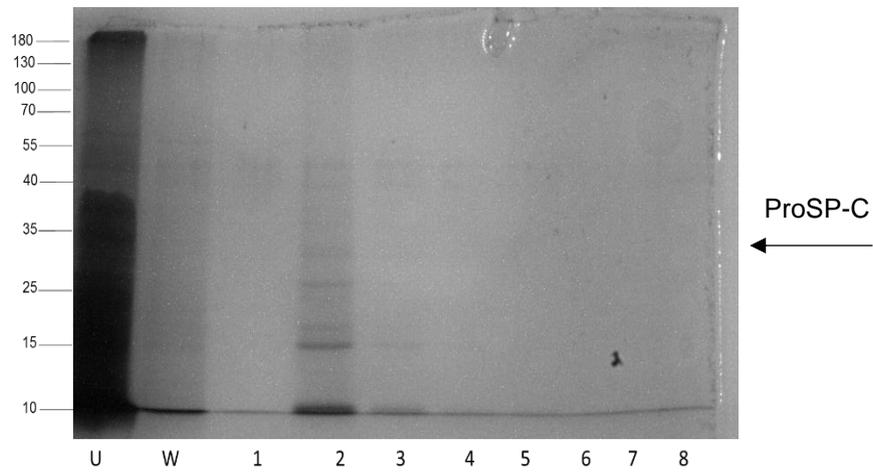
This preliminarily “screening” test would determine whether switching to Co<sup>2+</sup> purification would yield purer proSP-C protein.

Sodium phosphate buffers were used to wash and elute according to protocol guidelines in (Björner *et al.* 2012).

Proteins were eluted at 150 mM imidazole and examination of the purification fractions reveals a 25 kDa and 15 kDa band.

Whilst the 25 kDa band was closer to the expected size of proSP-C, but a number of other bands were also co-eluted, see below in Figure 4.11, the intensity of contaminating bands suggests that this would be a poor purification process to carry forward.

As the quality of the overall purification was poor, it was therefore abandoned as a purification option.



**Figure 4.11 ProSP-C Purification using  $\text{Co}^{2+}$ -NTA**

ProSP-C was solubilised using NLSS, and eluted using 150 mM imidazole.

Protein bands were stained with Coomassie and can be seen just under 55 kDa, 25 kDa, 15 kDa and 10 kDa.

This demonstrated that the overall purification quality using  $\text{Co}^{2+}$ -NTA was poor.

This was therefore abandoned as a purification method.

U=unbound; W=wash; Lanes 1-8: elution fractions.

## 4.9 *ProSFTPC10* Synthesis

The task of trying to obtain pure fractions of proSP-C had proved particularly challenging, but it was important to continue trying to obtain highly pure samples of proSP-C.

This was to reduce/eliminate any effect of contaminating proteins in later investigation of surface activity.

Due to problems encountered in preparing highly-pure fractions of proSP-C, it was hypothesised that the construct would benefit from a longer polyhistidine-tag.

A His<sub>10</sub>-tag at the C-terminus of proSP-C was chosen because it was thought it might help the proSP-C protein bind more tightly to Ni-NTA and so would also enable more stringent washing.

This was anticipated to successfully remove the contaminating bands, and so would help improve the purity of proSP-C samples.

The new construct was identified as *ProSFTPC10* and translated to give proSP-C10.

*In silico* cloning of *ProSFTPC10* was carried out as previously described, the vector map for *ProSFTPC10* is illustrated, see below in Figure 4.12.

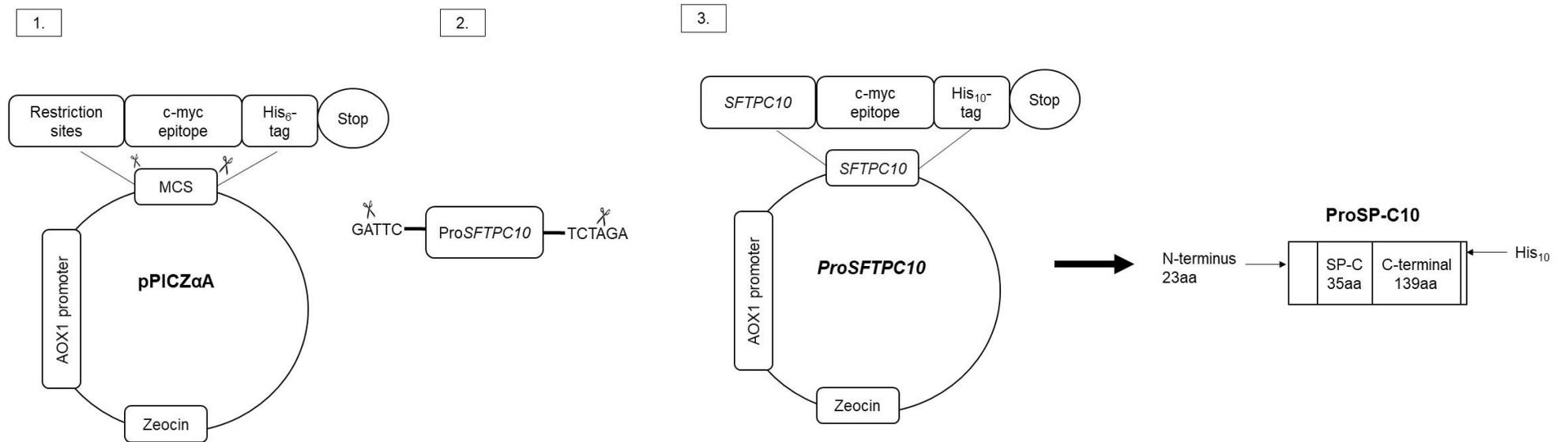
Using ExPASy, the protein product was predicted to be at 24.5 kDa.

*Eco*R1 and *Xba*1 were used to digest both the *ProSFTPC10* DNA and the pPICZαA vector.

The insert and vector fragments were ligated and transformed into *E.coli* XL10 Gold cells.

To confirm that the His<sub>10</sub> tagged construct was successfully cloned, it was digested again with *Eco*R1 and *Xba*1.

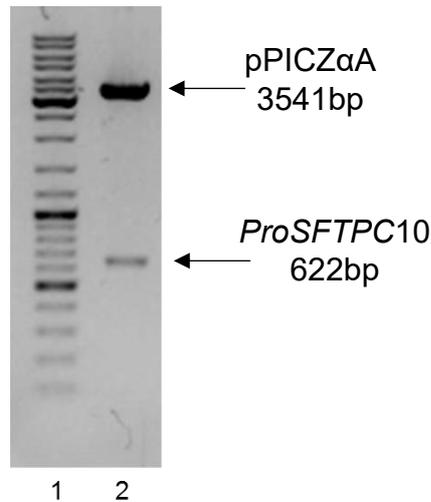
Two bands could be seen: *ProSFTPC10* fragment (622bp) and the vector backbone (3541bp), again, see below in Figure 4.13 lane 2.



**Figure 4.12 ProSFTPC10 Vector Map Design**

*ProSFTPC10* and pPICZαA were digested with *EcoR1* and *Xba1* (1 & 2) and ligated into the multiple cloning site (MCS) (3).

*ProSFTPC10* encoded the 24.5 kDa full length proSP-C10 protein.



**Figure 4.13** *ProSFTPC10* synthesis

Restriction digestion of transformed *ProSFTPC10* revealed an insert of 622 bp (*proSFTPC10*) and a pPICZαA vector backbone of 3541 bp in lane 2.

Lane 1: Molecular weight ladder.

The transformed vector was linearised and transformed into yeast strain X-33.

Transformation into X-33 and the subsequent expression screen were carried out as described previously.

Figure 4.14, see below, illustrates the results of the expression screen to identify a high-yielding colony.

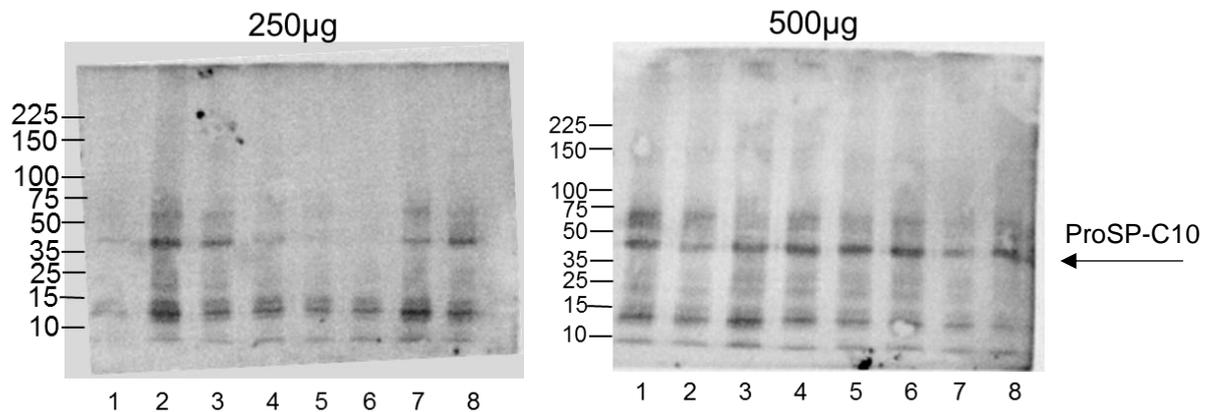
In both blots, protein bands at 50 kDa, 40 kDa, 15 kDa and 10 kDa can be seen.

The expected size of proSP-C10 was 24.5 kDa but the protein bands seen were either rather higher at 40 kDa and 50 kDa or much lower at 15 kDa.

ProSP-C10 was identified to be the band at 40 kDa, later confirmed by mass spectrometry, see below in Tables 4.3 and 4.4.

The other bands may be breakdown products or other non-specific binding products.

The 40 kDa band is present in all the colonies grown on the 500 µg/ml YPDS plates. In contrast, it is present in only four of the colonies from the 250 µg/ml plates



**Figure 4.14** Transformation of *ProSFTPC10* Vector into X-33

Transformed colonies were spread onto YPDS plates containing 250 µg/ml and 500 µg/ml zeocin concentrations.

Membrane fractions from expressed colonies were isolated and run on an SDS-PAGE gel and then visualised using a Western blot.

There are protein bands at 15 kDa and 40 kDa in all the expressed colonies; those from the 500 µg/ml plate had a darker intensity, particularly the 40 kDa band.

An anti-His<sub>6</sub> monoclonal antibody and anti-mouse IgG-HRP-linked secondary antibody were used.

ProSP-C10 is indicated by the labelled arrow above.

ProSP-C10 was confirmed (mass spectrometry), to be the 40 kDa protein, see Tables 4.3 and 4.4.

25 µg protein was loaded into each well.

#### 4.10 Solubilisation and Purification of ProSP-C10 using SMA2000P

Colonies 4, 5 and 6 from the 500 µg YPDS plate, seen above in Figure 4.14, had the highest band intensities at the 40 kDa band.

Colony 6 was chosen for scale up in a bioreactor because it appeared to have this greater intensity compared to the others.

*ProSFTPC10*-expressing yeast was cultured in a bioreactor and the membrane fractions isolated and solubilised with SMA2000P.

The supernatant was then purified using Ni-NTA agarose beads.

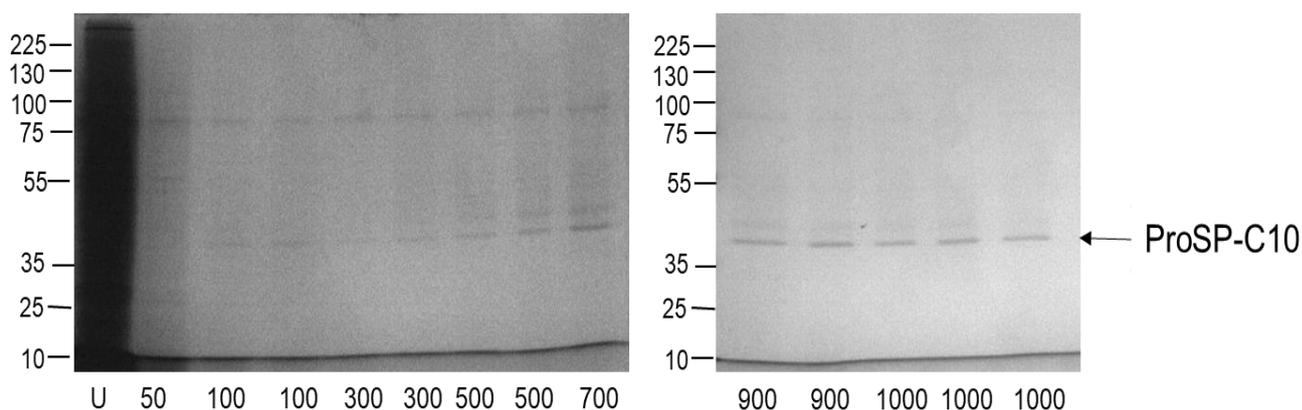
ProSP-C10 was eluted using a stepped imidazole gradient (50 mM-1000 mM), shown below in Figure 4.15.

A contaminating band at the 75 kDa molecular weight started being eluted at 50 mM and is seen in each elution fraction as the imidazole concentration is increased.

The contaminating band was faintest at an imidazole concentration of 900 mM.

ProSP-C10 was faintly eluted at an imidazole concentration of 100 mM; the band increased in intensity as the concentration of imidazole was increased to 1000 mM.

Note: The molecular weight of proSP-C10 is slightly lower (37 kDa) than previously observed (40 kDa), but compares with 35 kDa from other production routes.



**Figure 4.15 Purification of ProSP-C10-SMALP**

ProSP-C10 membranes were solubilised using SMA 2000P.

ProSP-C10 was eluted using a stepped imidazole gradient of 50 mM-1000 mM in 1ml fractions.

ProSP-C10 was eluted faintly at 100 mM and increased in intensity as the stepped imidazole gradient increased to 1000 mM.

As the concentration of imidazole was increased, the contaminating proteins (100 kDa) were eluted. This is particularly evident after 900 mM, where very pure fractions of proSP-C10 can be seen.

Addition of the longer His<sub>10</sub>-tag, combined with a step-wise imidazole gradient greatly improved proSP-C10 purity.

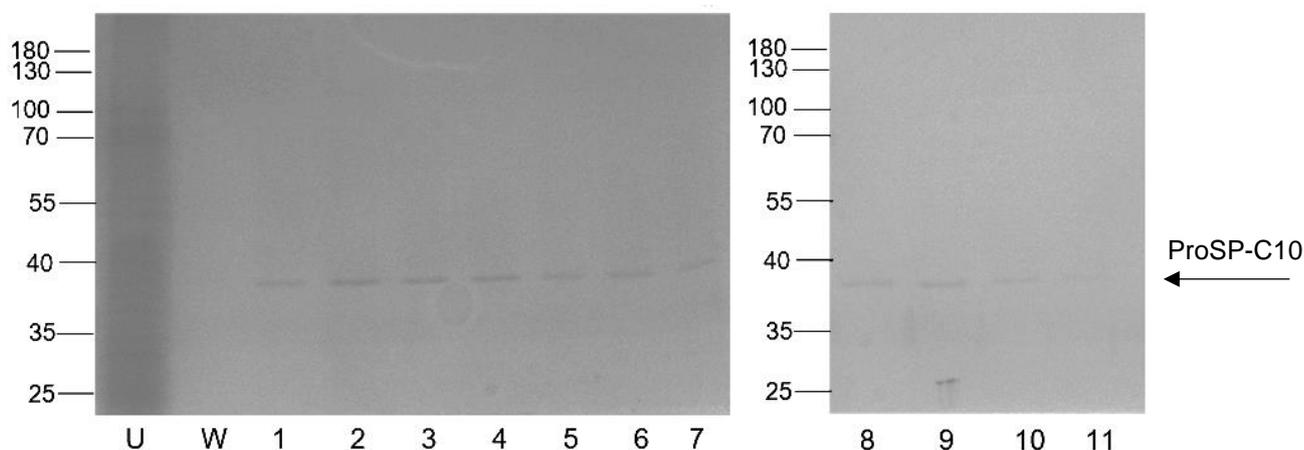
Encouraged by these results, the solubilisation and purification was repeated, but with the modification that the washes were carried out using a purification buffer containing 50 mM imidazole (50ml); Elution fractions were numbered 1-11 as they were all eluted using 1000 mM imidazole.

The purification process yielded pure proSP-C10, seen at the 37 kDa marker, see below in Figure 4.16.

No other contaminating protein bands could be seen, underlining that this was the ideal point to reach in terms of purity.

It was important to optimise purification in this way so as to be able to later identify that any surface activity seen was from proSP-C10 only.

This marked a significant turning point in the progress of characterising proSP-C10 and so being able to confidently observe its surface activity.



**Figure 4.16 Purification of ProSP-C10-SMALP**

ProSP-C10 was solubilised using SMA 2000P.

The solubilised fraction was purified with Ni-NTA beads in a polypropylene column.

A low imidazole wash (W) was carried out using 50 mM imidazole, and proteins were eluted with 1000 mM imidazole (lanes 1-11).

Proteins were run on an SDS-PAGE and stained with Coomassie.

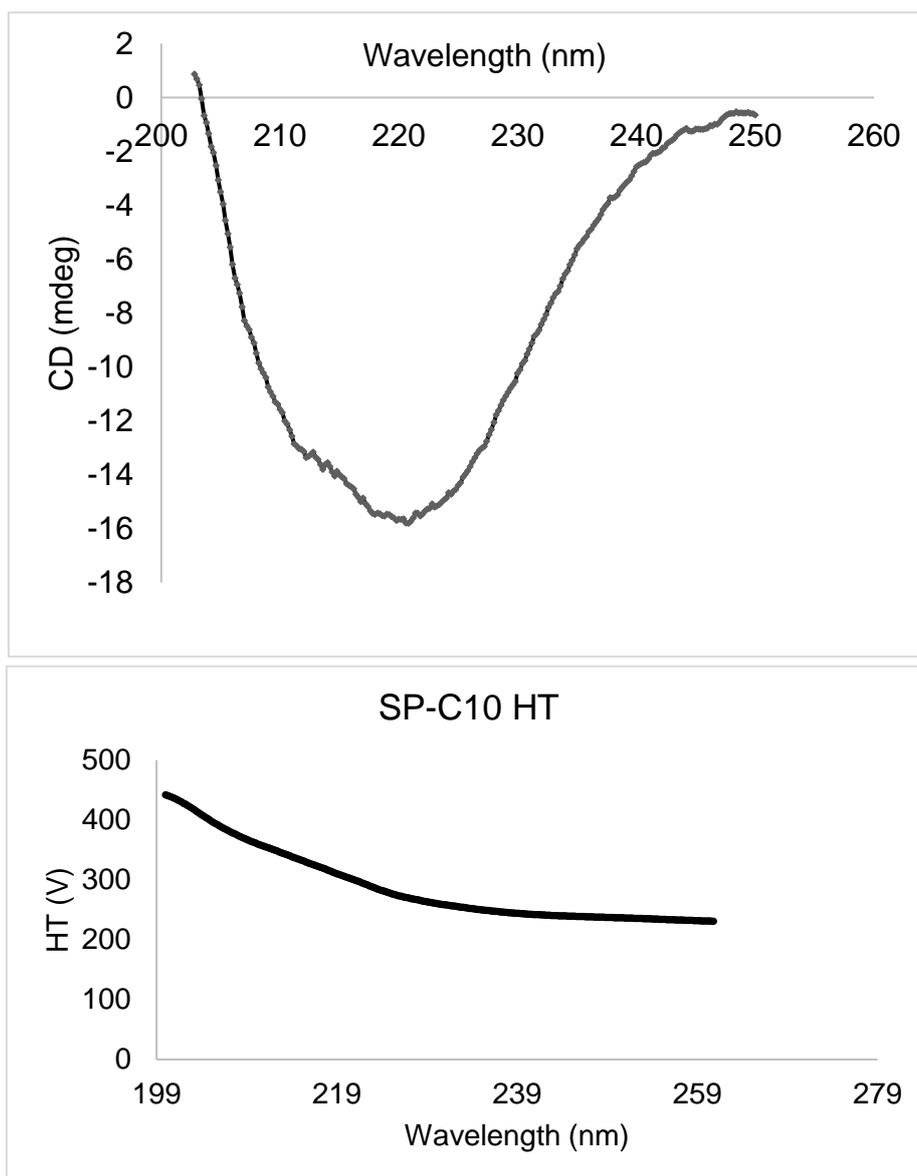
The band at 37 kDa was confirmed to be proSP-C10, see below in Tables 4.2 and 4.3.

To check that the protein was correctly folded, purified protein sample was taken to Birmingham University and investigated using Circular Dichroism (CD).

The secondary structure of this purified proSP-C10 protein product from the SDS-PAGE, shown in Figure 4.16 above, was analysed using CD.

ProSP-C10 showed strong negative peaks at 208 nm and 222 nm, indicating proSP-C10 to be composed of alpha helices, see Figure 4.17 below. The HT trace in the bottom panel shows that below 200nm the sample absorbed strongly, again, this likely due to the SMA.

This alpha helical secondary structure indicated, was fully as expected (Johansson *et al.* 1994).



**Figure 4.17 CD Spectrum and Accompanying HT trace of ProSPC-10 in a SMALP**

The CD spectrum of proSPC-10 indicates an alpha helical secondary structure for the protein.

Analysis indicated that proSP-C10 had a minimum at 208 nm and 222 nm;

ProSP-C10 protein concentration is 0.1 mg/ml.

CD was carried out by Charles Moore-Kelly at Birmingham University. JASCO J-1500 spectropolarimeter used. CD carried out in a 1 mm path length cuvette.

#### 4.11 Functional Characterisation of ProSP-C10 in SMALP

To next investigate whether any surface activity of proSP-C10 could be observed, it was solubilised and purified using SMA 2000P; before then being tested via a Langmuir trough.

See section 1.1.5 and 1.1.6, earlier in Chapter 1, for a full explanation of Langmuir trough and surface tension derivation.

ProSP-C10-SMALP and BmrA (an alternative control membrane protein) were investigated via the Langmuir trough.

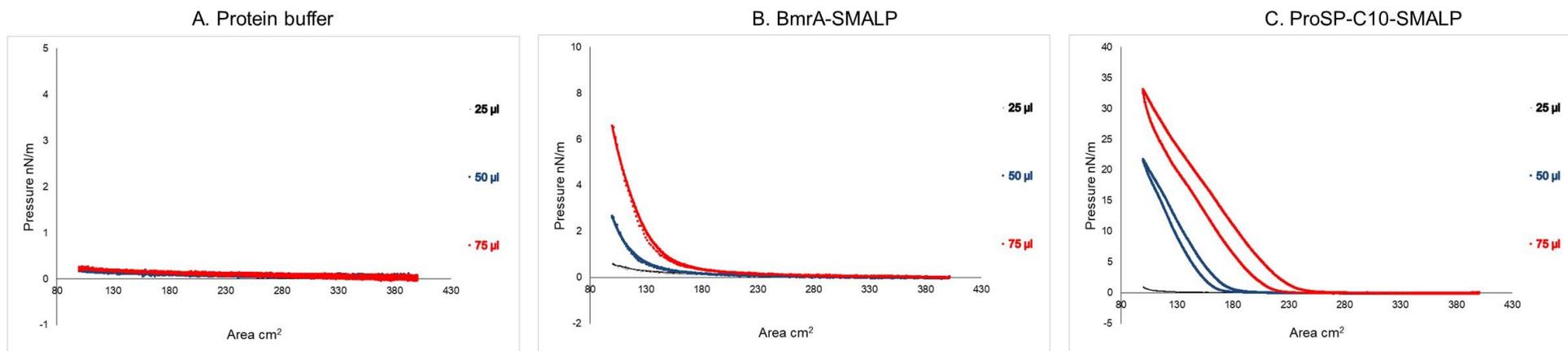
BmrA was chosen in this case because, as a membrane protein, it is closer in structure to proSP-C10 than such as BSA (SP-B, see Figure 3.24). It is also known to lack surface activity.

The shape of the isotherm, shown in Figure 4.18 below, indicates hydrophobic regions of proSP-C10 have been forced to interact under compression.

The three curves representing different concentrations of blank (BmrA) and proSP-C10 (25, 50 and 75  $\mu$ l), as shown below in Figure 4.18 show;

Surprisingly there does appear to be a small amount of surface activity from BmrA.

The experimentally repeated hysteresis curves on the verified, purified proSP-C10 product appear to show it to be surface active.



**Figure 4.18 Langmuir Trough Isotherms of ProSP-C10 and BmrA (control)**

Isotherms generated on a Langmuir trough highlighted the behaviour of proSP-C10 and BmrA molecules at the surface.

There is unexpected surface activity, albeit low indicated for BmrA.

Surface pressure measurements from protein buffer; this was the buffer within which proSP-C10-SMALP and BmrA molecules were suspended.

BmrA protein concentration was at 1.5  $\mu\text{M}$ .

ProSP-C10 in SMA 2000P (3  $\mu\text{M}$ ).

The increasing volumes cause an inflection of the curve suggesting that under pressure the hydrophobic regions are interacting with each other.

Figure 4.19 shows maximum surface pressure measurements vs volume added.

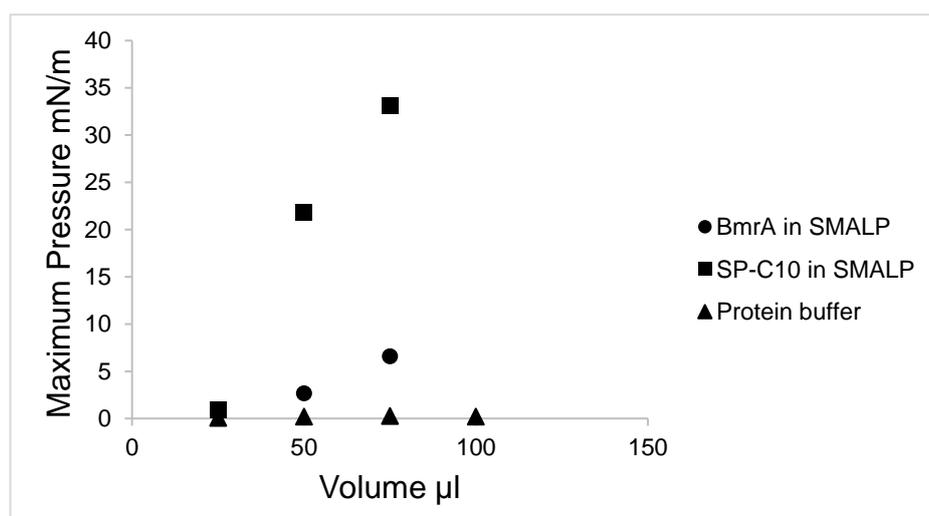
SMA-solubilised BmrA initially had a low surface pressure of 0 mN/m at 25  $\mu$ l, the surface pressure increasing to 2 mN/m at 50  $\mu$ l and finally to 6 mN/m at 75  $\mu$ l.

The surface pressure values seen from these samples shows slightly higher activity readings, than the surrogate 0 mN/m levels shown with BSA (Figure 3.24).

SP-C10 also had a low maximum surface pressure at 25  $\mu$ l of 0 mN/m, this increased to a higher surface pressure of 21 mN/m at 50  $\mu$ l and 33 mN/m at 75  $\mu$ l.

These results indicate that the proSP-C10 remains at the surface and exhibiting a surface pressure increasing with number of molecules added.

This activity differentiates proSP-C10 from other surface active molecules because it remains at the surface. This seen from the increase in maximum surface activity when the number of proSP-C10 molecules is increased.



**Figure 4.19** Maximum Surface Activity of ProSP-C10-SMALP

ProSP-C10-SMALP (3  $\mu$ M) was injected onto the surface of a Langmuir trough and the surface pressure was plotted as a function of volume added to the experiment.

The SMALP-solubilised BmrA starts at a low surface activity of 0 mN/m at 25  $\mu$ l which steadily increases as the volume added to the trough increases. ProSP-C10 also had a low maximum activity at 25  $\mu$ l of 0 mN/m; this increased to 33 mN/m at 75  $\mu$ l. The ability to remain at the surface is one of the distinguishing features of SP-C. ProSP-C10 protein concentration was at 3  $\mu$ M. BmrA protein concentration was 1.5  $\mu$ M.

The Langmuir trough experiments were carried out with proSP-C10-SMALP, using BmrA as a control.

However, as SMA 2000P is amphipathic, it is possible that some surface activity seen in both the proSP-C10-SMALP and BmrA-SMALP samples could be attributed to the behaviour of the polymer.

The decision to reconstitute proSP-C10 into liposomes was therefore taken; this way only the lipids surrounding the protein would be present in the sample.

To get a closer representation of the *in vivo* situation e.g. avoiding such as the potential risk of polymer effect, detergent solubilisation of proSP-C10 with a view to reconstitution was revisited.

#### **4.12 Solubilisation and Purification of ProSP-C10 using Conventional Detergents**

A detergent screen was carried out to determine the detergents which could be used to solubilise proSP-C10, based on knowledge from the previous screen before use with proSP-C, and shown above in Figure 4.8.

Detergents: Anz, Fos, CHAPS, MNG, DDM,  $\beta$ -OG and Triton X-100 were all used to solubilise proSP-C10.

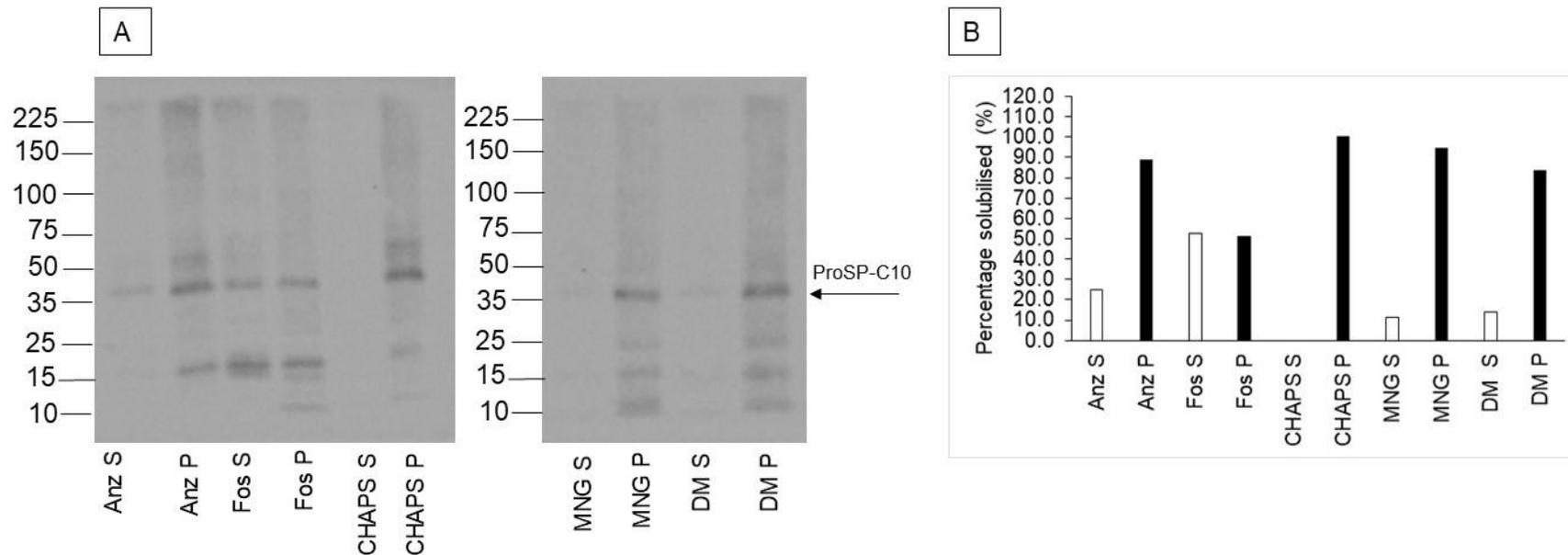
The solubilised supernatant and pellet were loaded onto an SDS-PAGE gel and transferred to a Western blot, see below in Figure 4.20 panel A.

The Fos supernatant had the strongest proSP-C band compared to all the other detergents.

This indicated that Fos was the most effective solubilising agent.

Anz, MNG and DM solubilised weakly, indicated by the faint 37 kDa band, whilst CHAPS,  $\beta$ -OG and Triton X-100 had strong 37 kDa bands in the pellets.

The solubilisation results were supported by the solubilisation analysis carried out using Image J seen in Figure 4.20 panel - B. The analysis indicated solubilisation efficiencies of the proSP-C10 membrane were: Fos (50%), Anzergent (24%), MNG (11%) and DM (14%) CHAPS (0%).



**Figure 4.20 Detergent Solubilisation Screen of ProSP-C10**

ProSP-C10-containing membranes were solubilised using 5% Anz, Fos, CHAPS, MNG and DM. The solubilised supernatant (S) separated from the pellet (P) was loaded onto a SDS-PAGE gel and transferred to a Western blot. Fos solubilisation resulted in a strong band in the supernatant fraction (Fos S) compared to all the other detergent solubilisations.

(B) Solubilisation analysis of proSPC-10 using Image J. Solubilisation analysis examined the band intensities of each of the solubilised (S) and unsolubilised (P) proSP-C10 fractions. Fos solubilised 50% of the proSP-C10 membrane, Anz, MNG and DM all solubilised 10-20% of the membrane. CHAPS failed to solubilise proSP-C10. Solubilisations were normalised to the CHAPS pellet (CHAPS P). An anti-His<sub>6</sub> monoclonal antibody and an anti-mouse IgG-HRP-linked antibody were used for detection.

ProSP-C10 membranes were solubilised using Fos choline-12 because this process showed the greatest protein extraction, see Figure 4.20 above.

The detergent concentration was lowered because larger solubilisations require much more detergent and these reagents are costly within a brief of a targeted, low cost alternative source of the target proteins.

This concentration is 43 times greater than the critical micelle concentration (CMC), this is the point at which there is excess detergent favouring the formation of micelles.

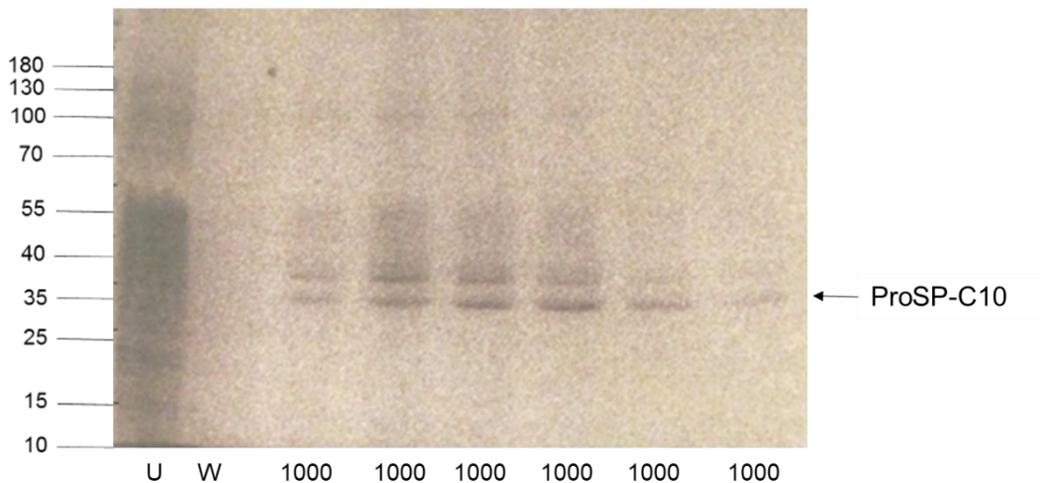
However, parallel detergent experiments with SP-C also indicated that lowering the detergent concentration would result in solubilised protein.

Solubilised proSP-C10 was thus purified, the fractions being eluted using 1000 mM imidazole.

A double band is seen near the 35 kDa molecular weight marker with a faint band at 100 kDa also being seen.

A band can be seen in supernatant of the Fos solubilisation below in Figure 4.21, but not as a double band.

This was considered to be a contaminating protein, or perhaps a consequence of the denaturing effects of the detergent used to solubilise the protein.



**Figure 4.21**      **Coomassie Stained Gel of ProSP-C10 Purification**

The solubilised proSP-C10 fraction was mixed with Ni-NTA, and purified fractions were eluted using 1000 mM imidazole.

ProSP-C is considered to be the 35 kDa band seen in the elution fractions.

The double band wasn't seen in the small scale detergent screen and was therefore assumed to be a contaminating protein or a consequence of denaturing by the solubilising detergent.

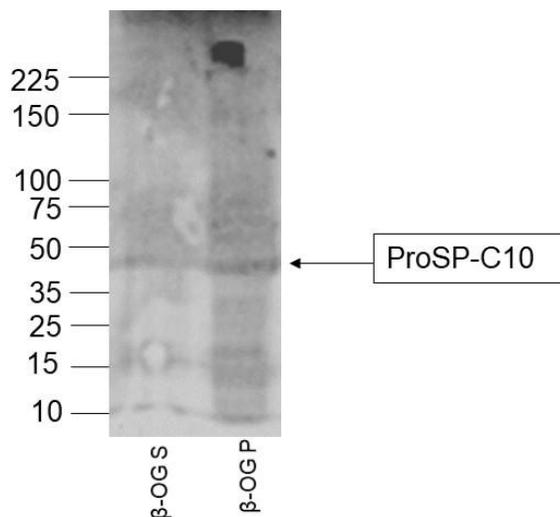
Faint bands at 100 kDa band could also be seen.

The solubilisation of proSP-C10 was repeated and the double band was seen again.

A milder detergent was therefore sought that would disrupt protein-lipid bonds only.

An SP-C solubilisation and purification screen had revealed that  $\beta$ -OG was the best detergent for extraction of SP-C; so given the similarity of the two proteins, a solubilisation using  $\beta$ -OG was repeated, see below in Figure 4.22.

ProSPC-10 at a final protein concentration of 2 mg/ml was solubilised with  $\beta$ -OG (2%) The supernatant and pellet are also shown below in a Western blot.



**Figure 4.22 Solubilisation of ProSP-C10 with  $\beta$ -OG**

ProSP-C10 was solubilised using  $\beta$ -OG.

The solubilised fraction (S) and unsolubilised fraction (P) were isolated, separated on an SDS-PAGE gel and visualised with a Western blot.

ProSP-C10 is the 40 kDa protein shown by the labelled arrow, and is present in the solubilised and unsolubilised fraction.

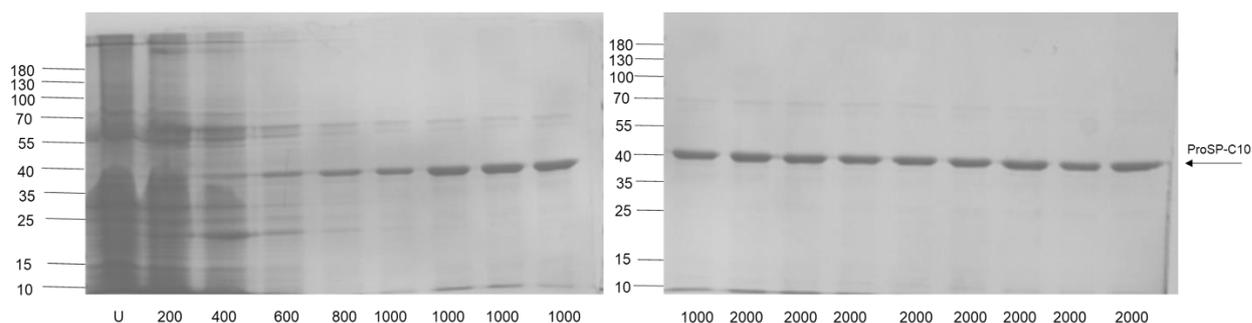
An anti-His<sub>6</sub> monoclonal antibody and anti-mouse IgG-HRP-linked secondary antibody were used.

ProSPC-10 was then solubilised at a larger scale, on the basis of its extraction with  $\beta$ -OG before being purified.

An imidazole gradient ranging from 200 mM to 2000 mM imidazole was used to purify proSP-C10.

The purification gel showed a purer eluted protein at 40 kDa than had been previously, as seen in Figure 4.23 below.

ProSP-C10 yield, calculated using densitometry, was 0.18 mg/L.



**Figure 4.23 Coomassie Stained Gel of ProSP-C10 Purification with  $\beta$ -OG**

ProSP-C10 was solubilised using a  $\beta$ -OG detergent concentration of 2% for 1 hour.

A stepped imidazole gradient, ranging from 200 mM-2000 mM, was carried out to determine the optimal concentration of imidazole to elute proSP-C10.

A predominant pure band can be seen at 40 kDa, at an optimal imidazole concentration of 2000 mM, but even from 800 mM the purity of proSP-C10 was only marginally poorer.

On the basis of the findings of its purity, the sample was used to investigate reconstitution of proSP-C10 into a proteoliposome.

Note:

ProSP-C10 was excised from a Coomassie stained gel, see above in Figure 4.23 and sent for mass spectrometry analysis at Birmingham University.

Mass spectrometry on the provided samples indicated that proSP-C10 was present with a coverage score of 13.88, as shown below in Tables 4.2 and 4.3.

As also reported earlier in this chapter, this indicated that proSP-C10 had a low match between expected and identified peptides from the spectra. Despite other more abundant peptides being identified as Keratin, proSP-C10 was *indeed present*.

The low match specificity may have been due to poor cleavage during sample preparation.

Arginine (Arg) and Lysine (Lys) have been cited as amino acid sites that result in poor cleavage (Keil 1992).

Trypsin cleavage of proSP-C resulted in cleavage at the following amino acid positions 6;Lys, 21;Arg, 23;Arg, 34;Lys, 35;Arg, 63;Lys, 81;Arg, 125;Lys, 139;Arg, 160;Leu, 167; Arg. Cleavage sites have been included, see below in Table 4.2.

This suggested that whilst proSP-C10 was indeed present in the sample, the sample was apparently contaminated, notably with Keratin.

**Table 4.3 Identified ProSP-C10 Peptide Fragments and their Location in the ProSP-C Sequence**

ProSP-C sequence:
MDVGSK <sup>6</sup> <b>EVLME<span style="background-color: #cccccc;">SPDYSAAPR</span></b> <sup>21</sup> GR <sup>23</sup> <b>FGIPCCPVHLK</b> <sup>34</sup> RLLIVVVVVVLLIVVVIVGALLMG LHMSQK <sup>63</sup> HTEMLVLEMSIGAPEAQQR <sup>81</sup> LALSEHLVTTATFSIGSTGLVVYDYQQLLIAYKP APGTCCYIMK <sup>125</sup> <b>IAPESIP<span style="background-color: #cccccc;">SLEALNR</span></b> <sup>139</sup> KVHNFQMECSLQAKPAVPTSK <sup>160</sup> LGQAEGR <sup>167</sup> D AGSAPSGGDPFLGMAVNTLCG EVPLYYI
EVLME <span style="background-color: #cccccc;">SPDYSAAPR</span>
IAPESIP <span style="background-color: #cccccc;">SLEALNR</span>

**Table 4.4 Identified Peptide Fragments from Mass Spectrometry Analysis of ProSP-C10 (40 kDa)**

ProSP-C10 was present with a coverage score of 13.88.

Sequencing was carried out at University of Birmingham.

Description	Score
1. Keratin 1 GN=KRT1 [H6VRF8_HUMAN]	63.53
2. Keratin, type I cytoskeletal 10 GN=KRT10 [K1C10_HUMAN]	41.20
3. Keratin, type I cytoskeletal 9 GN=KRT9 [K1C9_HUMAN]	34.39
4. Keratin, type II cytoskeletal 2 epidermal GN=KRT2 [K22E_HUMAN]	28.60
5. Keratin, type I cytoskeletal 14 GN=KRT14 [K1C14_HUMAN]	19.31
6. ATP synthase subunit beta, mitochondrial GN=ATP2 [ATPB_KLULA]	18.84
7. Elongation factor 1-alpha GN=TEF1 [EF1A_PICGU]	16.45
8. Keratin 77 GN=KRT77 [Q0IIN1_HUMAN]	15.79
9. Isoform 2 of Pulmonary surfactant-associated protein C GN=SFTPC [PSPC_HUMAN]	13.88

#### 4.13 Attempts to Reconstitute ProSP-C10 into Proteoliposomes

The surface activity of proSP-C10-SMALP is shown above in Figure 4.18.

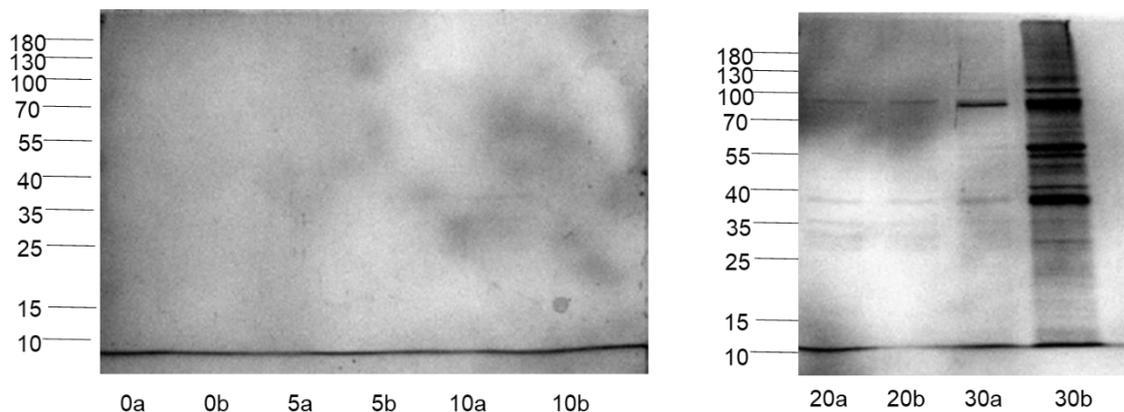
Encouraged by these results, the possibility of analysing the activity of proSP-C10 reconstituted into proteoliposomes, was investigated.

SP-C is found as a protein-lipid complex *in vivo*, and so reconstitution experiments better reflecting this physiological reality *in vitro* would be very valuable.

POPC was initially chosen as the lipid to use in reconstitution studies because it was both suitable and readily available.

A sucrose gradient was carried out on the reconstituted sample to inform relative success of the reconstitution procedure.

Results from a reconstitution are shown below in Figure 4.24.



**Figure 4.24 Sucrose Gradient of Reconstituted ProSP-C10**

Sucrose layers ranging from 5-30% concentrations were used to form sucrose gradients, this was performed to confirm the reconstitution of proSP-C10

The Coomassie stained gel shows a faint 40 kDa band is present in the and 20a, 20b, 30a and 30b fractions

The majority of SP-C10 protein can be seen in the 30b fraction, which is indicative of a poor reconstitution.

Limited co-migration of the SP-C10 proteins and lipids to fraction 20 along with associated molecular weight on a Western blot, represented a low yield of reconstituted material.

A heavy band just under 40 kDa can be seen in fraction 30 but the band gets significantly lighter towards fraction 20.

SP-C10 protein that remained in the 30 fraction represented a poor reconstitution.

Unfortunately, much of SP-C10 was in this 30 fraction, indicating it had aggregated during the reconstitution process.

The aggregation may have also indicated an overall failure of SP-C10 to reconstitute at all, one of the reasons causing this may well be the potential lipid disrupting characteristics of SP-C.

A further three, slightly varied, attempts to reconstitute proSP-C10 were undertaken but each was unsuccessful.

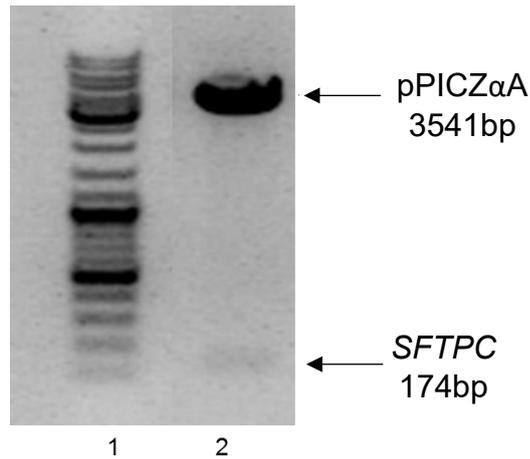
As SP-C is the mature active protein, the next experiments focussed on applying the lessons learned from the proSP-C10 synthesis and expression, and applying them to the expression of SP-C.

#### **4.14 SFTPC Synthesis**

SP-C is the active peptide in pulmonary surfactant, formed by the cleavage of proSP-C (Weaver 1998).

The cloning of the gene encoding SP-C (*SFTPC*) was carried out using the previously-described strategy, shown earlier in Figure 4.1.

Transformed vectors were digested to check for successful ligation, and revealed a 174 bp *SFTPC* insert and a 3,451 bp vector backbone, as seen below in Figure 4.25.



**Figure 4.25 SFTPC Synthesis**

The DNA sequence encoding *SFTPC* was amplified using polymerase chain reaction (PCR).

The PCR fragments and pPICZαA vector were digested with *EcoR1* and *Xba1*, ligated and transformed into super-competent *E. coli* XL10 -Gold.

The transformed vectors were digested again with *EcoR1* and *Xba1* and revealed *SFTPC* (174 bp) inserts and vector backbone (3541 bp).

The transformed *SFTPC* vector was analysed externally by Eurofins.

**The translated amino acid sequence aligned completely to the NCBI SP-C sequence, see below in Figure 4.26.**

Recomb SFTPC	FGIPCCPVHL	KRL LIVVVVV	VLIVVVIVGA	LLMGL
SFTPC NCBI	.....	.....	.....	.....
Identity	*****	*****	*****	*****

**Figure 4.26 SP-C Sequence Alignment**

The *SFTPC* sequence was independently analysed by Eurofins and translated into its corresponding recombinant amino acid sequence (Recomb SFTPC).

This sequence was aligned against the NCBI SP-C sequence.

The amino acid sequence of recombinant SP-C is identical (denoted by \*) to the amino acid sequences of that in the NCBI databank.

#### 4.15 Isolation of a High-Yielding SP-C Colony

Protein yield from transformed colonies has been carried out as described previously, and was assessed using a Western blot, see below in Figure 4.27.

A 10 kDa band can be seen from colonies grown on the 100  $\mu\text{g/ml}$  and 200  $\mu\text{g/ml}$  YPDS plate. This is close to the expected size of 6kDa.

This band appears to be SP-C, but there could be the possibility that it is a contaminating protein that is binding to the His<sub>6</sub>-tag.

A 15 kDa band and 10 kDa band could be seen from colonies grown on the 500  $\mu\text{g/ml}$  plate.

As the zeocin concentration in each plate increased, the band intensity also increased.

Colony 5 from the 500  $\mu\text{g/ml}$  plate was chosen for further scale up experiments as the dominant band was the 10 kDa band.

Also, there appeared to be greater purity compared to the bands from the other colonies.

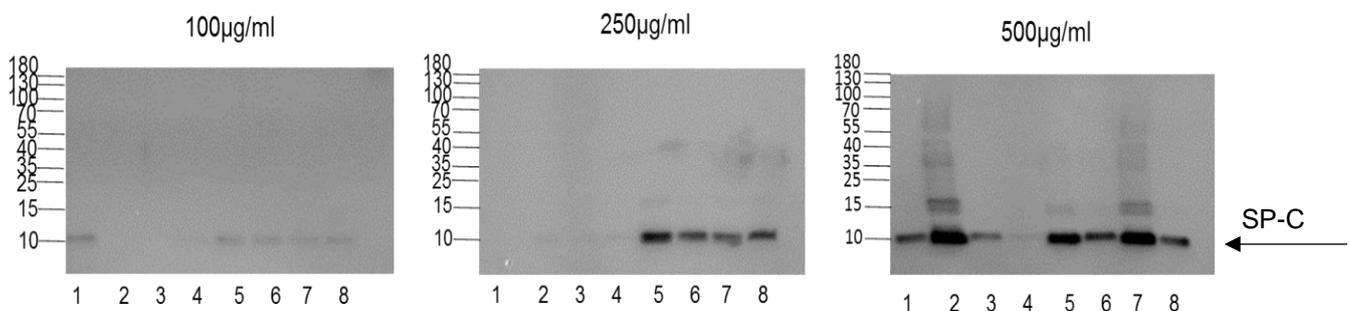


Figure 4.27 SP-C Expression Screen

A 10 kDa band was expressed by all colonies screened.

There is an additional 15 kDa protein band in colonies from the 500  $\mu\text{g/ml}$  plates.

10 kDa compares to the expected 6 kDa size of SP-C.

Protein yield was assessed using a Western blot.

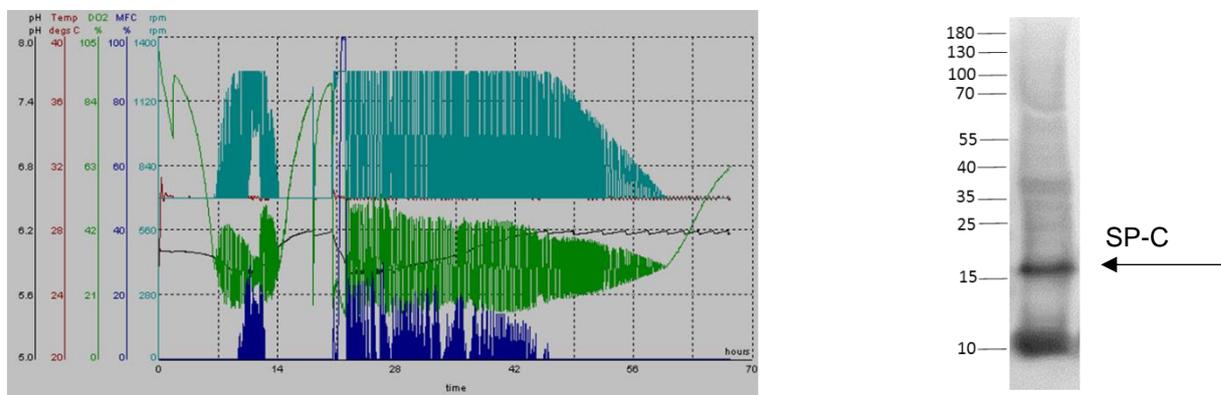
Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used.

25  $\mu\text{g}$  membrane preparation was loaded into each lane.

#### 4.16 Scale-Up of a High Expressing SP-C Colony

Colony 5 was cultured in a bioreactor using the previously developed, standard protein expression conditions for *P. pastoris*.

Figure 4.28 shown below, is a representative bioreactor trace; whilst alongside is a Western blot showing a membrane preparation from the culture after some 50 hours.



**Figure 4.28 Scale-Up of a High Expressing SP-C Colony**

A bioreactor trace representing growth of SP-C-producing colony 5 over 50 hours.

The growth of the culture can be monitored by the green dissolved oxygen (DO) trace.

The RPM is represented by the turquoise trace and pH is shown by the line in black; it was at pH 6.0 consistently throughout the run.

The trace in red shows the temperature was maintained at 30°C for the duration of the run.

A membrane prep from a bioreactor run was isolated from the culture and has been analysed by both SDS-PAGE and Western blot.

Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used.

SP-C is presumed to be the 10 kDa protein shown by the arrow.

40 µg of total protein was loaded per lane.

#### **4.17 SP-C Detergent Solubilisation Screen**

A detergent solubilisation screen to assess extraction of recombinant SP-C was performed.

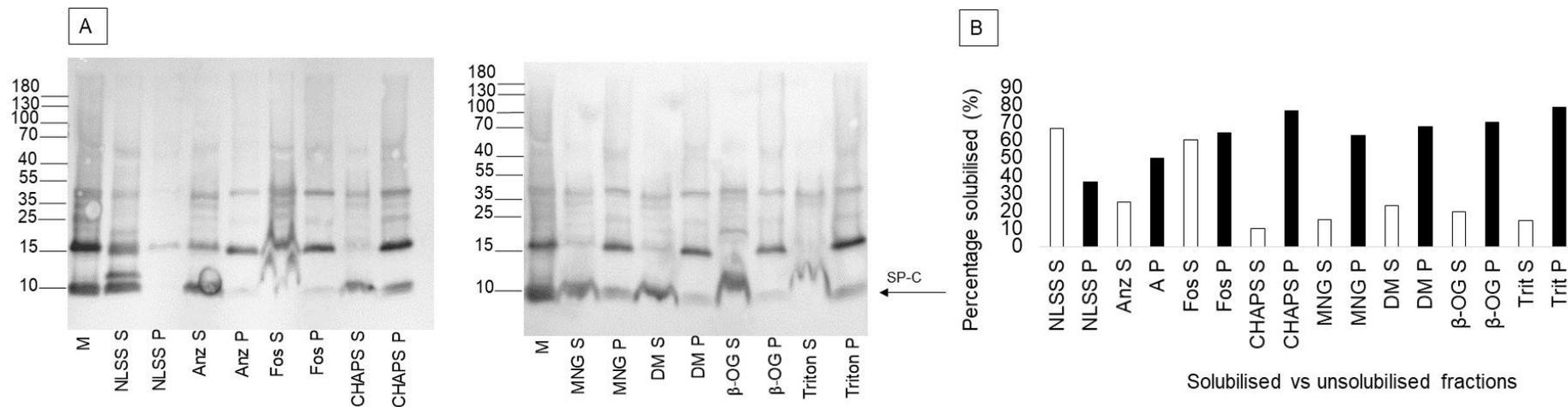
The detergents used were the same mix of ionic, non-ionic and zwitter ionic detergents used previously in Figure 4.8.

The solubilised supernatants and pellets were loaded into SDS-PAGE and visualised using a Western blot.

The Western blot analysis identified Fos and NLSS as the best solubilising agents as seen below in Figure 4.29 panel - A.

NLSS had been used initially, but it caused the Nickel to be stripped away from Ni-NTA during purification.

Fos was therefore the next best extractant.



**Figure 4.29 SP-C Detergent Solubilisation Screen**

Solubilisation of SP-C with detergents. The solubilised supernatants were loaded onto a Tris-Tricine gel and transferred to a Western blot.

Anti-His<sub>6</sub> monoclonal and anti-mouse IgG-HRP-linked secondary antibodies were used to detect SP-C. Key: M membrane preparation, S supernatant, P pellet. The supernatant and pellet were loaded respectively to accommodate some six slightly modified (varied detergents) replications for the following detergents: NLSS, Anz, Fos, CHAPS, MNG, DM, β-OG and Triton.

Solubilisation efficiency of detergent in extraction of SP-C.

The detergents that gave the best solubilising efficiency were NLSS (66%) and Fos (60%).

#### 4.18 Purification of Detergent Solubilised SP-C

SP-C was solubilised with Fos-12 detergent.

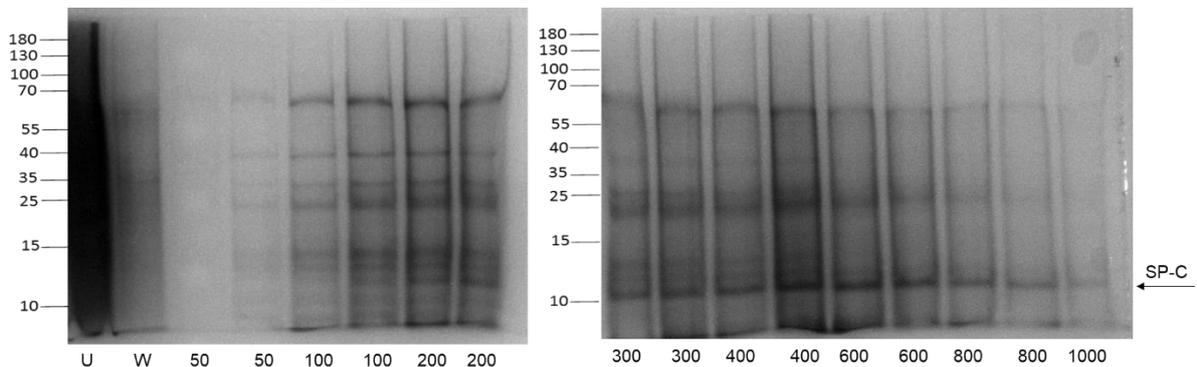
The solubilised fraction was separated and purified using Ni-NTA.

A stepped imidazole gradient, ranging from 50 mM to 1000 mM was used to assess the optimum amount of imidazole needed for elution.

A high amount of contaminating protein was eluted at 100 mM.

Only at 800 mM were the contaminating proteins removed.

SP-C is presumed to be the 10 kDa band shown below by the arrow in Figure 4.30.



**Figure 4.30 Solubilisation and Purification of SP-C with Fos-Choline 12**

SP-C was purified using an imidazole gradient.

Contaminating proteins around 15 kDa, 25 kDa and 70 kDa were eluted between 50 mM and 200 mM.

SP-C (10 kDa), began to be eluted at 300 mM but other bands around 25 kDa and 70 kDa were also eluted.

As the imidazole concentration is increased, the purity of SP-C increases, and many of the bands around 25 kDa and 70 kDa are removed.

A pure fraction of SP-C is seen in the 1000 mM imidazole fraction.

The solubilisation and purification of SP-C using Fos, resulted in the elution of other contaminating proteins alongside SP-C, see above in Figure 4.30.

As the concentration of imidazole was increased from 100 to 200 mM, several contaminating bands were eluted.

More contaminants around 70 kDa and 25 kDa were eluted at 300 mM.

SP-C was also eluted at this concentration (10 kDa).

All 3 of these bands continued being eluted as the concentration of imidazole was increased from 300 to 800 mM.

At 1000 mM imidazole, a substantially pure sample of SP-C can be seen, although the yield is low.

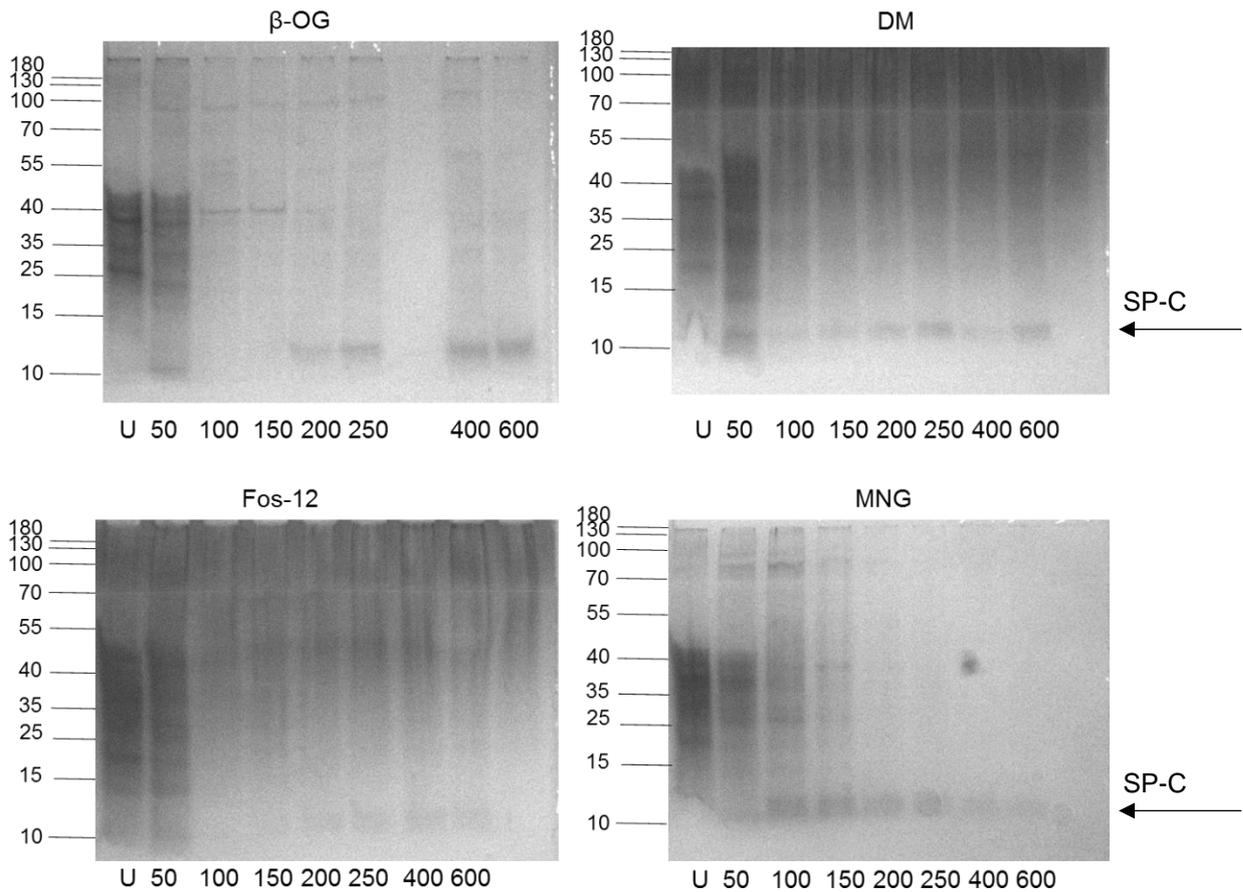
Despite Fos having the best solubilising efficiency for SP-C, when it was used to purify SP-C, it resulted in eluents with other contaminating proteins. These were present at a high concentration, and thus overall purified yield was low, see above in Figure 4.30.

A purification screen was therefore carried out to examine whether the yield of pure SP-C could be improved by using yet another different detergent.

Detergents  $\beta$ -OG, DM and MNG were used in a purification screen.

These detergents were selected because they had also been shown to solubilise SP-C previously, as seen above in Figure 4.30.

A stepped imidazole gradient that ranged from 50-600 mM was used to determine which detergent yielded pure SP-C, as seen below in Figure 4.31.



**Figure 4.31 Purification Screen of SP-C Solubilized in  $\beta$ -OG, DM, Fos and MNG**

A detergent concentration of 2% was used to carry out both solubilisations and purifications to determine the optimum detergent with which to extract SP-C.

A stepped imidazole gradient (50 mM-600mM) was used to elute purified proteins.

Solubilisation of SP-C using  $\beta$ -OG gave the best purification yield.

Observation of the purification fractions that were eluted using  $\beta$ -OG, showed most of the contaminating proteins were eluted at 50 mM.

A further contaminating protein at a molecular weight of 40 kDa was eluted at both 100 mM and 150 mM.

SP-C started to be eluted at 250 mM; a faint band at 100 kDa can also be seen, but SP-C is the predominant band.

This 100 kDa band was also present in the 400 mM and 600 mM imidazole elutions.

Within the purification fractions using DM to elute, the majority of the contaminating proteins that had co-purified were eluted in the 50 mM imidazole purification fraction.

SP-C was also seen in this fraction (10 kDa) and continued to be eluted throughout the 100 mM to 600 mM fractions; each of these fractions had little or no contaminating protein.

The Fos purification contained contaminating bands that were eluted at 50 mM.

Other contaminating proteins were also seen between 40 kDa and 55 kDa throughout the stepped imidazole gradient.

SP-C was eluted between 150 mM and 200 mM.

In the purification fractions that used MNG, most of the contaminating proteins were eluted between 50 mM and 100 mM.

SP-C was also seen in these fractions (10 kDa); other contaminating bands at 25 kDa and 40 kDa were also eluted at the same point.

Pure SP-C fractions could be seen in the imidazole fractions eluted between 200 mM and 250 mM.

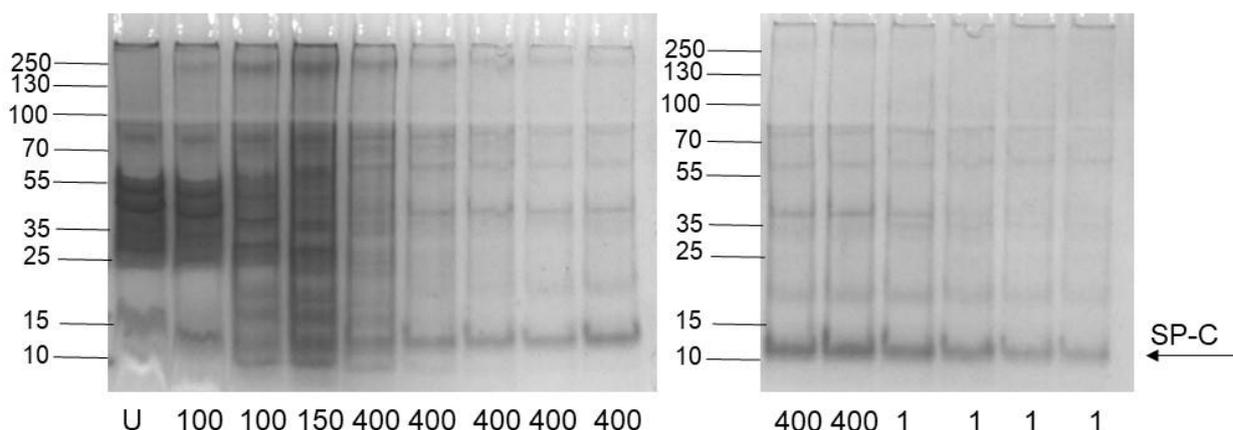
Based on the purification screen shown above in Figure 4.31,  $\beta$ -OG was concluded to be the most suitable detergent for solubilisation and purification of SP-C.

This was because it had the highest pure band intensity.

A large scale solubilisation, followed by purification was next carried out.

An imidazole gradient ranging from 100-150 mM was used to elute contaminating proteins.

SP-C was eluted at an imidazole concentration of 400 mM to 1000 mM.



**Figure 4.32 Solubilisation and Purification of SP-C with  $\beta$ -OG**

SP-C was solubilised and purified using  $\beta$ -OG and a stepped imidazole gradient was carried out to determine the optimal imidazole concentration for SP-C elution.

Many of the contaminating proteins were eluted at an imidazole concentration of 100 mM-150 mM.

A protein, considered to be SP-C, was eluted at an imidazole concentration of 400 mM.

SP-C elution fractions were concentrated, and then run on a Tris-Tricine gel, see above in Figure 4.32.

The gel was stained with Coomassie and those proteins at molecular weights of 100 kDa, 70 kDa, 40 kDa and 12 kDa were all sent off for independent mass spectrometry analysis at Birmingham and St Andrews Universities.

Despite some six differing attempts, a positive ID for SP-C could not be attained; the possible reasons behind this could be poor cleavage of the peptide during sample preparation.

This may have had an effect on how well the peptides were detected.

Trypsin was used to prepare SP-C samples for mass spectrometry analysis. The cleavage sites of SP-C using trypsin are found at amino acid positions 1 (Arg), 3 (Arg), 13 (Lys) and 14 (Arg). See Figure 1.12 for detailed amino acid positions.

Trypsin cleavage is blocked by arginine (Arg) and Lysine (Lys) (Keil 1992), these amino acids correspond to the cleavage site and provide an explanation for the multiple negative identification attempts.

Whilst these various differing sample productions and their mass spectrometry were being analysed, the product sample was being assessed as a proof of principle for *in vitro* reconstitution.

Despite three differing approaches to reconstitution; combining purified SP-C protein and POPC lipids and their loading for assessment into a sucrose gradient, unfortunately rather as earlier, but rather more disappointingly, the reconstitutions could not be viewed as successful.

All of the reconstituted protein was found to be present only in the pellet component, none had separated into the higher sucrose fractions.

#### **4.19 Effect of SP-C Associated with POPC Lipids on Surface Tension**

SP-C (0.26  $\mu$ M) was mixed with POPC lipids (2.2 mg/ml), but reconstitution was unable to be confirmed on a sucrose gradient (empty gel).

Using the Langmuir trough, the activity of SP-C mixed with POPC lipids was assessed, see below in Figure 4.33.

POPC alone has a low surface pressure of 0mN/m at 25  $\mu$ l; the surface pressure increases slightly to 2 mN/m at 50  $\mu$ l.

Unexpectedly, the surface pressure of SP-C mixed with POPC was 17 mN/m at 25  $\mu$ l but increased to 25 mN/m at 50  $\mu$ l.

Note: When added to phospholipids, SP-C should lower the surface pressure; the results found are therefore unexpected.

## DISCUSSION

A rising surface pressure may indicate problems with the secondary structure e.g. folding of the helices that may have arisen during the association of purified protein and POPC lipids.

It has also been reported that SP-C has a tendency to aggregate when removed from a native membrane (Lukovic *et al.* 2006).

This tendency of SP-C to misfold, forms the basis in of its involvement in alveolar proteinosis (Gustaffson *et al.* 2001).

Removal of SP-C from the membrane causes the native lipids to be replaced with detergents. During this process SP-C may temporarily be 'exposed' to the buffer environment where the hydrophobic regions may prefer to associate with other hydrophobic regions.

A further process where the SP-C structure is 'vulnerable' is the reconstitution into lipid vesicles.

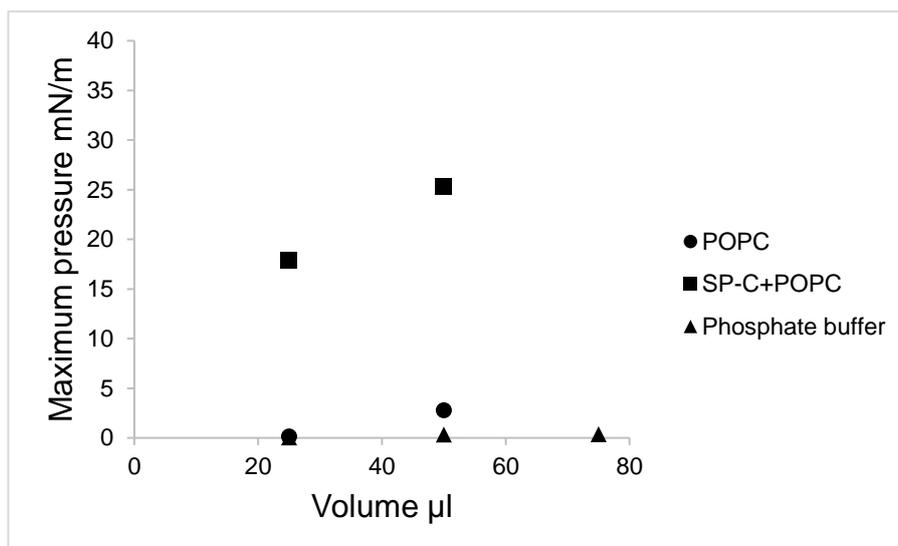
The replacement of detergents with lipids may render hydrophobic parts of the structure to again aggregate.

In light of the extreme hydrophobicity of SP-C, there are many instances at which this could occur in the structure.

Six consecutive valine residues form part of the SP-C amino acid sequence, see earlier in Figure 1.12, together form an alpha helical secondary structure.

Valine residues tend to be underrepresented in alpha helices (Berg, Tomoko and Stryer 2002) but form alpha helices in a lipid environment (Lukovic *et al.* 2006).

Removal of SP-C from the native membrane may cause the hydrophobic valine residues to revert to a  $\beta$ -pleated sheet conformation (Gustaffson *et al.* 2001).



**Figure 4.33 Maximum Surface Pressure of SP-C Mixed with POPC Lipids**

POPC lipids alone have a low surface pressure of 0mN/m at 25 μl, with the surface pressure increasing slightly to 2 mN/m at 50 μl.

The surface pressure of SP-C with POPC started at 17 mN/m at 25 μl and increased to 25 mN/m at 50 μl.

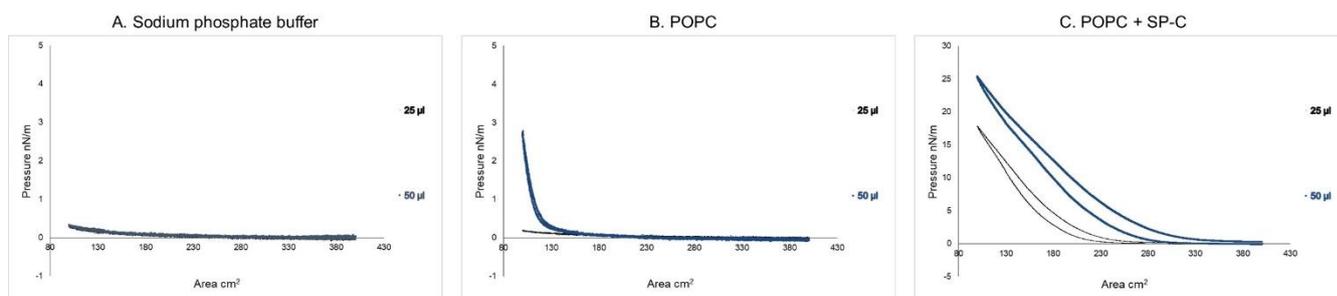
SP-C concentration was 0.26 μM and POPC lipid concentration was 2.2 mg/ml.

The pressure-area isotherms generated for POPC show 'typical' isotherms from a surface active material, as shown below in Figure 4.35.

SP-C mixed with POPC show a different shape to that for POPC alone, with the inflection greater when the volume added is increased.

SP-C mixed with POPC also showed differences in the shape of the curve after compression, again especially after the added volume was increased.

Denaturation of the alpha helical properties of SP-C have been shown to reduce lipid spreading (Johansson 1995). The possible aggregation could have been a further reason SP-C didn't get to low surface activity in these functionality experiments.



**Figure 4.34 Langmuir Trough Isotherms of SP-C Mixed with POPC Lipids**

The behaviour of POPC (2.2 mg/ml) and SPC 0.26  $\mu$ M mixed with POPC under compression and expansion was examined.

The shape of the POPC isotherm was very different to the shape of the isotherm when POPC and SP-C are jointly added.

#### **4.20 Conclusions – Recombinant Synthesis of ProSP-C and SP-C**

***For the first time, ProSFTPC was synthesised in a yeast vector, having been confirmed using sequence alignment.***

The new synthesised vector was expressed in *Pichia pastoris* yeast, and cultured at a large scale in a bioreactor.

***Positive identification of proSP-C using mass spectrometry confirmed the successful synthesis and expression of ProSP-C for the first time.***

***Optimization of proSP-C to a construct termed proSP-C10, containing a longer His<sub>10</sub>-tag, was achieved.***

***ProSP-C10 was extracted by SMA2000P for the first time.***

ProSP-C10 purification was optimised and the correct secondary structure confirmed using CD

***Preliminary surface activity measurements, using the Langmuir trough, indicate proSP-C10 (encapsulated in a SMALP) shows evidence of surface activity.***

The synthesis of SFTPC in a *Pichia pastoris* vector was confirmed via sequence alignment.

***Expression and evidence of activity of proSP-C10 has generated a reproducible protocol that can be used in therapeutic treatments involving SP-C***

## Chapter 5

### **Analysis of the Lipid Context of Recombinant Surfactant Proteins: Using Mass Spectrometry to assist Molecular Dynamics Simulations**

Chapters 3 and 4 indicated that the expressed recombinant proteins SP-B and SP-C, had surface activity; both on their own and in the presence of lipids.

Whilst SP-B and SP-C are involved in lipid modulation, their involvement in re-spreading the tear film lipid layer, in particular during blinking, is not well understood.

Currently, differences in surface tension between the polar layer and the aqueous layer help to drive the spreading of the tear film after a blink (Coren and Berger 1974).

The contributory role of SP-B and SP-C in the tear film lipid spreading mechanism hasn't yet been explored, this work aims to aid further understanding of this issue.

SP-B and SP-C interact with phospholipids of the lipid polar layer.

A detailed phospholipid composition was required in order to observe how these proteins influence lipid behaviour. Ideally this would have been a phospholipid composition from the tear film but there was no agreement on the exact composition. The yeast phospholipid composition was therefore used as a 'model membrane'

Thus the key objective of this Chapter 5 is to further build on the independently confirmed analyses-by such as mass spectrometry and sequencing production- of the purification and surface activity work of Chapters 3 and 4.

These inputs are now combined to elucidate a phospholipid composition which will subsequently also further inform the modelling efforts in the following Chapter 6.

As mentioned earlier, an additional particular challenge in investigating the behaviour of SP-B and SP-C in the tear film, is that there is no current agreement on the composition of the polar layer phospholipids, see earlier in Chapter 1, Table 1.2.

To address this particular obstacle, the phospholipids surrounding SP-B and SP-C in the *Pichia pastoris* membrane were used as a model for the tear film polar layer.

The types of phospholipids present in both the *Pichia pastoris* membrane, and the lipid polar layer, are the same; but the amounts of each phospholipid vary.

This model helps to inform the behaviour of SP-B and SP-C, with some of the findings potentially translating to understanding the tear film dynamics.

To gain detailed information about the phospholipids interacting with SP-B and SP-C in the yeast membrane, lipids around SP-B and SP-C were extracted using SMA 2000P.

The SMALP discs encapsulated SP-B and SP-C proteins, as well as the surrounding phospholipids.

These were extracted and analysed using mass spectrometry analysis (Dr Ivana Milic Aston University).

The identified lipids helped formulate explicit membranes, see later in Chapter 6, so as to observe the effect of SP-B and SP-C on lipid behaviour.

## **5.1 Lipid Extraction of ProSP-C, ProSP-C10, SP-C, ProSP-B and SP-B**

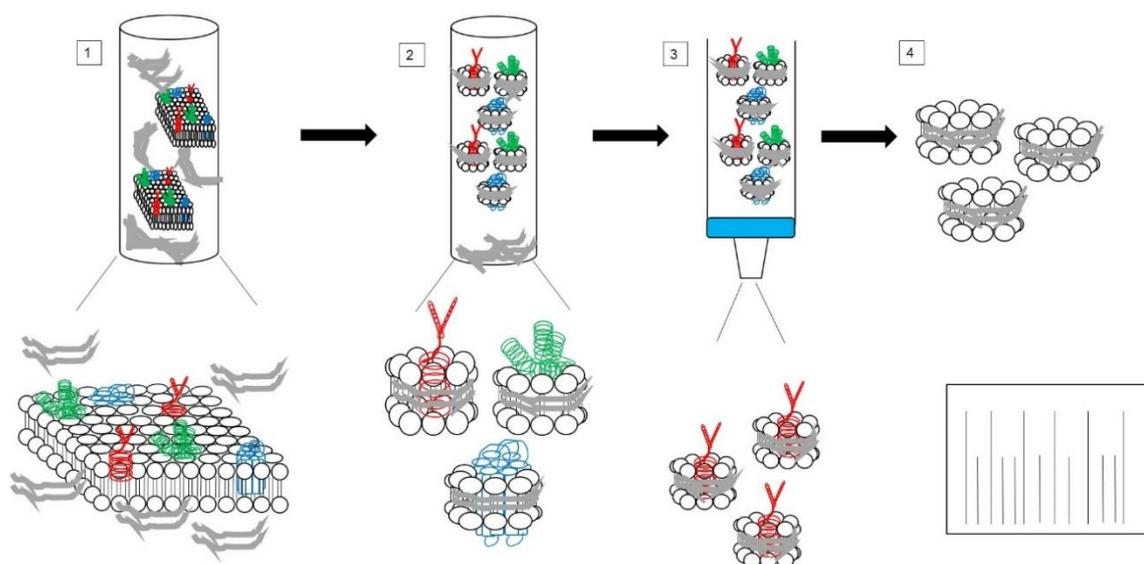
To separate SP-B, SP-C proteins and surrounding phospholipids, 2.5% w/v SMA2000P polymer was added to membrane fractions, as illustrated below in Figure 5.1 panel 1.

Solubilised fractions were purified.

The remaining protein and lipids were associated with either the SP-B or SP-C surrounded by the SMALP.

Isolation of pure protein and associated lipids is shown below in Figure 5, panel 2 and Figure 5.1, panel 3.

Lipids were extracted using the method of (Matyash *et al.* 2008) and analysed using mass spectrometry.



**Figure 5.1 Extraction of Lipids Encapsulated in a SMALP for Analysis by Mass Spectrometry**

2.5% w/v SMA2000P polymer was added to membranes at a final concentration of 40 mg/ml and left shaking for 1 hour (panel 1).

The solubilised fraction was separated (panel 2) and purified (panel 3).

Extracted lipids (Matyash *et al.* 2008) were analysed using mass spectrometry (panel 4).

## 5.2 Mass Spectrometric Analysis of the Lipid Compositions of SMALPs

### Containing ProSP-C, ProSP-C10, SP-C, ProSP-B, and SP-B

Analysis of the lipid composition was performed by comparison of the native lipid membrane to SMA extracted purified membranes.

Examination of differences between the samples highlighted the important protein-lipid interactions.

ProSP-C was the representative control sample.

The protein was extracted using the SMA2000P polymer but not purified, and so contained membrane proteins and associated lipids.

Preparation of the proSP-C control sample provided an understanding of the phospholipid context before purification.

The control presented a reference for phospholipids extracted from around purified proSP-C10, SP-C, proSP-B, and SP-B.

The lipid compositions of SMA-extracted and purified proSP-C10, SP-C, proSP-B and SP-B were obtained from samples that had been prepared according to the method described above in Figure 5.1.

The lipid species identified are shown in the corresponding mass spectra, see below in Figure 5.8.

Overlapping lipid species between the different proteins are also illustrated in a Venn diagram, see below in Figure 5.9.

Seven lipid species, at 943.7366 and 971.7689, could not be identified

### **5.3 Lipids Associated with Recombinant ProSP-C;**

#### **The Control Reference Condition**

Crude proSP-C membranes, extracted with SMA, were used as a control.

The phospholipids identified were phosphatidylcholines (PC), phosphatidylethanolamine (PE), phosphatidylglycerol (PG), phosphatidylinositol (PI) and triglycerides;

PC and PE, have been reported to be highly abundant in other mass spectrometry analysis of *Pichia pastoris* membrane; in contrast, PG was not reported (Grillitsch *et al.* 2014).

Triglycerides were also detected.

#### **5.4 Lipids Associated with Recombinant, SMA-Extracted, Purified ProSP-C10**

The phospholipids identified around purified proSP-C10 saw the introduction of more species of PI, PA and fewer species of TG.

Other phospholipids that were also present were PE, PC and PG.

The presence of PC, PE, PG, PI and PA is considered to indicate that these phospholipids are close protein-lipid associations.

PI and PA are anionic lipids.

The comparisons and implications of this information are discussed below.

#### **5.5 Some Implications of Lipids Associated with**

##### **Recombinant, SMA-Extracted, Purified SP-C**

The phospholipids identified around purified SP-C saw the introduction of more species of PI, PE and PA.

There were fewer species of PC and TG.

PI, PE and PA were also reported in the full length proSP-C10, this being further potential evidence of close protein-lipid interactions.

As already stated, PI and PA are anionic phospholipids, see below.

The N-terminus of SP-C is charged and this may be indicative of mutual attraction between the negatively charged lipids and positively charged N-terminus.

PE is involved in membrane curvature; this characteristic helping to provide supporting evidence of curvature in relation to lipid spreading.

Refer to Figure 1.14 earlier for a fuller description of the role of SP-B and SP-C in lipid vesicle formation.

The presence of PE around proSP-C10 and SP-C fractions may potentially implicate and support the involvement of SP-C in lipid spreading.

There are also fewer species of TG.

SP-C is a highly hydrophobic protein and a retention in TG may have implied a shielding of the protein by these lipids; the reduction in TG species is therefore surprising.

The fewer TG species around the proSP-C10, again may also be supportive of close protein-lipid associations.

## **5.6 Lipids Associated with Recombinant, SMA-Extracted, Purified ProSP-B**

Again, there is a notable increase in the number of PI species around purified proSP-B, compared to the proSP-C control.

There is also a decrease in the number of TG species.

Different species of PG, PI and PA have been introduced.

As discussed earlier, these are anionic (negatively charged) lipids and so their association may be driven by a positive charge, of +7 from the SP-B protein (Gill *et al.* 1995).

## 5.7 Lipids Associated with Recombinant, SMA-Extracted, Purified SP-B

A similar composition of phospholipid species is present in SP-B compared to SP-C.

Large amounts of additional PI and PG were also present.

As noted earlier, PI and PG are negatively charged lipids. The overall protein structure of SP-B has a positive charge of +7 and may indicate that the opposing charges between the lipids and protein are the basis for their interaction (Gill *et al.* 1995).

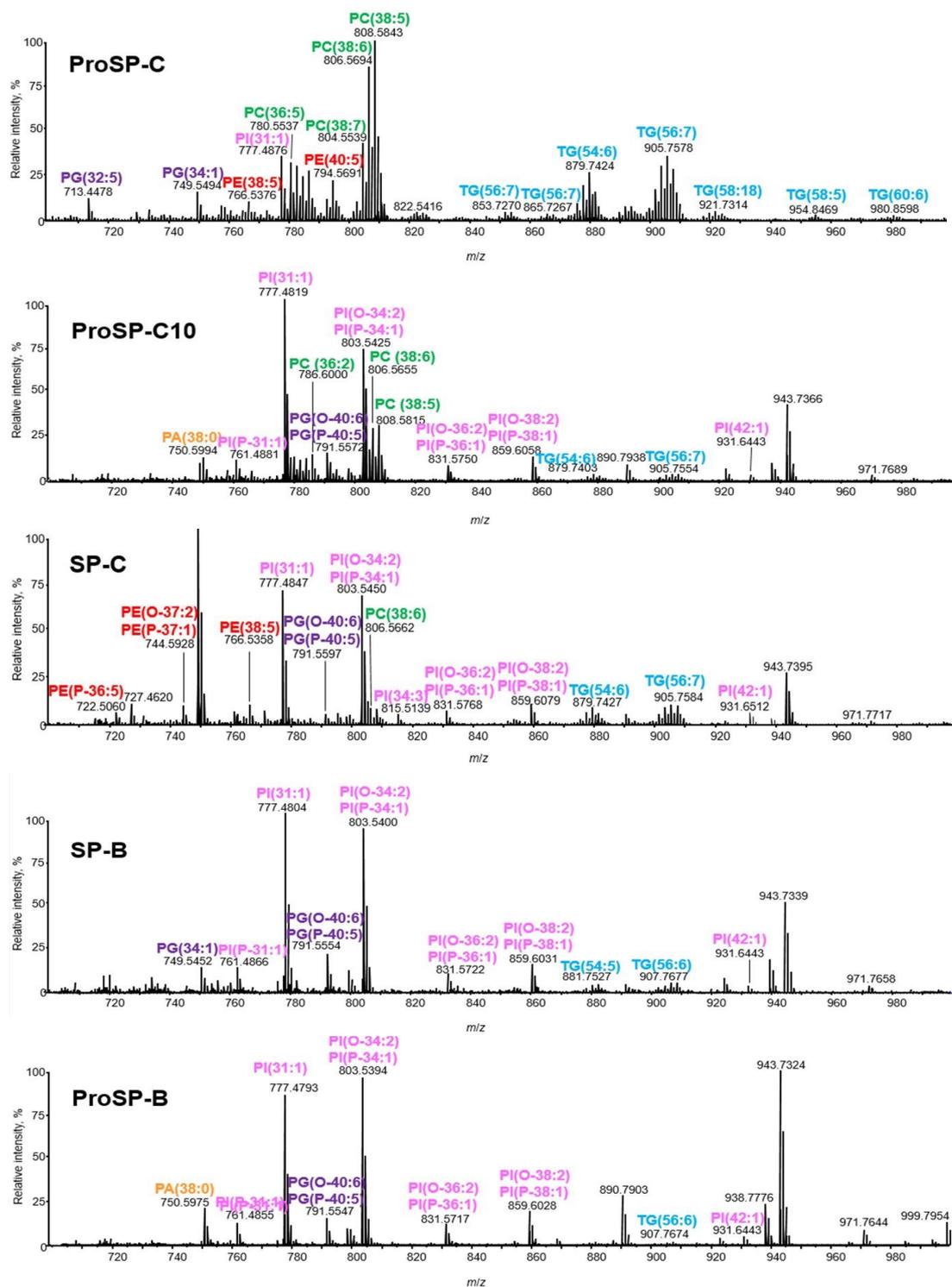
Again, the identification of PI around both proSP-B and SP-B may indicate the preference of these lipids for the full length and mature proteins.

There were fewer TG species present.

An interesting point to note is the similarity between the phospholipid compositions of SP-B and SP-C.

This could be further supporting information that SP-B and SP-C work in an interactive manner. The mass spectra, see Figure 5.2 below, evidence this.

Note: PI is an important lipid involved in signalling. Such lipids may initiate or receive signals to reorganise the membrane lipids.



**Figure 5.2 Mass Spectrometry Analysis of Lipids Around ProSP-C, ProSP-C10, ProSP-B, SP-B and SP-C**

Lipids around proSP-C were used as a control for the SMALP extracted lipids in this analysis.

The phospholipid types identified were POPC, POPA, POPG, POPI, POPE, and TG. These have been abbreviated in the spectra. There were also seven lipid species that were unable to be identified.

## 5.8 Further Discussion and Evaluation of Lipid Associations with Recombinant Surfactant Proteins

This overall study is the first which has attempted to examine the lipid environment of SP-B and SP-C within a yeast membrane.

Identified lipids were used to form 'membrane models' in which to study the behaviour of SP-B and SP-C.

A greater understanding of the lipid context should help to contribute to future improvement of therapeutics involving SP-B and SP-C.

Pulmonary surfactant is similarly composed of phospholipids that are enriched predominantly with PC and smaller amounts of PG, PE and PI (Georke 1992).

The mass spectrometry analyses identified the same phospholipids in the *Pichia pastoris* membrane as the pulmonary surfactant lipids.

Interestingly, the lipids encapsulated from each of the samples had small differences between full length and mature constructs, with larger differences between the proteins.

PG, PE, PI and PC were isolated from ProSP-C10 and SP-C, again this is the same as the phospholipids in the pulmonary surfactant.

As PC makes up a large proportion of lung phospholipids, it is surprising that PC and PE were not evident in lipids from around ProSP-B or SP-B.

PG has been suggested previously to interact with SP-B (Baatz et al 1990; Gill et al 1995); the work in this thesis which has identified PG from around ProSP-B and SP-B, supports these findings.

Speculation of the interaction between the charged N-terminus of SP-C with anionic phospholipids (Gill *et al.* 1995) is also supported by these findings.

Of these anionic phospholipids, this study also finds PI to be the dominant phospholipid interacting with SP-C.

PI was present in all samples, in fact, the identification of PI distinguished all purified samples. The control sample, Pro-SP-C, was free of PI.

PI accounts for a small percentage of both the *Pichia pastoris* (Grillitsch *et al.* 2014) and phospholipid membranes (Perez-Gil, Casals & Marsh 1995), suggesting the interactions with PI could be preferential.

***This is the first study proposing SP-B and SP-C interact preferentially with PI.***

Figure 5.3 also shows an overlap of PI and PG phospholipids between SP-B and SP-C.

This, together with earlier observations, add to the support of theories indicating the two proteins to be working in concert.

All lipid species are shown in Table 5.1 and as part of Venn diagrams below in Figure 5.3.

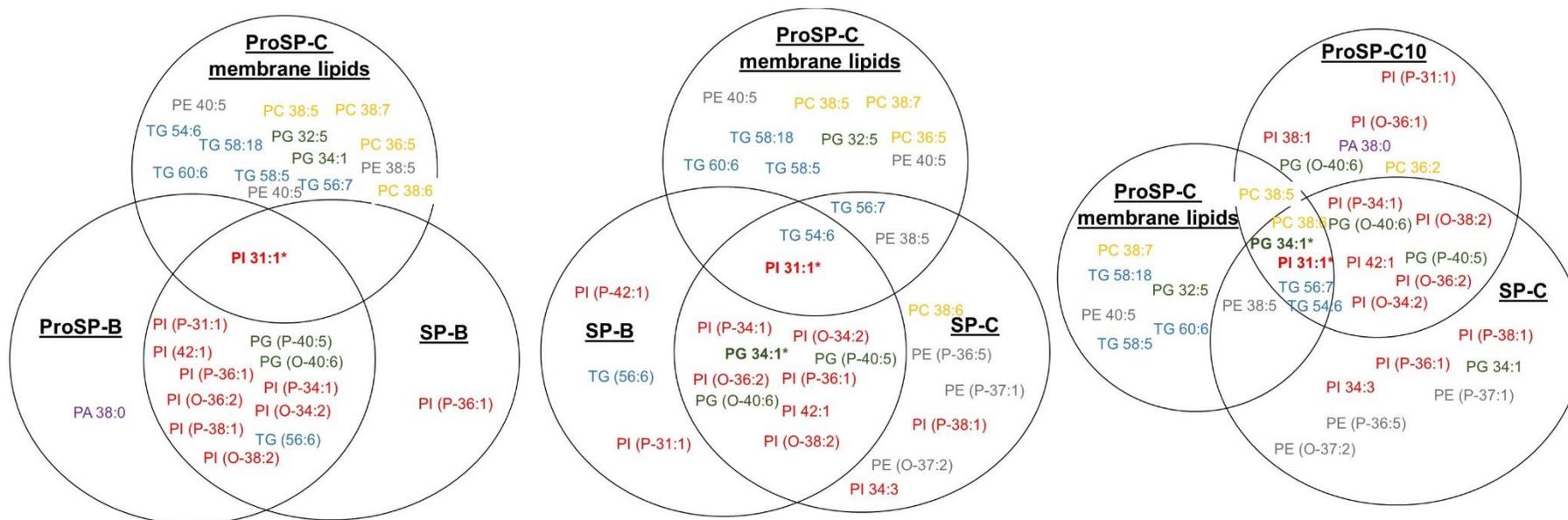
The identification of PC, PG, PE, PI and PC around proSP-C10, SPC, ProSP-B, SP-B, and their reported presence in the lung, strengthens the validity of the 'model membrane' they will be used to generate.

**Table 5.1: Identified Phospholipids from Mass Spectrometry**

Five different groups of phospholipids PI, PC, PG, PA and PE were notably present, along with triglycerides (TG) in every sample. Full names of the phospholipids can be found in the abbreviations list. PI 31:1 could be identified in every sample including the control.

The \* indicates phospholipids found in the control sample and other samples.

<b>ProSP-C (Control)</b>	<b>ProSP-C10</b>	<b>SP-C</b>	<b>ProSP-B</b>	<b>SP-B</b>
PC 38:5	<b>PI 31:1*</b>	<b>PI 31:1*</b>	PI (O-34:2)	PI (O-34:2)
PC 38:6	PI (O-34:2)	PI (O-34:2)	PI (P-34:1)	PI (P-34:1)
PC 36:5	PI (P-34:1)	PI (P-34:1)	<b>PI 31:1*</b>	<b>PI 31:1*</b>
PC 38:7	PC 36:2	PE (O-37:2)	PA 38:0	PI (P-31:1)
<b>PI 31:1*</b>	PC 38:6	PE (P-37:1)	PI (P-31:1)	PG (O-40:6)
TG 56:7	PC 38:5	PE 38:5	PG (O-40:6)	PG (P-40:5)
TG 54:6	PG (O-40:6)	PG (O-40:6)	PG (P-40:5)	PI (O-36:2)
PE 40:5	PG (P-40:5)	PG (P-40:5)	PI (O-36:2)	PI (P-36:1)
PE 38:5	PI (P-31:1)	PC 38:6	PI (P-36:1)	PI (O-38:2)
PG 32:5	PI (O-36:2)	PE (P-36:5)	PI (O-38:2)	PI (P-38:1)
PG 34:1	PI (O 36:1)	PI 34:3	PI (P-38:1)	TG 56:6
TG 58:18	PI (O-38:2)	PI (O-36:2)	TG (56:6)	PI 42:1
TG 58:5	PI (P-38:1)	PI (P-36:1)	PI (42:1)	TG 54:5
TG 60:6	PA 38:0	PI (O-38:2)		PG 34:1
	PI 42:1	PI (P-38:1)		
	TG 54:6	TG 54:6		
	TG 56:7	TG 56:7		
		PI 42:1		



**Figure 5.3 Lipid Species Identified by Mass Spectrometry Represented in Venn Diagrams**

Lipids present in all samples were PI 31:1.

There are several overlapping lipid species in SP-B and proSP-B: PI 31:1 and (O-34:2), PI (P-34:1), PG (O-40:6), PG (P-40:5), PI (O-36:2), PI (P 31:1), PI (P-36:1), TG 56:6 and PI 42:1.

The major lipids present in the proSP-C control, SP-C and SP-C10 were PC (38:5), PC (38:6), PI (P-31:1) and TG 56:7.

There is also overlap between SP-B and SP-C, the species identified included: PI (O-34:2), PI (P-34:1), PG (P-40:5), PG 34:1, PI (O-36:2), PI (P-36:1), TG 56:6 and PI 42:1.

## **5.9 Conclusion - Analysis of the Lipid Context of Recombinant Surfactant Proteins: Using Mass spectrometry to assist Molecular Dynamics Simulations**

**This chapter has demonstrated for the first time the protein-lipid interactions of SP-B and SP-C in membranes belonging to *Pichia pastoris*.**

PI has been found to be present in both full length and mature SP-B and SP-C samples

This is supporting evidence of the two proteins influencing lipids together.

The lipids observed around SP-B and SP-C were then used in explicit membranes in the modelling work following in Chapter 6.

## Chapter 6

### **Molecular Dynamics Simulations of SP-B and SP-C**

Developing from Chapter 5 to help further understand the involvement of SP-B and SP-C in tear film lipid spreading at the molecular level, molecular modelling was then pursued.

The model simulations were led by John Simms (now Coventry University), with the simulation analysis and consequential structural and contextual analysis from myself.

The latter involved for instance, understanding the structural application of the models and simulations, with personal considered interpretation in respect of; lipid spreading, structure and functional efficacy

This chapter, uses models of SP-B and SP-C which have been generated to bridge the lack of structural understanding for the proteins in the tear film.

Previously, Homology Modelling, have been used in attempts to understand protein structure and function.

Homology modelling is carried out by aligning proteins of a similar protein sequence, which has a known atomic structure

Using this to get a true representation of protein function, the sequence similarity should be at least 50% (Rost 1999)

SP-B has previous homology models that are based on the structure of saposin B, but here the sequence similarity of saposin B was only 18% of that minimum 30% sequence similarity target (Khatami *et al.* 2016).

Models which were generated using this data may therefore be unreliable.

In light of the above, an alternative to homology models are those models generated using an *ab initio* folding method, see below.

The SP-B and SP-C models developed and now discussed were based on this method.

*Ab initio* models are based on the primary sequence of proteins.

Here, sections of the primary sequence are 'folded' using data based on the secondary structure of known peptides.

In the work reported in this Thesis, some 10,000 models were generated using membrane protein modelling software called 'Rosetta'.

This was then "refined" down to 100 models using 'Replica Exchange Monte Carlo' (REMC) simulation. REMC was applied in the presence of a hydrophobic region with lateral pressure and friction.

The final model was selected using an in-house developed scoring method (in publication Simms and Poyner).

The *ab initio* model methodology has been validated by X-ray crystallography data of a Receptor Activity-Modifying Protein (RAMP) (Simms *et al.* 2006).

Protein orientation was observed using an "Implicit" membrane. This is one which uses a hydrophobic area that mimics the phospholipid bilayer.

The function of SP-B was also studied in an "Explicit" membrane, which included specific phospholipids.

As discussed in Chapter 5, phospholipids that were identified around SP-B and SP-C in the *Pichia pastoris* membrane were used in Explicit membrane construction.

## 6.1 Modelling the Structural Dynamics of SP-B in Implicit Membranes

As mentioned previously, homology modelling has been used to examine the involvement of SP-B in lipid spreading (Khatami *et al.* 2016).

However, as already mentioned these models were based on low sequence similarity (18%), and which thus may well have resulted in an unrealistic final model.

To overcome this challenge of low sequence similarity, an *ab initio* model of SP-B was created.

As discussed in the introduction, generating a 'final model' involved:

- Sequential rounds of folding of the primary sequence of SP-B.
- A scoring process.

The SP-B primary sequence was found in 'UniProt', a protein database, (accession number: SP-B: P07988).

The *ab initio* model for SP-B comprised a structure made up of 5 alpha helices and is illustrated below in Figure 6.1.

Most importantly, the presence of 5 alpha helices is in agreement with the expected number of alpha helices for SP-B (Andersson *et al.* 1995).

Helix 1 (in yellow) represents residues 1-22; helix 2 (in red) represents residues 23-37; helix 3 (in green) represents residues 38-52; helix 4 (in purple) represents residues 53-59; whilst helix 5 residues (in blue) represents amino acids 60-79.

SP-B has 7 cysteine residues (see black), Cys8, Cys77, Cys11, Cys71, Cys35 and Cys46.



**Figure 6.1** *Ab Initio* Model of SP-B

SP-B partially folds into the 90 Å membrane creating a hairpin loop-type feature.

Helix 1 is in yellow and represents residues 1-22; helix 2 is in red and represents residues 23-37; helix 3, in green, represents residues 38-52; helix 4 residues 53-59 are in purple; helix 5 in blue represent amino acids 60-79.

SP-B has seven cysteine amino acid residues within its structure, see black Cys 11 to Cys 77.

SP-B also has 3 disulphide bonds within its dimeric structure.

Interestingly, the *ab initio* model also replicated the experimentally determined disulphide bonding pattern, but without that information having been part of the folding protocol (Johansson *et al.* 1991).

In addition, a further cysteine (Cys48) has been reported as the amino acid involved for dimer formation (Andersson *et al.* 1991).

The SP-B model generated was a monomeric model and here the Cys48 residue was surface exposed. This made it primed for dimer placement.

A 2D model of SP-B, illustrating the positions of the cysteine residues and the disulphide bonds they form, is shown below in Figure 6.2 (Khatami *et al.* 2016).



**Figure 6.2      2D Structure of SP-B**

SP-B has 7 cysteine residues (Cys8, Cys77, Cys11, Cys71, Cys35, Cys46 and Cys48).

Cys 8-77, Cys 11-71 and Cys 35-46 form disulphide bonds.

Cys48 is considered to be involved in an intermolecular disulphide bond with another SP-B monomer.

The 3 disulphide bonds in SP-B force the structure into a hairpin loop. (Image taken from (Khatami, Saika-Voivod and Booth 2016).

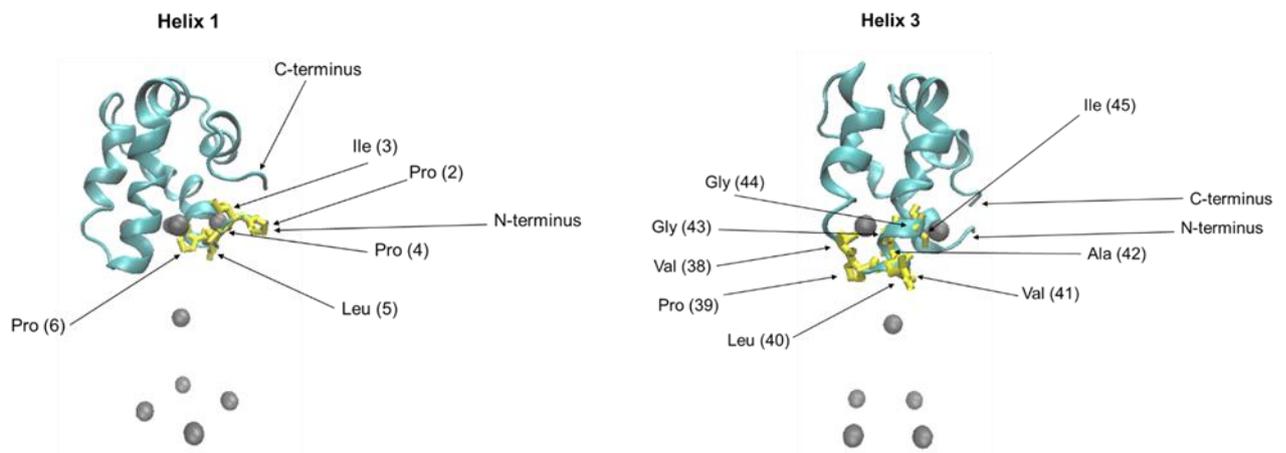
Membrane proteins can either span across the entire membrane, or be associated to a single side of the membrane leaflet only.

To be able to understand how SP-B was orientated with respect to the membrane, it was added to an implicit membrane (90 Å) and its behaviour then examined in simulation.

Understanding the orientation of SP-B within the membrane both helped to support the robustness of the *ab initio* model, and also aim to confirm current theories of SP-B orientation.

Currently it is thought that SP-B is associated with the membrane heads, but there is previous research that suggests it may be interacting with the acyl chains.

Snapshots of the Implicit membrane simulation are represented in Figure 6.3 below.



**Figure 6.3 Amino Acid Interaction of Helices 1 and 3 with the Membrane**

The hydrophobic amino acids at the tip of helix 1 and 3 are made up of Val, Pro, Leu, Ile, Ala and Gly; positions of the amino acids within the overall SP-B structure are illustrated.

This association of the amino acids with the membrane may help to explain the peripheral orientation of SP-B.

Note: The numbers in brackets refer to the amino acid number.

The Implicit membrane is represented by the grey dots, and represents an environment that gets progressively hydrophobic from the top of the membrane up to the central grey atom. After this point, the hydrophobicity reduces.

The areas of increasing and then decreasing hydrophobicity aims to simulate the conditions in an *in vivo* membrane.

A further advantage of this type of simulation is that it is computationally rapid.

Note: As the implicit membrane uses an area of hydrophobicity (hydrophobic slab), rather than including specific lipids, this approach also helped make the overall process more rapid.

A simulation of the SP-B model and the Implicit membrane was then carried out to examine the protein orientation along with the adopted conformations.

The resultant simulation is a 47 s simulation that begins with SP-B partially associated with the membrane.

Part of the structure then inserts into the membrane and returns back to be fully associated with the membrane.

Snapshots of the simulation are further represented below in Figure 6.4.

The starting conformation of SP-B is represented in a closed conformation and is partially associated with the top of the membrane.

Hydrophobic amino acids at the tips of helices 1 and 3 and are associated with the membrane.

An illustration of the location of these amino acids with respect to the membrane is given below, in Figure 6.3. Numbers in the illustration relate to the amino acid number.

Helix 3 (green) and helix 1 (yellow) are partially inserted into the membrane and these regions are made up of hydrophobic amino acids. The sequence of amino acids that make up helix 3 (38-52) are Val (38), Pro (39), Leu (40), Val (41), Ala (42), Gly (43), Gly (44), Ile (45), Cys, Gln, Cys, Leu, Ala, Glu and Arg.

The tip of helix 1 (2-22) is composed of: Pro (2), Ile (3), Pro (4), Leu (5), Pro (6), Tyr, Cys, Trp, Leu, Cys, Arg, Ala, Leu, Ile, Lys, Arg, Ile, Gln, Ala, Met and Ile.

Hydrophobic amino acids prefer hydrophobic environments, helping explain the partial insertion of SP-B in the membrane, see below in Figure 6.4 panel - A).

Figure 6.4 - B represents the open, expanded state where helix 3 (green) forms part of a hairpin loop.

Visual inspection of the simulation revealed that SP-B spontaneously inserted on 3 occasions during the simulation.

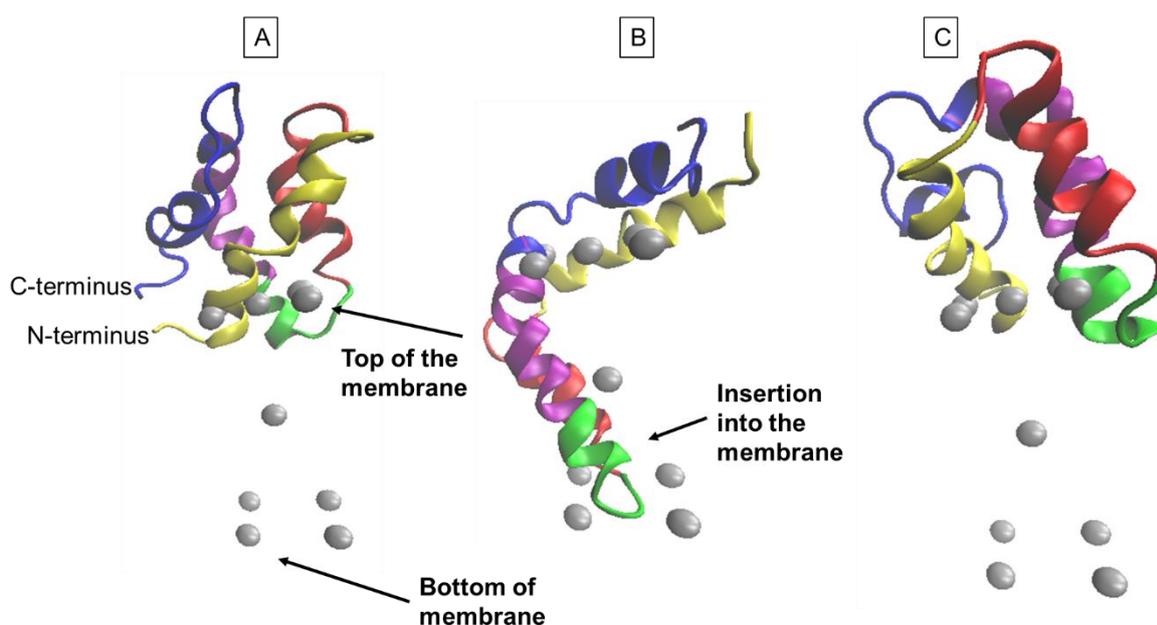
Further, whilst the simulation lasted 47 seconds, SP-B was inserted for a duration of 23 seconds.

The rapidity at which the process of alternating between a peripheral and a partly integral protein occurs, suggests that the free energy of insertion is small (~10kcal/mol).

It is thought that the lipid composition itself may provide an energy barrier to control this (Khatami *et al.* 2016).

Helices 2 (red) and 4 (purple) are fully inserted in the membrane; the amino acid residues around the tip of the helix are Arg, Val, Pro and Leu, Gly and Ala which are all hydrophobic amino acids.

SP-B then reverts back to the compressed state, with helices 1 and 3 partially inserted into the membrane again, see below in Figure 6.4 panel C.



**Figure 6.4 Snapshots of the Initial and Final Conformations of SP-B in an Implicit Membrane**

SP-B was oriented in closed conformation on top of the membrane (panel A);

The hairpin loop structure with helices 1 (yellow) and 3 (green) partially insert into the membrane (panel B).

In this intermediate conformation, the N-terminus hairpin loop (helix 1) of SP-B extends out, whilst the other loop (helix 3) fully inserts into the membrane.

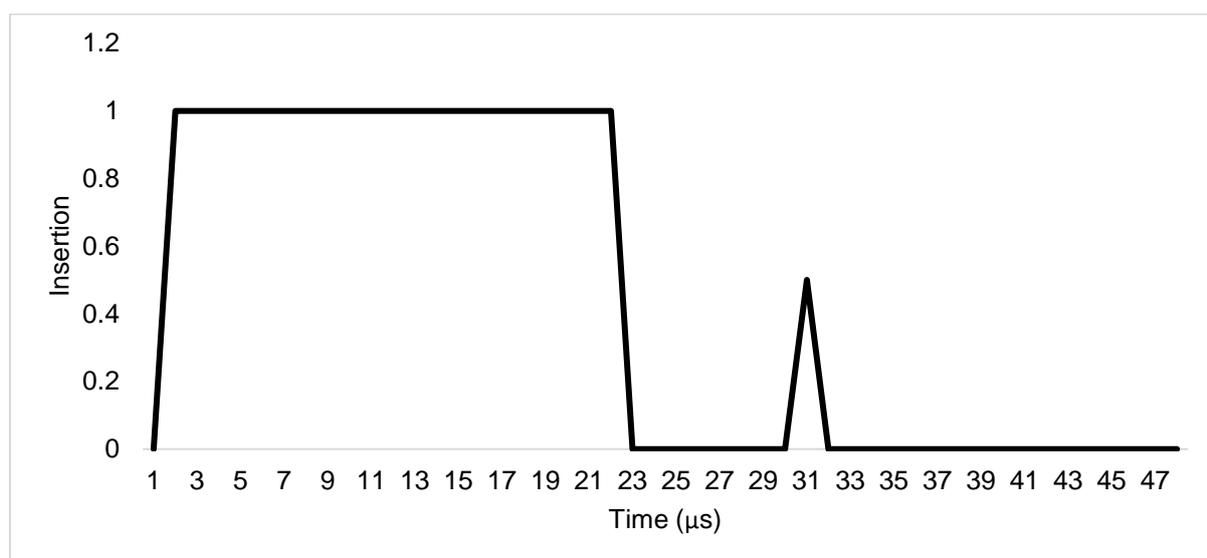
SP-B remains in this conformation for just above half of the simulation, before reverting to the closed conformation (panel C).

SP-B moves from a peripheral association to being partially (approx. 50%) inserted into the Implicit membranes, as shown in Figure 6.4 above.

This insertion of SP-B into the membrane has been previously reported (Perez-Gil *et al* 1995), as has the association with the membrane (Cabre *et al* 2012).

Perhaps the dynamic nature of SP-B structure had not been previously considered because our current computational modelling indicates that either conformations are valid.

The insertion of SP-B has been plotted on a graph to demonstrate the time (microseconds) spent inserted fully into the membrane and the time it is associated to the top of the membrane. A value of 1 represents the full insertion of SP-B into the membrane, 0.5 represents a partial insertion, see below in Figure 6.5.



**Figure 6.5 SP-B Membrane Insertion**

The insertion of SP-B has been plotted against time (microseconds), to represent the amount of time spent by SP-B either inserted into the membrane as compared to the amount of time it is associated to the top of the membrane.

A value of 1 represents the full insertion of SP-B into the membrane, 0.5 represents a partial insertion.

SP-B is fully inserted into the membrane for 23 μs, partially inserted for 2 μs and is associated with the membrane for the remainder of the time.

The Implicit membrane modelling data, shown above in Figure 6.4, began with a superficial orientation of SP-B, where it is partially inserted into the membrane.

SP-B was then fully inserted into the membrane before reverting to the superficial orientation.

The *ab initio* model of SP-B also supports other data suggesting SP-B is permanently associated with membranes (Olmeda, García-Álvarez and Pérez-Gil 2013), and also that it has a superficial orientation (Cabré *et al.* 2012).

Interestingly, there are also suggestions SP-B is able to “permeabilise” membranes (Parra *et al.* 2013) as well as perturb acyl chains (Perez-Gil, Casals and Marsh 1995).

The insertion of SP-B, see above in Figure 6.4 panel – B, further supports this.

***The ab initio model of SP-B confirms the association of SP-B with phospholipid membranes but also provided evidence of SP-B insertion into the membrane.***

***The role of SP-B insertion in membranes can help to further understand its role in the tear film.***

## 6.2 Modelling the Structural Dynamics of SP-C in Implicit Membranes

An *ab initio* model of SP-C was also then generated using the same methodology as previously described for SP-B.

SP-C is palmitoylated at the N-terminus (two chains of palmitic acid are attached to the Cysteine residues, Cys5 and Cys6).

SP-C is very hydrophobic; the residues in green, see below in Figure 6.6 are hydrophobic amino acids that form a helix.

Interestingly, SP-C contains two stretches of Val residues, which are known to be helix breaking, in its sequence (Berg, Tymoczko and Stryer 2002).

The stretch of red Val residues seen in Figure 6.6 are in the middle of the SP-C primary amino acid sequence.

Very importantly, the *ab initio* model of SP-C is also in agreement with the NMR structure of SP-C (Figure 1.14) (Johansson *et al.* 1994).



**Figure 6.6** *Ab Initio* Model of SP-C

SP-C is a 35 amino acid peptide that forms an alpha helical secondary structure (Johansson *et al.* 1994).

It is palmitoylated at residues Cys5 and Cys6, at the N-terminus.

A high proportion of the amino acids that make up SP-C are hydrophobic; indicated by the green and red residues on the SP-C model.

The red residues represent the Val residues that make up the helix.

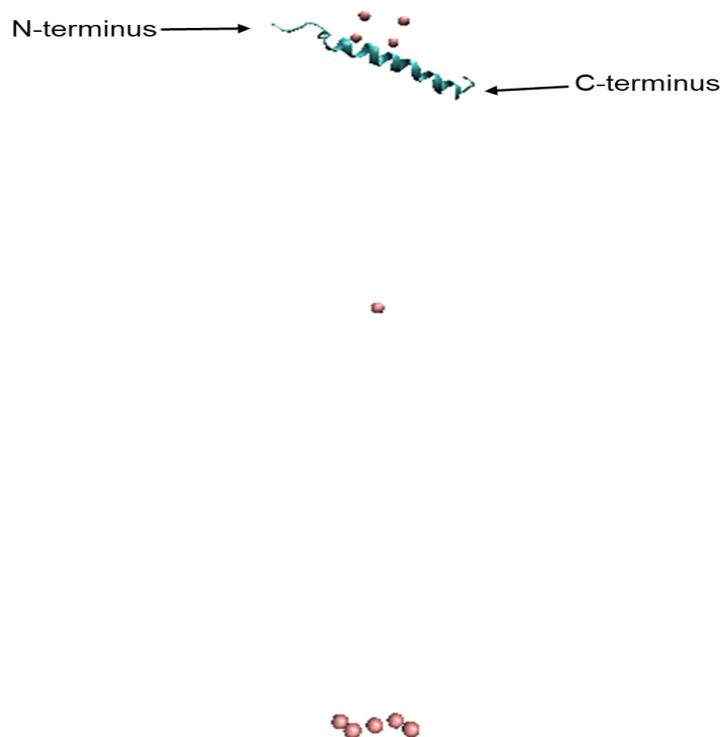
As with the SP-B model, the orientation of SP-C with respect to the membrane and the conformations it adopts, was simulated.

The Implicit membrane had a membrane depth of 90 Å; to simulate the size of the tear film lipid layer.

Thus the size of the Implicit membrane reflects the *in vivo* conditions.

This model representation of the orientation of SP-C with respect to the membrane is given below in Figure 6.7.

The model representation of SP-C indicates a tilting of the protein within the membrane.



**Figure 6.7 SP-C Membrane Orientation**

The *ab initio* model of SP-C was inserted into an implicit membrane to examine the orientation of the protein.

The size of the membrane reflects the size of the non-polar lipid layer in the tear film (90 Å).

The transmembrane region of SP-C is fully inserted in the hydrophobic portion of the membrane and sits at a tilt.

The N-terminus is associated with the membrane.

On closer examination of a simulation of SP-C in an Implicit membrane, SP-C begins in an extended conformation, see below in Figure 6.8 panel -A.

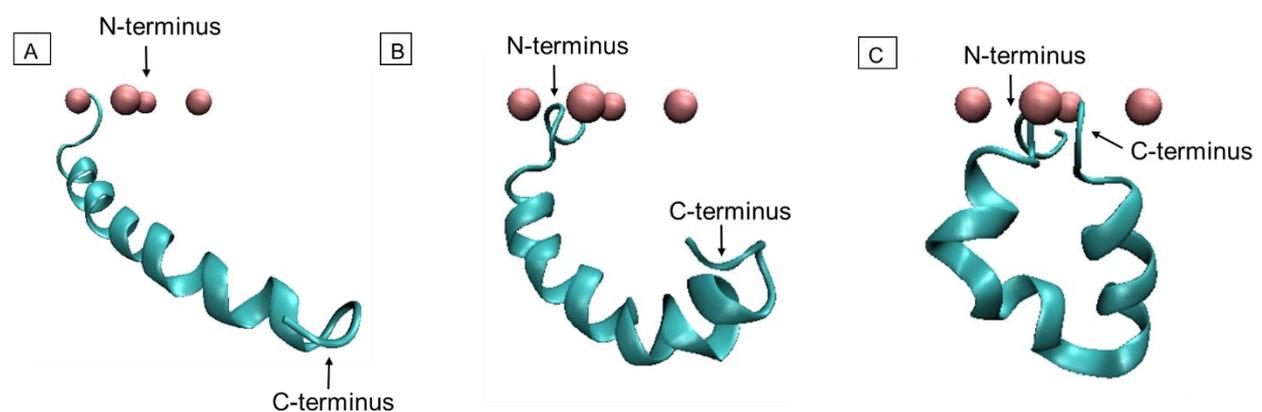
The N-terminus region is charged and remains associated with the phospholipid head groups, again see Figure 6.8 panel -A.

The next action has not been documented previously, namely, here the **C-terminus region bends inwards until it is close to the N-terminus**, see below in Figure 6.8 panel – B.

Further originality is seen via the structure **forming a V shape**.

At this closed conformation, the C-terminus becomes associated with the head groups.

The bending is caused by the branching of valines at the beta carbon atom. To offset any steric clashes in the alpha helices that this may cause, the Val molecules form a bend, see below in Figure 6.8 panel - C.



**Figure 6.8 Snapshots of the Initial and Final Conformations of SP-C in an Implicit Membrane**

The N-terminus of SP-C is closely associated with the membrane. It adopts an extended tilted conformation in the membrane (panel A).

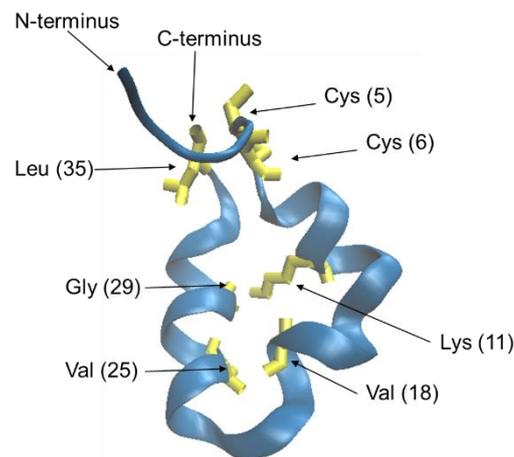
The extended conformation starts to bend inwards from the C-terminus (panel B). The SP-C structure forms a bend at the region rich in valines (panel C).

When the structure of SP-C closes in on itself, Leu (35) at the C-terminus comes into close contact with Cys (5) and Cys (6) at the N-terminus.

These are both hydrophobic amino acids. Lys (11) is close to Gly (29), which are also each hydrophobic amino acids.

Val (18) and Val (25), which are small aliphatic hydrophobic amino acids, also come into close contact.

Figure 6.9, below illustrates the positions of the highlighted amino acids and their interactions.



**Figure 6.9 Close Amino Acid Interactions of SP-C in 'V' Conformation**

A 'bending' of SP-C can be seen between positions 18-25, all of which are made up of Val.

Steric clashes between the Val amino acids induce a bend in the SP-C structure.

The interactions of amino acids which come into close contact, as a result of the bend in SP-C structure, have been labelled in the above illustration.

Once again, hydrophobic amino acids such as Val, Gly, Leu, Cys and Lys are present.

***These snapshots illustrate SP-C in a 'V' shaped conformation, and represent the potential orientation of SP-C in the tear film; this is work that has never been carried out previously.***

Other models of SP-C have examined the behaviour of SP-C within a bilayer, however the tear film polar layer is a monolayer.

Thus, the above snapshots give tremendous insight into how SP-C might be functioning in the tear film.

SP-C adopts an expanded conformation in the tear film polar layer, wherein SP-C is believed to be associated.

The extended and closed V shaped conformations reflect the role of SP-C during compression and expansion of the tear film.

The extended conformation of SP-C are also likely be involved in helping to anchor vesicles of phospholipids that have been expelled from the membrane during the expansion phase.

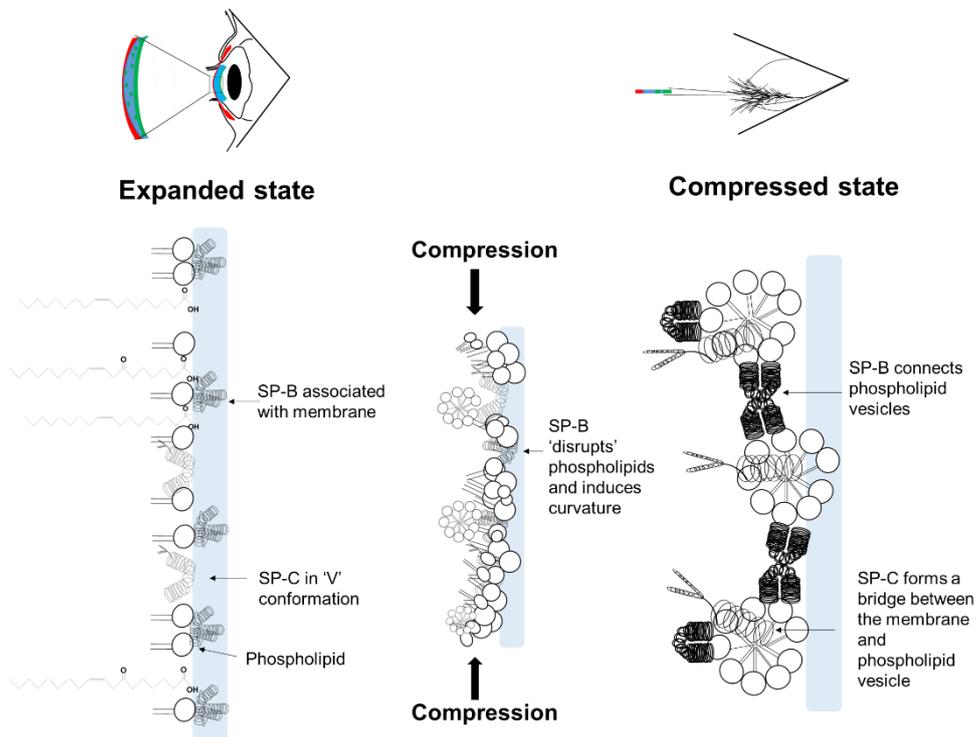
Once the vesicles have re-spread, SP-C may again adopt a V shaped conformation.

***These are the first computational images that examine the behaviour of SP-C in a membrane which mimics the tear film.***

Implicit membranes help to give insight into the different conformations that a protein can adopt in a membrane. Whilst these Implicit membranes are devoid of lipid molecules, Explicit membranes use specific lipids.

To gain additional, more detailed information, the behaviour of SP-C was examined in an Explicit membrane.

To add further context and overall understanding of the total picture, it would seem appropriate at this stage to provide a figure, see Figure 6.10 to demonstrate at the high level, the key interactions discussed throughout this Chapter.



**Figure 6.10 Schematic Illustration of SP-B and SP-C in Tear Film Lipid Spreading**

In the expanded state, SP-B is associated with phospholipids that form the membrane. SP-C adopts a V shaped conformation and is orientated adjacent to the acyl phospholipid tails.

Tear film break up initiates blinking, leading to compression of the tear film. SP-B induces phospholipid curvature leading to vesicle formation.

In the compressed state, phospholipid vesicles are held together via SP-B.

SP-C forms a bridge between the membrane and the phospholipid vesicle. Through their interactions with phospholipid vesicles and the membrane, both SP-B and SP-C help to stabilise the membrane.

This indicative schematic attempts to explain the influence of SP-B and SP-C on tear film phospholipids.

Conjecture about the final compressed state re the role of phospholipids may well over anticipate the level of phospholipid micellar formation because the level of reported phospholipids is so low (Campbell, Griffiths & Tighe 2011).

Alternatively the phospholipids may in reality be 'super' compressed, with SP-B and SP-C sitting interspersed amongst the phospholipids heads and tails, and facilitating re-spreading.

### 6.3 Modelling the Structural Dynamics of SP-B and SP-C in Explicit Membranes

To be able to get an accurate representation of the behaviour of SP-B and SP-C in the tear film, it was important to try to use a phospholipid composition that reflected the *in vivo* polar layer composition.

However, there is a lack of agreement on the precise lipid composition of the tear film polar layer, see earlier in Table 1.2 in Chapter 1.

This obviously made it rather challenging to select a lipid composition for Explicit modelling simulations.

To overcome the problem of not having a defined tear film polar layer composition, the phospholipid composition of *Pichia pastoris* yeast was substituted instead.

Extracted lipids from around SP-B and SP-C, using SMA 2000P polymers, were analysed using mass spectrometry.

Explicit models were based on concentrations composed of phospholipids (Grillitsch *et al.* 2014).

Previous studies have speculated about the type of phospholipid that SP-B and SP-C are interacting with, but **this work includes the first study to examine the actual interactions, albeit in a yeast membrane.**

The interactions of SP-B and SP-C were examined in Explicit period simulations which included POPS, POPE, POPC, and POPI.

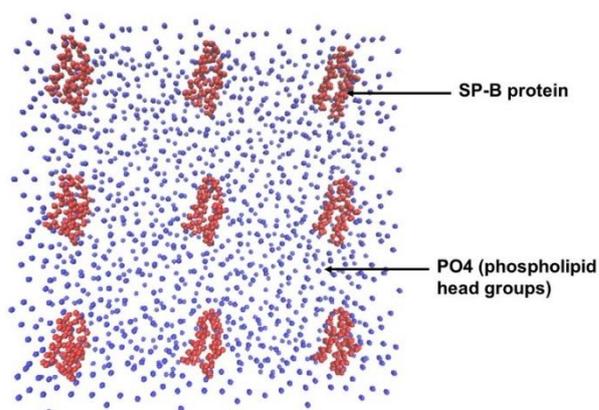
Period simulations were used to look at the interactions of SP-B and SP-C with POPI, POPS, POPE and POPC.

Such Period simulations allow protein-lipid interactions to be examined.

The central SP-B atom represents the simulation, the other proteins around the central SP-B protein are mirror simulations.

Conditions called periodic boundary conditions are set, whereby the number of molecules, temperature and pressure are all programmed to stay within a set limit, so aiding simulation conditions in the membrane.

The period simulations of SP-B and SP-C are illustrated below, in Figure 6.11 and 6.12



**Figure 6.11 Period Simulation of SP-B**

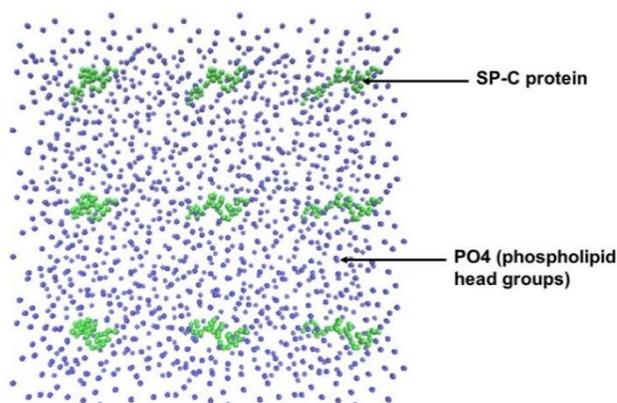
Periodic simulations allow protein-lipid interactions to be closely examined.

The central SP-B atom represents the simulation, the other proteins around the central SP-B protein are mirror simulations.

SP-B adopts a superficial position on top of the membrane.

SP-B protein is represented by the molecules in red, phospholipid head groups are represented by the blue molecules.

The period simulation of SP-C is shown below.



**Figure 6.12**      **Period Simulation of SP-C**

The central SP-C protein represents the simulation, the other proteins around the central SP-C protein are mirror simulations.

SP-C adopts a superficial position on top of the membrane.

SP-C protein is represented by the molecules in green, phospholipid head groups are represented by the blue molecules.

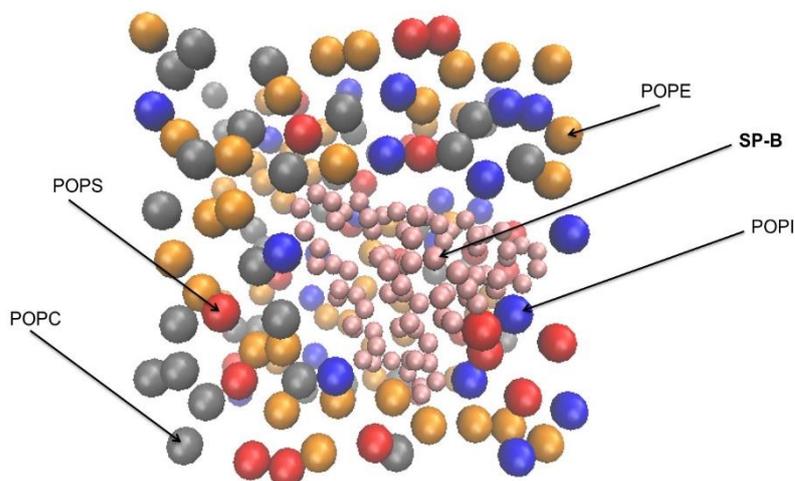
Mass spectrometric analysis of phospholipids around SP-B identified PI and PG lipid species (see results in Chapter 5 section 5.3.5).

The distance of the phospholipids from the SP-B protein has been used to support the lipid analysis carried out using mass spectrometry.

The protein-lipid interactions of SP-B indicate POPI is closely associated with SP-B.

Also present in a slightly greater abundance is POPE, see below in Figure 6.13.

POPC is present in similar levels to POPI but is located more externally to SP-B than POPI and POPE. POPE was not identified in the mass spectrometry analysis.



**Figure 6.13 Protein-Lipid Interactions of SP-B**

An explicit membrane composed of POPI (blue), POPS (red), POPE (orange) and POPC (grey) was used to examine the lipid-protein interactions of SP-B.

Examination of the lipids around SP-B in this membrane illustrate a close interaction with POPI and POPE.

Mass spectrometry analysis of SP-C identified POPE, POPI, POPC and POPG from SMALP extracted lipids.

Explicit simulations illustrate POPI and POPE are closely associated with SP-C, see below in Figure 6.14.

POPC and POPS also interact, but not as frequently as POPI or POPE.

PS was found to be present in SP-C samples from mass spectrometry.

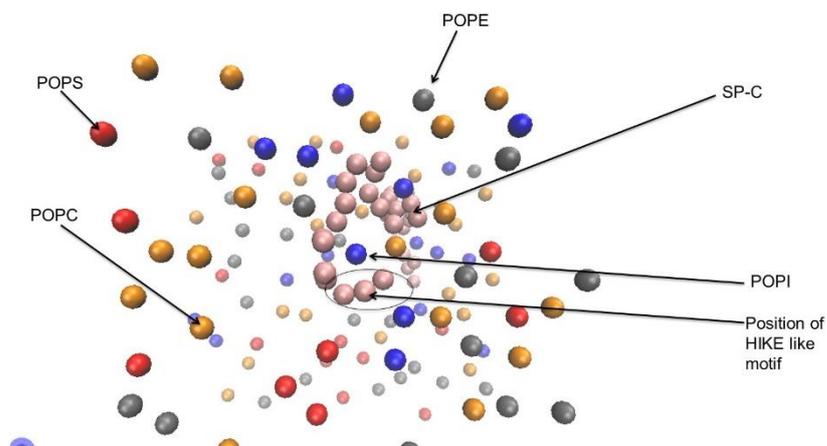
Explicit membranes did not reflect POPG species found from the mass spectrometry, but this is one an area where additional lipid analysis would give value.

A HIKE motif made up of: His, Iso, Lys and Glu.

It has been found to be present in SP-C as a conserved protein domain that can bind to POPI molecules (Ciccarelli, Acciarito and Alberti 2000).

The presence of a **HIKE domain** is a novel finding, and it helps to provide yet more information about how SP-C may be involved in lipid modulation.

The findings also support the high abundance of PI in the lipid analysis.



**Figure 6.14 Protein-Lipid Interactions of SP-C**

An explicit membrane composed of POPI (blue), POPS (red), POPE (orange) and POPC (grey) was used to examine the lipid-protein interactions of SP-C.

POPC and POPS also interact, but not as frequently as POPI or POPE.

POPS was found to be present in SP-C samples from mass spectrometry.

A HIKE motif was discovered in SP-C, these motifs are known to bind POPI lipids.

SP-B causes modulations to membranes.

To understand how this affects membranes, it is important to observe the membrane in the absence of any proteins.

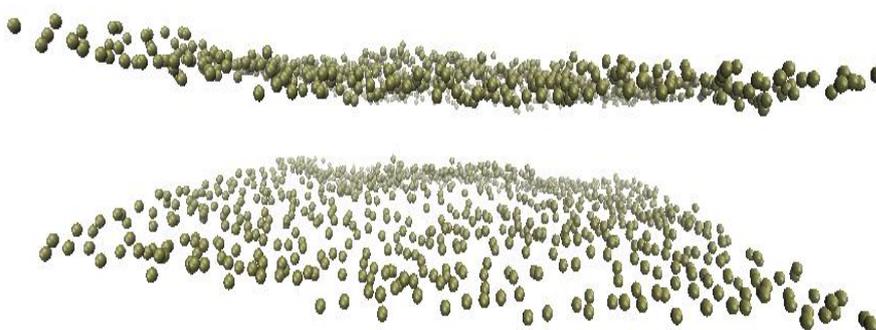
Coarse grain simulations, using phospholipids identified from mass spectrometry analysis, were incorporated.

The types of phospholipids used in explicit membranes were therefore based on mass spectrometry analysis, and the following phospholipid abundances were:

POPC (30%), POPE, (30%), POPI, (15%), POPS (15%) and ergosterol (10%) (Grillitsch *et al.* 2014).

A lipid bilayer, with a smooth bilayer surface on both sides of the membrane, can be observed.

The lipid layer generated is shown below in Figure 6.15.



### **6.15 An Explicit Bilayer Membrane in the Absence of SP-B**

The following phospholipid abundances were used to generate the bilayer membrane: POPC (30%), POPE, (30%), POPI, (15%), POPS (15%) and 10% ergosterol (Grillitsch *et al.* 2014).

Small fluctuations in the surfaces of the bilayer can be seen.

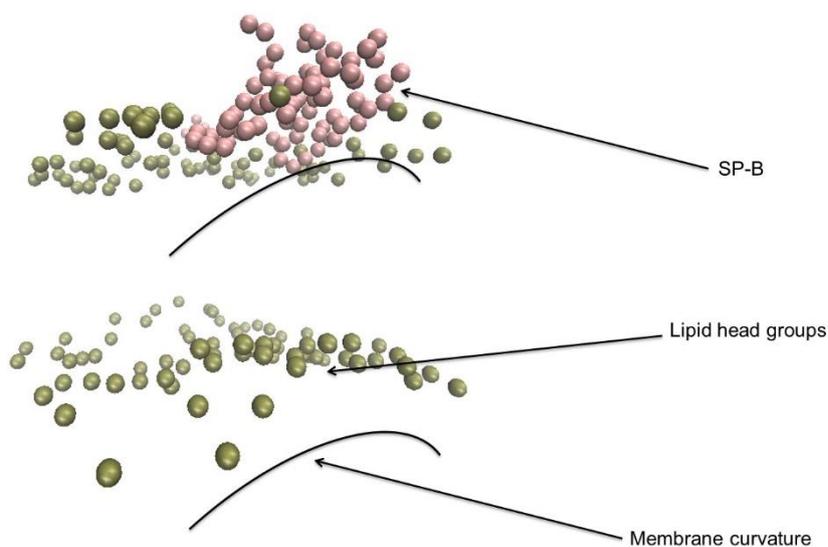
This represents the lipid membrane in the absence of any protein

Addition of SP-B to the same phospholipid membrane, see below in Figure 6.16, causes membrane curvature in the bottom leaflet of the bilayer.

This curvature is more pronounced in the presence of SP-B.

Such curvature has been reported by others (Parra *et al.* 2013), and importantly, the new SP-B model created in this work shows a similar activity in a membrane with a *different phospholipid composition* to that of the lungs.

Although the results are based on the effect of SP-B in a yeast phospholipid composition, these results do indicate that the effect of SP-B could be potentially transposed to the tear film.



**Figure 6.16 An Explicit Bilayer Membrane with SP-B**

The addition of SP-B to the phospholipid bilayer causes membrane curvature.

The curvature may be induced during compression, with the role of SP-B being to help stabilise the polar layer.

The bottom leaflet of the bilayer is curved due to the addition of SP-B.

These effects highlight the possible function of SP-B in the tear film lipid spreading.

The same phospholipid bilayer, composed of POPS, POPE, POPC and POPI, was used to generate an Explicit membrane with a bilayer height of (30 Å), see above in Figure 6.16.

Adding SP-C to the Explicit phospholipid membrane causes undulations of the lipid membrane at the top and bottom of the lipid bilayer.

This is demonstrated below in Figure 6.17 where SP-C can be seen pulling both sides of the membrane inwards, namely;

***SP-C appears to act as bridge between the two bilayer leaflets.***

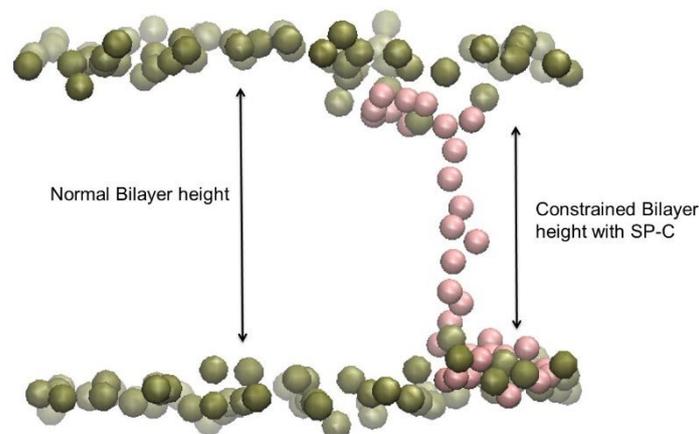
This model supports the reported function of SP-C being vital for helping to stabilise the membrane by anchoring vesicles to the membrane under compression.

It is probable that SP-C has this very function in the tear film also.

***The wider role of SP-C in the tear film has not been examined until now.***

These computational modelling images indicate SP-C is associated with phospholipids as well as showing that SP-C functions to modulate lipids.

Computational modelling demonstrates that this applies even in these different phospholipid compositions to those of the lung, see below in Figure 6.17.



**Figure 6.17 An Explicit Bilayer Membrane with SP-C**

SP-C spans the entire length of the membrane and acts as membrane bridge.

In an explicit membrane, in the absence of protein, the bilayer shown exhibits 'normal' height.

The addition of SP-C pulls the two bilayer leaflets in towards each other, and SP-C keeps a tight association between the two bilayer leaflets.

## **6.4 Considering the Function of SP-B and SP-C in Tear Film Lipid Spreading**

To date, the role of SP-B and SP-C in tear film lipid spreading had not been investigated.

This chapter outlines several novel discoveries relating to the orientation and function of SP-B and SP-C in the tear film lipid layer.

These findings have been incorporated in an previous schematic illustration (Figure 6.10 above) which conjectures how SP-B and SP-C *may well be functioning*.

The expansion and compression stages in Figure 6.10 reflect data from the Implicit and Explicit computation modelling simulations.

There remains conjecture about any particular detailed representation of the compressed state; namely the reported level of phospholipids in the tear film.

At this point, it may add value to discuss the implications of this modelling chapter (Chapter 6) alongside Chapters 3 and 4 and the previously reported understanding of tear film spreading.

## **6.5 Context of SP-B and SP-C in Tear Film Spreading**

Findings of this work are differentiated by the use of italics

Soluble mucins of the aqueous-mucin layer traverse the tear film, causing instability and tear film break up (Holly and Lemp 1977; Brown and Dervichian 1969).

This event causes a cycle of expansion and compression of the tear film over the duration of a blink.

The 'open eye' represents a relaxed or 'expanded' state, at this point the tear film is spread out over the entire ocular surface.

Once tear film break up is initiated, the lipid layer is compressed and expanded.

*SP-B adopts a 'peripheral association' and an 'inserting conformation' as shown in Figure 6.4, above.*

*The peripheral SP-B conformation and its association with phospholipids may represent the expanded state.*

*In this state, SP-C adopts a V shaped conformation and is orientated adjacent to the phospholipid acyl tails.*

Tear film break up initiates blinking, leading to compression of the tear film.

*SP-B may have several functions during lipid spreading.*

*During 'compression' of the lipid layer, the 'inserting conformation' may have a disrupting effect, inducing phospholipid curvature, leading to vesicular formation.*

*Eventually, at the point of maximum compression, with the eyes being fully closed phospholipid vesicles are formed, these vesicles now being held together via SP-B.*

*SP-C forms a bridge between the membrane and the phospholipid vesicle.*

*The association of SP-B between vesicles is further reinforced by SP-C forming a bridge between the membrane and the phospholipid vesicles in the compressed state.*

*Through their interactions with phospholipid vesicles and the membrane, both SP-B and SP-C help to stabilise the membrane.*

*During 'expansion' of the lipid layer, the 'peripheral conformation' of SP-B may encourage re-spreading of the polar layer.*

*The re-spreading may also be aided by SP-C in its 'V' shaped conformation.*

*This explanation is based on the findings from this and previous chapters.*

The range and novelty of this work helps to contribute more knowledge, understanding and limited conjecture of the mechanistic role of SP-B and SP-C in lipid spreading.

## 6.6 Conclusions - Molecular Dynamics Simulations of SP-B and SP-C

The first *ab initio* models of SP-B and SP-C have been reported in this chapter.

The *ab initio* model of SP-B is in agreement with published literature about the orientation of SP-B with respect to the membrane (Cabre *et al.* 2012) and disulphide bond formation (Johansson *et al.* 1991).

Protein-lipid interaction studies indicate POPI is closely associated with SP-B and SP-C.

A 'HIKE' domain was discovered in the SP-C structure and supports the strong association of SP-C with PI.

***Implicit simulation of SP-B indicates a simultaneous capacity to associate with, and insert into the membranes.***

***Explicit membrane models investigating SP-B function, demonstrated an ability to induce curvature.***

Simulations of SP-C in an implicit membrane demonstrated a novel 'V shaped conformation.

***This is the first demonstration of the mechanism of SP-C in the tear film and may indicate that SP-C adopts an extended conformation in the expansion state, (which helps anchor vesicles expelled from the membrane), and a closed V shaped conformation during compression.***

***Explicit membrane models support the function of SP-C as an anchor between the membrane and lipid vesicles (Lopez-Rodriguez and Pérez-Gil 2014).***

***Overall SP-B and SP-C help to stabilise the tear film layer via organisation of the phospholipids at the surface during compression, and re-spreading during expansion.***

## Chapter 7

### **Discussion Highlights and Overall Conclusions**

The overall aim of this thesis was to express proSP-B, SP-B, proSP-C, proSP-C10 and SP-C as recombinant proteins in *P. pastoris*; and to understand the role the proteins play in lipid spreading in the tear film.

This discussion, and reprise of resultant key conclusions, builds from the various, previous Chapters'; information, contributory discussions and conclusions.

These context the overall principal findings of this thesis.

#### **7.1 Production of Recombinant ProSP-B and SP-B**

##### **ProSP-B Expression**

ProSP-B has been successfully synthesised and expressed in *Pichia pastoris* yeast, for the first time as part of the experimental elements of this thesis.

Chapter 3 outlined the particulars of the experimental processes, beginning with the cloning techniques used to clone the gene encoding proSP-B into *P. pastoris* and on through to extraction.

Eurofins were able to independently confirm that the correct *proSFTPB* DNA sequence had been inserted into the pPICZαA vector.

ProSP-B aggregation and migration to different molecular weights were particular challenges encountered during the expression, separation and purification processes, albeit reported literature notes similar issues.

Culturing ProSP-B within bioreactors, rather than baffled shake flasks, led to a reduction in the amount of aggregation seen. This methodology, which allowed tighter control, provided conditions for the production of higher-quality protein.

For instance, the various different extraction agents employed to isolate ProSP-B from the membrane were a contributory factor in the differing molecular weights seen.

In detergent solubilisation, the surrounding annular lipids which have been shown to be important for the integrity of protein structure (Dawaliby *et al.* 2016) are replaced with detergent.

This process could well have resulted in proSP-B aggregation, as a result of fusion of hydrophobic domains with each other resulting in proSP-B remaining in the loading wells.

Solubilisation of proSP-B, using guanidine hydrochloride (GH) cited in a patent (US 6031075A), was therefore explored.

This method improved the extraction of proSP-B and yielded a protein band around 75 kDa.

Purification of proSP-B with GH and  $\beta$ -OG extracted a 35 kDa protein band.

This band was sent to the mass spectrometry department at Birmingham University to be identified. Resultant peptide analysis confirmed the presence of proSP-B.

Comparable differences seen in the proSP-B molecular weight has also been reported in other studies (Palacios *et al.* 2006).

Such differences have been an underlying theme, results and explanations of which are detailed in particular, in Chapters 3 and 4.

For instance, incomplete folding of proSP-B by SDS and a tendency for the protein to form higher molecular weight aggregation structures also affected proSP-B molecular weight determination.

## **7.2 SP-B Expression and Extraction**

As mentioned, successful integration of *SFTPB* in *P. pastoris* was followed by expression and scale up in bioreactors.

All these experiments mark the first reported successful synthesis in *Pichia pastoris* yeast.

SP-B extraction was carried out using detergents and SMA2000P polymer.

Circular Dichroism (CD) analysis of purified SP-B confirmed an alpha helical secondary structure.

Functional analysis using the Langmuir trough demonstrated the surface activity of SP-B in: precipitated protein form; in detergents, and as a protein-lipid mixture.

This is the first time that SP-B has been both successfully produced as a recombinant protein in any host; and shown to have relevant surface activity.

Previous reported attempts to overcome the hydrophobicity of SP-B used methods that involved fusing the N-terminus and mature SP-B regions to a maltose binding protein (MBP) in *E. coli* (Weaver 6031075 A). The authors cite very poor yields.

Another reported attempt at expressing SP-B as a fusion protein involved fusing SP-B to a urokinase domain and expressed in Chinese hamster ovary (CHO) cells. The authors again cited low yields but also little surface activity (Clemens *et al.* 2008).

### **7.3 Mass Spectrometry Analysis and Modelling of Lipids Around SP-B**

Analysis of phospholipids around SP-B, using mass spectrometry analysis, identified nine different species of PI.

Interactions with PI were also confirmed in computational modelling simulations.

PI has an overall negative charge (Cooper 2000) and may well form electrostatic interactions with SP-B as a result of its +7 positive charge.

The PI abundance in *P. pastoris* membranes is 4.20% (Grillitsch *et al.* 2014); this high abundance of reported PI is likely indicative of specific protein-lipid interactions.

Analysis of phospholipids around SP-B also identified two species of PG and two species of TG.

The interaction of SP-B with POPG has been previously reported (Batz, Elledge and Whitsett 1990); other studies failed to demonstrate any selectivity (Cabr  *et al.* 2012).

POPG has an overall net negative charge (Yeagle 2016). This also supports theories that positively-charged SP-B protein would prefer to interact with negatively charged lipids (Perez-Gil 2008).

## 7.4 Computational Modelling

### SP-B in Implicit Membrane

To explore the mechanism of SP-B in a lipid membrane, an Implicit computational model of SP-B was generated using an *ab initio* approach.

SP-B modelled in an Implicit membrane demonstrated that a possible mechanism of SP-B efficacy involved insertion of part of the hairpin loop structure.

This then remained associated with the membrane for a time, before reverting to an original position of sitting partially inserted into the membrane.

Some studies have also suggested SP-B forms pores in membranes (Parra *et al.* 2013).

The orientation of SP-B, on top of the membrane and insertion, has also recently been shown in a computational modelling study of SP-B (Khatami *et al.* 2016).

The permanent association of SP-B with the membrane; and with another conformation whereby the SP-B lies in the middle of the acyl chains, were both captured in Implicit model snapshots.

## 7.5 SP-B in Explicit Membrane

SP-B induced lipid curvature when added to an Explicit membrane.

SP-B is also thought to help connect lipid vesicles (Baoukina and Tieleman 2011).

Defects in lipid membranes caused by SP-B was also reported by (Chavarha *et al.* 2015);

The lung lipid layer under compression has shown evidence of forming vesicles (Bachofen *et al.* 2005).

SP-B helps to connect those formed vesicles from then moving away from the membrane (Ryan *et al.* 2005), so helping to stabilise and facilitate re-spreading.

Other recent studies that examined lipid spreading using computation models, observe a similar phenomenon to that seen in the lung lipid layer (Wizert *et al.* 2014, Cwiklik 2016).

## **7.6 Production of Recombinant ProSP-C and SP-C - ProSP-C Expression, Extraction, Purification and Functional Analysis**

This thesis also demonstrated the production of proSP-C and SP-C in *P. pastoris* for the first time.

Synthesis of *ProSFTPC* was confirmed via independent sequence analysis (Eurofins).

Expression of proSP-C was transferred to large scale (Bioreactor) production with the protein identified using Western blot analysis in the first instance.

Differences between the actual and expected weight were principally considered to be as a result of poor unfolding by SDS detergent.

ProSP-C samples were further independently confirmed using mass spectrometry analysis, confirming successful expression in *Pichia pastoris* for the first time.

ProSP-C was solubilised and purified using SMA2000P, this being the first reported isolation of proSP-C using polymers.

Detergent solubilisation of proSP-C revealed NLSS detergent to be the best detergent to extract proSP-C. However, purification using this detergent failed to produce samples of a suitable quality, despite several different modifications to purification technique.

Purity issues of proSP-C were overcome through development of a new construct termed 'proSP-C10'.

The secondary structure of proSP-C10 in SMA 2000P was analysed by CD and found to be alpha helical.

ProSP-C10 in SMALP displayed surface activity, when analysed using a Langmuir trough.

The preference of proSP-C for PI, has been demonstrated by mass spectrometry analysis, and further supported using computational modelling data.

## **7.7 SP-C Expression and Extraction**

*SFTPC* insertion into the yeast vector pPICZ $\alpha$ A was confirmed via independent sequence analysis (Eurofins).

Purification using  $\beta$ -OG detergent gave the purest fractions of SP-C.

Multiple attempts were made to identify the SP-C protein using mass spectrometry, however a positive ID could not be made, with poor cleavage during trypsin digestion having been identified as the likely problem.

## **7.8 SP-C Mass Spectrometry**

The mass spectrometry analysis identified that the phospholipids around SP-C were predominantly PI.

A HIKE motif present in the sequence of SP-C, was also identified in the modelling simulations. HIKE motifs represent a conserved consensus sequence made up of Leu, Arg, Gly and His (Ciccarelli, Acciarito and Alberti 2000).

A similar sequence can be found in in SP-C: Lys, Arg, and Leu.

HIKE motifs interact with POPI lipids (Ciccarelli, Acciarito and Alberti 2000) and so could explain why so many POPI species were identified.

## **7.9 SP-C Computational Modelling**

The *ab initio* model of SP-C generated an alpha helical protein with a long stretch of Val residues in the middle.

## **7.10 SP-C Implicit Model**

Examination of the SP-C mechanism shows an extended conformation that spans the membrane; then the structure of SP-C then forms a V shaped conformation.

## **7.11 SP-C Explicit Model**

The addition of SP-C to an Explicit membrane revealed that SP-C formed a bridge between two phospholipid leaflets resulting in an association between the two membranes.

This gives a strong indication that SP-C helps anchor the protein to the membrane and possibly lipid vesicles, whilst also helping to stabilise the membrane.

## 7.12 Summary and Key Conclusions

ProSP-B, proSP-C, proSP-C10, SP-B and SP-C were all made as recombinant proteins in *P. pastoris* for the first time.

ProSP-C10 and SP-B proteins show surface activity.

Lipid selectivity of SP-B and SP-C has also been investigated; with both proteins preferring to interact almost exclusively with PI.

Computational models of SP-B demonstrated SP-B is able to induce curvature, supporting recent literature that SP-B helps to promote the formation of vesicles.

The SP-C model was shown to help keep the membrane together virtually bridging the two leaflets.

A novel mechanism of SP-C has also been proposed whereby the structure bends inwards forming a V shape.

As the polar layer of the tear film resembles a monolayer, SP-C may adopt a V shape during compression, after which the protein extends out and helps to anchor a vesicle to the membrane. In this way SP-B and SP-C may be working together to help stabilise the membrane.

### 7.13 Future Work

The synthesis and expression of proSP-B and SP-B has great potential for future studies.

The expression of SP-B can be improved upon in future work to help produce greater quantities of SP-B.

Further development to purification tags, such as the addition of longer His tags to the SP-B construct could help improve yield and purity. This was undertaken with proSP-C10 and both the yield and purity were vastly improved.

Improving yield and purity in this way should enable treatment of respiratory disorders to benefit from lower cross infection than from those animal derived sources currently used.

SP-B and SP-C are important for lipid spreading in the tear film, the addition of the two proteins helping reduce the increasing incidence of dry eye. This may allow wider addition of SP-B and SP-C to eye formulations which themselves should then also be further examined.

Both proSP-B and SP-B have been extracted using SMA2000P, providing a platform for future structural studies. This platform will further allow greater insights into the protein-lipid interactions and functional aspects of the protein.

The mass spectrometry analysis of various SP-B and SP-C protein-lipid interactions could be further investigated with the objective of creating surfactant treatment formulations (protein-lipid) that are closer to the *in vivo* protein lipid interactions.

In order to ascertain the optimal conditions for protein activity, confirming whether PI is contributory within cell signalling in the lungs could be a rather important factor in surfactant secretion. PI is already an important lipid involved in cell signalling, hence its role in helping to understand how lipid modulation is influenced.

Synthesis and expression of proSP-C10 has provided an effective method for the production of pure proSP-C10 in *Pichia pastoris*. The reported methodology in Chapter 4 could be most usefully further optimised.

As these samples have demonstrated surface activity, they may also be used in therapies not only for dry eye but also for RDS. Formulations and treatments as outlined above could also be pursued here.

The efficacy of using full length SP-C (proSP-C10) is another area for potential investigation in light of the difficulties with aggregation of SP-C during extraction from host membranes.

Surface activity experiments have provided strong evidence of the involvement of SP-B and SP-C in lipid spreading.

Particularly understanding how both the tear film polar and non-polar layers spread in conjunction can now be investigated further by carrying out additional Langmuir experiments which include combination of both layers.

## References

- Ahmad, M., M. Hirz, H. Pichler & H. Schwab (2014) Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Applied microbiology and biotechnology*, 98, 5301-5317.
- Alfrey. T., E. Lavin 1945 The copolymerization of styrene and maleic anhydride. *Journal of American Chemical Society*. 67, 2044–2045.
- Anandarajan, M., S. Paulraj & R. Tubman (2009) ABCA3 Deficiency: an unusual cause of respiratory distress in the newborn. *Ulster Medical Journal*, 78, 51-52.
- Andersson, M., T. Curstedt, H. Jornvall & J. Johansson (1995) An amphipathic helical motif common to tumourolytic polypeptide NK-lysin and pulmonary surfactant polypeptide SP-B. *Federation of European Biochemical Societies*, 362, 328-332.
- André, N., N. Cherouati, C. Prual, T. Steffan, G. Zeder-Lutz, T. Magnin, F. Pattus, H. Michel, R. Wagner & C. Reinhart (2006) Enhancing functional production of G protein-coupled receptors in *Pichia pastoris* to levels required for structural studies via a single expression screen. *Protein Science*, 15, 1115-1126.
- Argueso, P., S. Spurr-Michaud, C. L. Russo, A. Tisdale, I. K. Gipson (2003) MUC16 Mucin Is Expressed by the Human Ocular Surface Epithelia and Carries the H185 Carbohydrate Epitope *Investigative Ophthalmology & Visual Science*, 44, 2487-2495.
- Avery, M. & J. Mead (1959) Surface properties in relation to atelectasis and hyaline membrane disease. *The American Journal of Diseases of Children*, 97, 517-523.
- Aw, R. & K. M. Polizzi (2013) Can too many copies spoil the broth. *Microbial Cell Factories*, 12, 1-9.
- Baatz, J., B. Elledge & J. Whitsett (1990) Surfactant protein SP-B induces ordering at the surface of model membrane bilayers. *Biochemistry*, 29, 6714-6720.
- Bachofen, H., U. Gerber, P. Gehr, M. Amrein & S. Schurch (2005) Structures of pulmonary surfactant films adsorbed to an air–liquid interface in vitro. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 30, 59-72.
- Baoukina, S. & P. D. Tieleman (2011) Lung Surfactant Protein SP-B Promotes Formation of Bilayer Reservoirs from Monolayer and Lipid Transfer between the Interface and Subphase. *Biophysical Journal*, 100, 1678-1687.
- Barnes, G.T and I, R Gentle 2005 *Interfacial Science*. Oxford University Press. Oxford.
- Baumgart, F., O. L. Ospina, I. Mingarro, I. Rodriguez-Crespo & J. Perez-Gil (2010) Palmitoylation of Pulmonary Surfactant Protein SP-C Is Critical for Its Functional Cooperation with SP-B to Sustain Compression/Expansion Dynamics in Cholesterol-Containing Surfactant Films. *Biophysical Journal* 99, 3234-3243.
- Bawa, Z., S. J. Routledge, M. Jamshad, M. Clare, D. Sarkar, I. Dickerson, M. Ganzlin, D. Poyner & R. Bill (2014) Functional recombinant protein is present in the pre-induction phases of *Pichia pastoris* cultures when grown in bioreactors, but not shake-flasks Functional recombinant protein is present in the pre-induction phases of *Pichia*

- pastoris cultures when grown in bioreactors, but not shake-flasks. *Microbial Cell Factories*, 13, 1-13.
- Berg , J. M., J. L. Tymoczko & L. Stryer. 2002. *Biochemistry*. New York: W H Freeman.
- Berger R.E., S. Corrsin (1974) A surface tension gradient mechanism for driving the pre-corneal tear film after a blink. *Journal of Biomechanics*, 7, 225–238.
- Bill, R. (2014) Playing catch-up with Escherichia coli: using yeast to increase success rates in recombinant protein production experiments. *Frontiers in Microbiology*, 5, 1-5.
- Björner , M., J. Lundqvist & H. Hedlund (2012) High selectivity purification screening of histidine-tagged proteins using small sample preparation formats containing cobalt IMAC media. 26.
- Booth, V., A. J. Waring, F. J. Walther & K. M. Keough (2004) NMR Structures of the C-Terminal Segment of Surfactant Protein B in Detergent Micelles and Hexafluoro-2-propanol. *Biochemistry* , 2004 , 43 (48), pp 15187–15194, 43, 15187-15194.
- Brasch, F., M. Ochs, T. Kahne, S. Guttentag, S.-V. V., M. Derrick, G. Johnen, N. Kapp, K. Muller, J. Richter, T. Giller, S. Hawgood, F. Buhling (2003) Involvement of napsin A in the C- and N-terminal processing of surfactant protein B in type-II pneumocytes of the human lung. *Journal of Biological chemistry*, 278, 49006-49014.
- Brauer, L., J. Borgermann, U. Pleyer, M. Tsokos, F. Paulsen (2007) Detection and Localization of the Hydrophobic Surfactant Proteins B and C in Human Tear Fluid and the Human Lacrimal System. *Current Eye Research*, 32, 931-938.
- Brierley, R., R. Davis & C. Holtz. 1994 Production of insulin-like growth factor-I in methylotrophic yeast cells.
- Bron, A. J., J. M. Tiffany, S. M. Gouveia, N. Yokoi & L. W. Voon (2004) Functional aspects of the tear film lipid layer. *Experimental Eye Research*, 78, 347-360.
- Brown, S., C. Kunnen, E. Duchoslav, N. Dolla, M. Kelso, E. Papas, P. Lazon de la Jara, M. Willcox, S. Blanksby & M. TW. (2013) A comparison of patient matched meibum and tear lipidomes. *Investigative ophthalmology and visual science*, 54, 7417-7424.
- Bräuer, L., M. Schicht, F. Garreis & F. Paulsen. 2012. *Innate Immunity and the Eye: Innate Immunity of the External Eye*. Jaypee-Highlights Medical Publishers.
- Brown S.I & D.G Dervichian (1969) The oils of the Meibomian glands. *Archives of Ophthalmology*, 82, 537–540.
- Butovich, I. (2013) Tear Film Lipids. *Experimental Eye Research*, 117, 1-53.
- Butovich, I., E. Uchiyama, M. Pascuale, J. McCulley (2007) Liquid Chromatography–Mass Spectrometric Analysis of Lipids Present in Human Meibomian Gland Secretions. *Lipids*, 42, 765-776.
- Butovich, I. A., J. C. Wojtowicz & M. Molai (2009) Human tear film and meibum. Very long chain wax esters and (O-acyl)-omega-hydroxy fatty acids of meibum. *The Journal of Lipid Research*, 50, 2471-2485.

- Byrne, B. (2015) *Pichia pastoris* as an expression host for membrane protein structural biology. *Current opinion in structural biology*, 32, 9-17.
- Cabre, E. M., J., D. Sutherland, J. Perez-Gil & D. Otzen (2009) Surfactant Protein SP-B Strongly Modifies Surface Collapse of Phospholipid Vesicles: Insights from a Quartz Crystal Microbalance with Dissipation *Biophysical Journal*, 97, 768-776.
- Cabré, E. J., L. M. Loura, A. Fedorov, J. Perez-Gil & M. Prieto (2012) Topology and lipid selectivity of pulmonary surfactant protein SP-B in membranes: Answers from fluorescence. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1818, 1717-1725.
- Campbell, D., G. Griffiths & B. J. Tighe (2011) Tear Analysis and Lens-tear Interactions: Part II. Ocular Lipids—Nature and Fate of Meibomian Gland Phospholipids. *Cornea*, 30, 323-332.
- Chaga, G., J. Hopp & P. Nelson (1999) Immobilized metal ion affinity chromatography on Co<sup>2+</sup>-carboxymethylaspartate-agarose Superflow, as demonstrated by one-step purification of lactate dehydrogenase from chicken breast muscle. *Biotechnology and Applied Biochemistry*, 29, 19-24.
- Chavarha, M., R. Loney, S. Rananavare & S. Hall (2015) Hydrophobic surfactant proteins strongly induce negative curvature. *Biophysical Journal*, 109, 95-105.
- Ciccarelli, F., A. Acciarito & S. Alberti (2000) Large and diverse numbers of human diseases with HIKE mutations. *Human molecular genetics*, 9, 1001-1007.
- Clark, J. C., S. E. Wert, C. J. Bachurski, M. T. Stahlman, B. R. Stripp, T. E. Weaver & J. A. Whitsett (1995) Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. *Cell Biology*, 92, 7794-7798,.
- Clemens, R., M. Poornima, W. Malgorzata, E. W. Timothy, M. Viktor, I. Steven, P. Klaus T, S. Werner, G. Andreas & M. Philipp (2008) Recombinant production of a hybrid plasminogen activator composed of surfactant protein B and low-molecular-weight urokinase. *Thrombosis and Haemostasis*, 100, 1185-1192.
- Clements, J. (1956) Dependence of pressure-volume characteristics of lungs on intrinsic surface active material. *American Journal of Physiology*, 187, 592.
- Cooper, G. M. 2000. *The Cell: A Molecular Approach Sunderland (MA) 2000. Structure of the Plasma Membrane*. Sunderland MA: Sinauer Associates.
- Corfield, A. P. (2015) Mucins: A biologically relevant glycan barrier in mucosal protection. *Biochimica et Biophysica Acta 1850 (2015) 236–252*, 1850, 236-252.
- Craig, J. P. & A. Tomlinson (1997) Importance of the lipid layer in human tear film stability and evaporation. *Optometry and Vision Science*, 74, 8-13.
- Curstedt, T., J. Johansson, P. Persson, A. Eklund, B. Robertson, B. Löwenadler & H. Jörnvall (1990) Hydrophobic surfactant-associated polypeptides: SP-C is a lipopeptide with two palmitoylated cysteine residues, whereas SP-B lacks covalently linked fatty acyl groups. *Proceedings of the National Academy of Sciences in the United States of America*, 87, 2985-2989.
- Cwiklik, L. (2016) Tear film lipid layer: A molecular level view. *Biochimica et Biophysica acta*, 1858, 2421-2430.

- Dawaliby, R., C. Trubbia, C. Delporte, M. Masureel, P. Van Antwerpen, B. K. Kobilka & C. Govaerts (2016) Allosteric regulation of G protein-coupled receptor activity by phospholipids. *Nat Chem Biol*, 12, 35-39.
- De Souza, G. A., L. M. Godoy, M. Mann (2006) Identification of 491 proteins in the tear fluid proteome reveals a large number of proteases and protease inhibitors. *Genome Biology*, 7.
- Dörr, J. M., S. Scheidelaar, M. C. Koorengel, J. J. Dominguez, M. Schäfer, C. A. van Walree, & J. A. Killian, (2016). The styrene–maleic acid copolymer: a versatile tool in membrane research. *European Biophysics Journal*, 45, 3–21.
- Dragosits, M., G. Frascotti, L. Bernard-Granger, F. Vázquez, M. Giuliani, K. Baumann, E. Rodríguez-Carmona, J. Tokkanen, E. Parrilli, M. Wiebe, R. Kunert, M. Maurer, B. Gasser, M. Sauer, P. Branduardi, T. Pakula, M. Saloheimo, M. Penttilä, P. Ferrer, M. Luisa Tutino, A. Villaverde, D. Porro & D. Mattanovich (2011) Influence of growth temperature on the production of antibody Fab fragments in different microbes: a host comparative analysis. *Biotechnology Progress*, 27, 38-46.
- Flanagan, J. & M. Wilcox (2009) Role of Lactoferrin in the tear film. *Biochimie*, 91, 35-43.
- Forrest, L.R., C.L Tang., B. Honig (2006) On the accuracy of homology modelling and sequence alignment methods applied to membrane proteins. *Biophysical Journal*, 91, 508-517.
- Foster, C., A. Aktar, D. Kopf, P. Zhang, S. Guttentag (2004) Pepsinogen C: a type 2 cell-specific protease. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 286, L382-L387.
- García-Montoya, I. A., T. Siqueiros Cendón, S. Arévalo-Gallegos & Q. Rascón-Cruz (2012) Lactoferrin a multiple bioactive protein: An overview. *Biochimica et Biophysica Acta*, 1820, 226-236.
- Gendler, S. J. & A. P. Spicer (1995) Epithelial Mucin Genes. *Annual review of Physiology*, 57, 607-634.
- Gil, F., E. Cabre, C. Pastor, P. Estrada & Perez-Gil. 2009a. *Expression of different recombinant forms of the precursor of human pulmonary surfactant protein B (pro SP-B) in Pichia pastoris*. Spain: World Scientific Publishing Co.
- Gipson, I. K. (2004) Distribution of mucins at the ocular surface. *Experimental Eye Research*, 78, 379–388.
- Glasser, S. W., T. R. Korfhagens, M. D. Bruno, C. Dey & J. A. Whitsett (1990) Structure and Expression of the Pulmonary Surfactant Protein SP-C Gene in the Mouse. *The Journal of Biological Chemistry*, 265, 21986-21991.
- Glasser, S. W., T. R. Korfhagens, T. E. Weaver, L. J. Fox, T. Pilot-Matias & J. A. Whitsett (1987) cDNA and deduced amino acid sequence of human pulmonary surfactant-associated proteolipid SPL(Phe). *Proceedings of the National Academy of Science*, 84, 4007-4011.
- Goerke, J. (1998) Pulmonary surfactant: functions and molecular composition. *Biochimica et Biophysica Acta (BBA)-Molecular basis of Disease*, 1408, 79-89.

- Govindarajan, B. & I. K. Gipson (2010) Membrane-tethered mucins have multiple functions on the ocular surface. *Experimental Eye Research*, 90, 655-663.
- Green-Church, K. B., I. Butovich, M. Willcox, D. Borchman, F. Paulsen, S. Barabino & B. Glasgow (2011) The International Workshop on Meibomian Gland Dysfunction: Report of the Subcommittee on Tear Film Lipids and Lipid-Protein Interactions in Health and Disease. *IOVS*, 52, 15.
- Grigorian, A. L., J. J. Bustamante, P. Hernandez, A. O. Martinez & L. S. Haro (2004) Extraordinarily stable disulfide-linked homodimer of human growth hormone. *Protein Science*, 14, 902-913.
- Grillitsch, K., P. Tarazona, L. Klug, T. Wriessnegger, G. Zellnig, E. Leitner, I. Feussner & G. Daum (2014) Isolation and characterization of the plasma membrane from the yeast *Pichia pastoris*. *Biochimica et Biophysica acta*, 1838(7):1889-97, 1889-1897.
- Grisshammer, R. & J. Tucker (1997) Quantitative evaluation of neurotensin receptor purification by immobilized metal affinity chromatography. *Protein expression and purification*, 11, 53-60.
- Gulati, S., M. Jamshad, T. Knowles, K. Morrison, R. Downing, N. Cant, R. Collins, J. Koenderink, R. Ford, M. Overduin, I. Kerr, T. Dafforn & A. Rothnie (2014) Detergent-free purification of ABC (ATP-binding-cassette) transporters. *The Biochemical Journal*, 461, 269-278.
- Guttentag, S., L. Robinson, P. Zhang, F. Brasch, F. Buhling & M. Beers (2003) Cysteine Protease Activity Is Required for Surfactant Protein B Processing and Lamellar Body Genesis. *American Journal of Respiratory Cell and Molecular Biology*, 28, 69-79.
- Halliday, H. (2008) Surfactants: past, present and future. *Journal of Perinatology*, 28, S47-S56.
- Hattrup, C. & S. Gendler (2008) Structure and function of the cell surface (tethered) mucins. *Annual review of Physiology* 70, 431-457.
- Hickey, K. D. & M. M. Buhr (2010) Lipid Bilayer Composition Affects Transmembrane Protein Orientation and Function. *Journal of Lipids*, 2011, 9.
- Higgins, D. R. & J. M. Cregg. 1998. Introduction to *Pichia pastoris*. In *Pichia Protocols*, eds. D. R. Higgins & J. M. Cregg, 1-15. Totowa, NJ: Humana Press.
- Hodge, K., S. Have, L. Hutton, . & A. Lamond (2013) Cleaning up the masses: Exclusion lists to reduce contamination with HPLC-MS/MS. *Journal of Proteomics*, 88, 92-103.
- Holly, F.J (1973) Formation and rupture of the tear film. *Experimental Eye Research*, 15, 515-525.
- Holly , F. & M. Lemp (1977) Tear physiology and Dry eyes. *Survey of ophthalmology*, 22, 69-87.
- Jamshad, M., J. Charlton, Y. Lin, S. Routledge, Z. Bawa, T. Knowles, M. Overduin, N. Dekker, T. Dafforn, R. Bill, D. Poyner & M. Wheatley (2015) G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Bioscience reports*, 35, 1-10.

- Johansson, J. (1998) Structure and properties of surfactant protein C. *Biochimica et Biophysica Acta*, 1408, 161-172.
- Johansson, J., T. Curstedt & H. Jornvall (1991) Surfactant Protein B: Disulfide Bridges, Structural Properties, and Kringle Similarities. *Biochemistry*, 30, 6917-6921.
- Johansson, J., G. Nilsson., R. Strömberg., B. Robertson., H. Jörnvall and T Curstedt (1995) Secondary structure and biophysical activity of synthetic analogues of the pulmonary surfactant polypeptide SP-C. 307, 535-541.
- Johansson, J., T. Szyperski, T. Curstedt & K. Wiithrich (1994) The NMR Structure of the Pulmonary Surf actant-Associated Polypeptide SP-C in an Apolar Solvent Contains a Valyl-Rich  $\alpha$ -Helix. *Biochemistry*, 6015-6023.
- Khatami, M. H., I. Saika-Voivod & V. Booth (2016) All-atom molecular dynamics simulations of lung surfactant protein B: Structural features of SP-B promote lipid reorganization. *Biochimica et Biophysica Acta*, 1858, 3082-3092.
- Keil, B. 1992 Specificity of Proteolysis. Springer-Verlag Berlin-Heidelberg – New York, pp.335
- King, R. J., D. J. Klass, E. G. Gikas & J. A. Clements (1973) Isolation of apoproteins from canine surface active material. *American Journal of Physiology*, 224, 788-795.
- King-Smith, P. E., E. A. Hinel & J. J. Nichols (2010) Application of a Novel Interferometric Method to Investigate the Relation between Lipid Layer Thickness and Tear Film Thinning. *Investigative Ophthalmology & Visual Science*, 51.
- Kishimoto, Y., M. Hiraiwa & J. S. O'Brien (1992) Saposins: structure, function, distribution, and molecular genetics. *The Journal of Lipid Research*, 33, 1255-1267.
- Knowles, T., F. R., C. Smith, Y.-P. Lin, T. Dafforn, M. Overduin (2009) Membrane Proteins Solubilized Intact in Lipid Containing Nanoparticles Bounded by Styrene Maleic Acid Copolymer. *Journal of American Chemical Society*, 131, 7484-7485.
- Kronqvist, N., M. Sarr, A. Lindqvist, K. Nordling, M. Otkovs, L. Venturi, B. Pioselli, P. Purhonen, M. Landreh, H. Biverstal, Z. Toleikis, L. Sjoberg, C. V. Robinson, N. Pelizzi, H. Jornvall, H. Herbert, K. Jaudzems, T. Curstedt, A. Rising & J. Johansson (2017) Efficient protein production inspired by how spiders make silk. *Nature comminucations*, 8, 1-15.
- Kulovesi, P., J. Telenius, A. Koivuniemi, G. Brezesinski, A. Rantama ki, T. Viitala, E. Puukilainen, M. Ritala, S. K. Wiedmer, I. Vattulainen & J. M. Holopainen (2010) Molecular Organization of the Tear Fluid Lipid Layer. *Biophysical Journal*, 99, 2559-2567.
- Kurutz, J. & K. Lee (2002) NMR structure of lung surfactant peptide SP-B(11-25). *Biochemistry*, 41, 9627-9636.
- Le Maire, M., P. Champeil & J. Moller (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et biophysica acta*, 1508, 86-111.
- Lee, S. C., S. Khalide, N. L. Pollock, K. T. J., K. Edlerd, A. J. Rothnie, O. R. Thomas & T. Dafforn (2016) Encapsulated Membrane Proteins: A Simplified System For Molecular Simulation. 1858, 2549-2557.

- Lee, S. C., T. J. Knowles, V. L. Postis, M. Jamshad, R. A. Parslow, Y. P. Lin, A. Goldman, P. Sridhar, M. Overduin, S. P. Muench & T. R. Dafforn (2016b) A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols*, 11, 1150-1162.
- Li, N., N. Wang, J. Zheng, M. Liu, O. Lever, P. Erickson, L. Li (2005) Characterization of Human Tear Proteome Using Multiple Proteomic Analysis Techniques. *Journal of Proteome Research*, 4, 2052-2061.
- Lin, S., K. S. Phillips, M. R. Wilder & T. E. Weaver (1996) Structural requirements for intracellular transport of pulmonary surfactant protein B (SP-B). *Biochimica et Biophysica Acta*, 1312, 177-185.
- Lopez-Rodriguez, E. & J. Pérez-Gil (2014) Structure-function relationships in pulmonary surfactant membranes: From biophysics to therapy. *Biochimica et Biophysica Acta*, 1838, 1568-1585.
- Lukovic, D., A. Cruz, A. Gonzalez-Horta, A. Almlen, T. Curstedt, I. Mingarro & J. Pérez-Gil (2012) Interfacial Behavior of Recombinant Forms of Human Pulmonary Surfactant Protein SP-C. *Langmuir*, 28, 7811-7825.
- Lukovic, D., I. Plasencia, F. J. Taberner, J. Salgado, J. J. Calvete, J. Pérez-Gil & I. Mingarro (2006) Production and characterisation of recombinant forms of human pulmonary surfactant protein C (SP-C): Structure and surface activity. *Biochimica et Biophysica Acta* 1758, 1758, 509-518.
- Macauley-Patrick, S., M. Fazenda, B. McNeil & L. Harvey (2005) Heterologous protein production using the *Pichia pastoris* expression system. *Yeast*, 22, 249-270.
- Man Lam, S., L. Tong, D. Xinrui, A. Petznick, M. R. Wenk & G. Shui (2014) Extensive characterization of human tear fluid collected using different techniques unravels the presence of novel lipid amphiphiles. *Journal of Lipid Research*, 55, 289-298.
- Matyash, V., G. Liebisch, T. V. Kurzchalia, A. Shevchenko & D. Schwudke (2008) Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics. *Journal of Lipid Research*, 49, 1137-1146.
- McCulley, J. P. & W. Shine (1997) A compositional based model for the tear film lipid layer *Transactions of the American Ophthalmological Society*, 95, 79-93.
- Millar, T. J. (2013) A Mechanism to Explain the Behaviour of Spread Films of Meibomian Lipids. *Current Eye Research*, 38, 220-223.
- Millodot, M. 2009. *Dictionary of Optometry and Visual Science*. Butterworth Heinemann Elsevier.
- Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, P. D. Tieleman & S.-J. Marrink (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation*, 4, 818-834.
- Mulugeta, S., S.-I. Nureki & M. F. Beers (2015) Lost after translation: insights from pulmonary surfactant for understanding the role of alveolar epithelial dysfunction and cellular quality control in fibrotic lung disease. *American Journal of Physiology. Lung Cellular and Molecular Physiology*, 309, L507-L525.

- Munoz-Hernandez, A., C. Galbis-Estrada, E. Santos-Bueso, R. Cuina-Sardina, D. Diaz-Valle & J. Gegundez-Fernandez (2016) Human tear metabolome. *Arch Soc Esp Oftalmol* 2016;91:157e9. *Archivos De La Sociedad Espanola De Oftalmologia*, 91, 157-159.
- Nash, J. A., T. Nicole S Ballard, T. E. Weaver & H. T. Akinbi (2006) The Peptidoglycan-Degrading Property of Lysozyme Is Not Required for Bactericidal Activity In Vivo. *The Journal of Immunology*, 177, 519-526.
- Neergaard, K. (1929) New conceptions on a basic concept of respiratory mechanics. *Journal for all experimental medicine*, 66, 373-394.
- Nicolaidis, N., J. K. Kaitaranta, T. N. Rawdah, J. I. Macy, F. M. Boswell & R. E. Smith (1981) Meibomian gland studies: comparison of steer and human lipids. *Investigative Ophthalmology. and Vision Science*, 20, 522-536.
- Nielsen, P. A., E. P. Bennett, H. H. Wandall, M. H. Therkildsen<sup>1</sup>, H. Jens & H. Clausen (1997) Identification of a major human high molecular weight salivary mucin (MG1) tracheobronchial mucin MUC5B. *Glycobiology vol. 7 no. 3 pp. 413-419, 1997, 7, 413-419.*
- Nogee, L. M., D. E. deMello, L. Dehner & H. R. Colten (1993) Deficiency of Pulmonary Surfactant Protein B in Congenital Alveolar Proteinosis. *The New England Journal of Medicine*, 328, 406-410.
- Nogee, L. M., A. E. Dunbar, S. E. Wert, F. Askin, A. Hamvas & J. A. Whitsett (2001) A Mutation in the Surfactant Protein C Gene Associated with Familial Interstitial Lung Disease. *The New England Journal of Medicine*, 344, 573-579.
- Olmeda, B., b. Garcia alvarez, M. J. Gómez, M. Martínez-Calle, A. Cruz & J. Perez-Gil (2015) A model for the structure and mechanism of action of pulmonary surfactant protein B. *FASEB journal*, 29, 1-13.
- Olmeda, B., B. García-Álvarez & J. Pérez-Gil (2013) Structure-function correlations of pulmonary surfactant protein SP-B and the saposin-like family of proteins. *European Biophysics Journal*, 42, 209-222.
- Oluwole, A. O., B. Danielczak, A. Meister, O. J. Babalola, C. Vargas & S. Keller (2017) Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angewandte Chemie International Edition*, 56, 1919-1924.
- Palacios, A., B. Gonzalez, S. Alonso, J. Perez-Gil & P. Estrada (2006) Production of a recombinant form of the propeptide NH<sub>2</sub>-terminal of the precursor of pulmonary surfactant protein B. *Enzyme and Microbial technology*, 40, 85-92.
- Parra, E., A. Alcaraz, A. Cruz, V. M. Aguilera & J. Perez-Gil (2013) Hydrophobic Pulmonary Surfactant Proteins SP-B and SP-C Induce Pore Formation in Planar Lipid Membranes: Evidence for Proteolipid Pores. *Biophysical Journal*, 104, 146-155.
- Patterson, M., H. J. Vogel & E. J. Prenner (2016) Biophysical characterization of monofilm model systems composed of selected tear film phospholipids. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1858, 403-414.
- Pattle, R. (1955) Properties, function and origin of the alveolar lining layer. *Nature*, 175, 1125-1126.

- Perez-Gil, J. (2008) Structure of pulmonary surfactant membranes and films: The role of proteins and lipid–protein interactions. *Biochimica et Biophysica Acta*, 1778, 1676-1695.
- Perez-Gil, J., C. Casals & D. Marsh (1995) Interactions of hydrophobic lung surfactant proteins SP-B and SP-C with dipalmitoylphosphatidylcholine and dipalmitoylphosphatidylglycerol bilayers studied by electron spin resonance spectroscopy. *Biochemistry*, 34, 3964-3971.
- Perez-Gil, J., A. Cruz, C. Casals (1993) Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in aqueous organic solvents and lipids. *Biochimica et Biophysica Acta*, 1168, 261-270.
- Perez-Gil, J. & T. E. Weaver (2010) Pulmonary Surfactant Pathophysiology: Current Models and Open Questions. *Physiology*, 25, 132-141.
- Persson, A., D. Chang, K. Rust, M. Moxley, W. Longmore & E. Crouch (1989) Purification and biochemical characterization of CP4 (SP-D), a collagenous surfactant-associated protein. *Biochemistry*, 28, 6361–6367.
- Philippot, J. R. & F. Schuber. 1994. *Liposomes as tools in Basic research and Industry*. USA: CRC Press.
- Phillip, A. (1995) Neonatal mortality rate: is further improvement possible. *The Journal of Paediatrics*, 126, 427-433.
- Posa, A., F. Paulsen., R. Dietz., F. Garreis., R. Sanderd., M. Schicht., S. Sele., M Scholz., C. M. Hammer & Lars Bräuer (2017) Quantification of surfactant proteins in tears of patients suffering from dry eye disease compared to healthy subjects. *Annals of Anatomy*, 216, 90–94.
- Possmayer, F., S. Yu, J. Weber & P. Harding (1984) Pulmonary Surfactant. *Canadian Journal of Cell Biology and Biochemistry*, 62, 1121-1133.
- Pérez-Gil, J., Cruz A & C. C. (1993) Solubility of hydrophobic surfactant proteins in organic solvent/water mixtures. Structural studies on SP-B and SP-C in aqueous organic solvents and lipids. *Biochimica et Biophysica Acta*, 1168, 261-70.
- Rath, A., M. Glibowicka, V. G Nadeau, G. Chen & C. M Deber (2009) Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 1760-1765.
- Rantamaki, A., J. Telenius, A. Koivuniemi, I. Vattulainen, H. JM (2011) Lessons from the biophysics of interfaces: Lung surfactant and tear fluid. *Progress in Retinal and Eye Research*, 30, 204-215.
- Rost B. 1999 Twilight zone of protein sequence alignments, *Protein Engineering*, 12, 85 – 94.
- Rothnie, A., J. Storm, J. Campbell, K. J. Linton, I. D. Kerr & R. Callaghan (2004) The Topography of Transmembrane Segment Six Is Altered during the Catalytic Cycle of P-glycoprotein. *The Journal of Biological Chemistry*, 279, 34913–34921.
- Routledge, S. & M. Clare. 2012. Setting up a bioreactor for recombinant protein production. In *Methods in Molecular Biology*, ed. R. M. Bill, 99-113. Humana Press.

- Ryan, M. A., Q. Xiaoyang, A. G. Serrano, M. Ikegami, J. Perez-Gil, J. Johansson & T. E. Weaver (2005) Mapping and Analysis of the Lytic and Fusogenic Domains of Surfactant Protein B. *Biochemistry*, 44, 861-872.
- Sambrook, J., E. Fritsch & T. Maniatis. 1989. *Molecular Cloning: A laboratory manual, 2nd Ed.* Cold Spring Harbor Laboratory Press Greene Publishing Associates and John Wiley & Sons. New York. pp 1.63-1.70
- Sarker, M., A. J. Waring, F. J. Walther, K. M. Keough & V. Booth (2007) Structure of Mini-B, a Functional Fragment of Surfactant Protein B, in Detergent Micelles. *Biochemistry*, 46, 11047-11056.
- Saville, J., Z. Zhao, M. Willcox, M. Ariyavidana, S. Blanksby & T. Mitchell (2011) Identification of phospholipids in human meibum by nano-electrospray ionisation tandem mass spectrometry. *Experimental eye research*, 92, 238-240.
- Schaumberg, D. A., D. A Sullivan, J. E Buring & M.R Dana (2003) Prevalance of dry eye syndrome among US women. *American Journal of Opthamology*, 136, 318-326.
- Schaumberg, D. A., M. R Dana., J. E Buring & D.A Sullivan (2009) Prevalance of dry eye syndrome among US men: estimates from the physicians' health studies. *Archives of Opthamology*, 127, 763-768.
- Scheidelaar, S., M. C. Koorengel, J. P. Dominguez, J. D. Meeldijk, E. Breukink & A. J. Killian (2015) Molecular Model for the Solubilization of Membranes into Nanodisks by Styrene Maleic Acid Copolymers. *Biophysical Journal*, 108, 279-290.
- Schurch, D., O. Ospina, A. Cruz & Perez-Gil. (2010) Combined and Independent Action of Proteins SP-B and SP-C in the Surface Behavior and Mechanical Stability of Pulmonary Surfactant Films *Biophysical Journal*, 99, 3290-3299.
- Schägger, H. 2006. Tricine-SDS-PAGE. 16-23. Nature protocols.
- Seddon, A. M., P. Curnow & P. J. Booth (2004) Membrane proteins, lipids and detergents: not just a soap opera. *Biochimica et Biophysica Acta (BBA) - Biomembranes*, 1666, 105-117.
- Serrano, A., C. EJ., Perez-Gil., Cabre (2007) Identification of a segment in the precursor of pulmonary surfactant protein SP-B, potentially involved in pH-dependent membrane assembly of the protein. *Biochimica et Biophysica Acta*, 1768, 1059-1069.
- Sharifahmadian, M., M. Sarker, D. Palleboina, A. J. Waring, F. J. Walther, M. Morrow & V. Booth (2013) Role of the N-Terminal Seven Residues of Surfactant Protein B (SP-B). *Plos one*, 8.
- Sibony, P. A., B. Walcott, C. McKeon & F. Jakobiec. 1988. Vasoactive Intestinal Polypeptide and the Innervation of the Human Lacrimal Gland. 1085-1088. *Arch Ophthalmol.* 1988;106(8):1085–1088.
- Simms, J., D. L. Hay, M. Wheatley & D. Poyner (2006) Characterization of the structure of RAMP1 by mutagenesis and molecular modeling. *Biophysical Journal*, 91, 662-669.
- Sunga, A. J., I. Tolstorukov & J. M Cregg (2008) Posttranslational vector amplification in the yeast *Pichia Pastoris*. *FEMS Yeast Research*, 8, 870-876.

- Sugita, Y. & Y. Okamoto (1999) Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314, 141-151.
- Taneva, S. G., J. Stewart, L. Taylor & K. M. W. Keough. 1998. Method of purification affects some interfacial properties of pulmonary surfactant proteins B and C and their mixtures with dipalmitoylphosphatidylcholine. 138-150. *Biochimica et Biophysica Acta*.
- The management of dry eye 2016 *BMJ* 353, 2333.
- Teh, S. H., M. Y Fong and Z. Mohamed (2011) Expression and analysis of the glycosylation properties of recombinant human erythropoietin expressed in *Pichia pastoris*. *Genetics and Molecular Biology*, 34, 464-470.
- Tiffany, J. (1978) Individual variations in human meibomian composition. *Experimental eye research*, 27, 289-300.
- Tonge, S. & B. Tighe (2001) Responsive hydrophobically associating polymers: a review of structure and properties. *Advanced Drug Delivery Reviews* *Advanced Drug Delivery Reviews*, 53, 109-122.
- Ueno, T., S. Linder, C.-L. Na, W. R. Rice, J. Johansson & T. E. Weaver (2004) Processing of Pulmonary Surfactant Protein B by Napsin and Cathepsin H. *The Journal of Biological Chemistry*, 279, 16178–16184.
- Ulmschneider, J. P. & M. B. Ulmschneider (2009) Sampling efficiency in explicit and implicit membrane environments studied by peptide folding simulations. *Proteins*, 75, 586-597.
- Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. Mark & H. Berendsen (2005) GROMACS: fast, flexible, and free. *Journal of computational chemistry*, 26, 1701-1718.
- Vandenbussche, G., A. Clercx, T. Curstedt, J. Johansson, H. Jornvall & J. Ruysschaertt (1992) Secondary Structure and Orientation of the Surfactant Protein SP-B in a Lipid Environment. A Fourier Transform Infrared Spectroscopy Study. *Biochemistry*, 31, 9169-9176.
- Veldhuizen, R., K. Nagb, S. Orgeigc, D. Possmayer (1998) The role of lipids in pulmonary surfactant. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1408, 90-108.
- Voorhout, W., T. Veendaal, H. Haagsman, T. Weaver, J. Whitsett, G. Van & H. Geuze (1992) Intracellular processing of pulmonary surfactant protein B in an endosomal/lysosomal compartment. *American Journal of Physiology*, 263,479-486.
- Vorbroker, D., S. Profitt, L. Noguee & J. Whitsett (1995) Aberrant processing of surfactant protein C in hereditary SP-B deficiency. *American Journal of Physiology - Lung Cellular and Molecular Physiology*, 268, L647-L656.
- Vyas, V., R. Ukawala, M. Ghate & C. Chintha (2012) Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. *Indian Journal of Pharmaceutical Sciences*, 74, 1-17.
- Waring, A., F. Walther, L. Gordon, J. Hernandez-Juviel, T. Hong, M. Sherman, C. Alonso, T. Alig, A. Braun, D. Bacon & J. Zasadzinski (2005) The role of charged amphipathic helices in the structure and function of surfactant protein B. *Journal of peptide research*, 66, 364-374.

- Weaver, T., V. Sarin, N. Sawtell, W. Hull & J. Whitsett (1988) Identification of surfactant proteolipid SP-B in human surfactant and fetal lung. *Journal of Applied Physiology*, 65, 982-987.
- Weaver, T. E. 1996. Mature alveolar SP-B and a process for producing the same (1998) Synthesis, processing and secretion of surfactant proteins B and C. 1408, 173-179.
- Whitsett, J. A., S. E. Wert & T. E. Weaver (2015) Diseases of Pulmonary Surfactant Homeostasis. *Annual review of Pathology*, 10, 371-393.
- Willcox, M. D., P. Argüeso, G. A. Georgiev, J. M. Holopainen, G. W. Laurie, T. J. Millar, E. B. Papas, J. P. Rolland, T. A. Schmidt, U. Stahl, T. Suarez, L. N. Subbaraman, O. O. Uçakhan & L. Jones (2017) TFOS DEWS II Tear Film Report. *The Ocular Surface*, 15, 366-403.
- Wizert, A., R. D. Iskander & L. Cwiklik (2014) Organization of Lipids in the Tear Film: A Molecular- Level View. *Plos one*, 9, 1-10.
- Wolff. 1954. *Wolff's The anatomy of the eye and orbit*. Spain: Chapman and Hall.
- Yeagle, P. L. 2016. *The membranes of cells*. CRC press Taylor and Francis Group
- Zaltash, S. & J. Johansson (1998) Secondary structure and limited proteolysis give experimental evidence that the precursor of pulmonary surfactant protein B contains three saposin-like domains. *FEBS letters*, 423, 1-4.
- Zaltash, S., M. Palmblad, T. Curstedt, J. Johansson & B. Persson (2000) Pulmonary surfactant protein B: a structural model and a functional analogue. *Biochimica et Biophysica Acta* 1466, 179-186.