Structural studies of surfactant protein D in complex with bacterial lipopolysaccharide ligands

Ruben Filipe da Silva

Thesis for submission towards the degree of Doctor in Philosophy

October 2017

Keele University

Abstract

This work is focused on the recognition of natural lipopolysaccharide (LPS) by the innate immune protein human lung surfactant protein D (hSP-D) in the form of a biologically active recombinant fragment (rfhSP-D), containing the α-helical coiled-coil and three carbohydrate recognition domains (CRD). Intact LPS from two bacterial strains, *S.* minnesota (R5 mutant) and *H. influenzae* type b Eagan (CA7 mutant), were delipidated by means of mild acid hydrolysis, leaving the purified polysaccharide (PS) to be used in X-ray diffraction studies by means of co-crystallisation with rfhSP-D. *S.* minnesota R7 full LPS was also investigated following development of a suitable solubilisation method which also utilised the LPS from *E. coli* O111:B4.

The structural studies of rfhSP-D bound to *H. influenzae* Eagan CA7 PS (solved and refined at 2.98 Å) and to *S.* minnesota rough mutant LPS/PS (solved and refined at 3.3 Å) reveal that rfhSP-D binds to LPS preferentially through the non-terminal inner core heptose HepI via the O6' and O7' hydroxyls. rfhSP-D recognition of *S.* minnesota HepI shows a similar bound heptose orientation to that previously reported for heptose binding by rfhSP-D in the literature with an indication of normal Kdo in the inner core Kdo-Hep-Hep trisaccharide.

rfhSP-D recognition of the HepI of *H. influenzae* Eagan CA7 reveals a novel bound heptose orientation, with the heptose rotated by 180° about C5-C6, resulting in the O6' and O7' hydroxyls being interchanged with respect to coordination to Ca1 and protein. The novel orientation of HepI is accompanied by a salt bridge being formed between the flanking residue Arg343 and Glu347, both of which adopt a previously unseen conformation. The novel binding mechanism of rfhSP-D for Eagan CA7 suggests flexibility in recognition and offers evidence to explain why this mutant binds more weakly than the Eagan 4A mutant to both rfhSP-D and hSP-D.

Contents

Chapter 1 - Introduction **1**

1.1 The human lung and the immune system 1

1.1.1 Innate Immunity in the lung 1

1.1.2 Primary antimicrobial defences in the alveolar lining layer 2

1.1.3 PRRs in the lung 3

1.1.4 Pulmonary surfactant- important in the lung normal physiology 4

1.1.4.1 Lamellar bodies and tubular myelin 5

1.1.4.2 Surfactant Lipids 5

1.1.4.3 Surfactant Proteins 6

1.2 Lipopolysaccharides (LPS) 7

1.2.1 Bacteria- gram-negative Vs gram-positive 7

1.2.2 Lipopolysaccharide (LPS) - Biological Role 8

1.2.3 LPS common structure 9

1.2.4 *Escherichia coli, Salmonella enterica* (*serotype minnesota*) and *Haemophilus influenzae* 11

1.2.4.1 *Enterobacteriaceae – Escherichia coli* and *Salmonella enterica* (serotype minnesota) 11

1.2.4.2 *E. coli* and *S. enterica* (minnesota) LPS 12

1.2.4.2.1 O-antigen/O-polysaccharide 13

1.2.4.2.2 Biosynthesis of *E. coli* and *S. enterica* LPS structure 14

1.2.4.2.3 Rough, rough mutant and smooth LPS 14

1.2.4.3 *Pasteurellaceae – Haemophilus influenzae* 16

1.2.4.3.1 LPS from *Haemophilus influenzae* 17

1.2.5 Phosphorylation 18

1.2.6 Sialic Acids 19

1.3 C-type lectins 20

1.3.1 C-type lectin domain (CTLD) and the C-type lectin like domain (CTLD*) 20

1.3.2 CTLD / CRD FOLD 21

1.4 Collectins 22

1.4.1 Collectins: Function 24

1.4.1.1 Mannose-Binding Protein (MBP) 24

1.4.1.2 Surfactant Protein A (SP-A) 25

1.4.1.3 Bovine/Cattle Collectins- CL-43, CL-46 and Conglutinin 26

1.4.1.4 Novel Collectins - CL-K1, CL-L1 and CL-P1 27

1.4.1.5 Surfactant protein D 28
1.4.1.5.1 Agglutination 29
1.4.1.5.2 Opsonisation 30
1.4.1.5.3 Phagocytosis 30
1.4.1.5.4 Direct Microbial Activity 31
1.4.1.5.5 Bacteria and mycobacteria 31
1.4.1.5.6 Viruses 33
1.4.1.5.7 Fungi and Parasites 34
1.4.1.5.8 Allergens 35
1.4.1.5.9 Necrotic/Apoptotic Cells 36
1.4.1.5.10 SP-D and the immune system 36
1.4.1.5.11 SP-D - Transgenic Studies and Pathogenesis 38
1.4.1.5.12 Disease - Cause or Consequence of SP-D levels? 40
1.4.2 Collectin structure and sequence 40
1.4.2.1 Collectin trimerisation and multimerisation 41
1.4.2.2 Sequence and structure 42
1.4.2.3 N-terminal region 46
1.4.2.4 Collagen Region 47
1.4.2.5 α-helical neck region 48
1.4.2.6 Carbohydrate Recognition Domain (CRD) 50
1.4.3 Distinction between Self and Non-self 53
1.4.4 Glycosylation of collectins 53
1.4.5 Sequence, structure and function 54
1.4.5.1 Collectin binding affinities 54
1.4.5.2 Ligand binding pocket 56
1.4.6 Structural studies 58
1.5 Structural studies of Surfactant Protein D (SP-D) 60
1.5.1 Recombinant fragment of human SP-D (rfhSP-D) 60
1.5.2 Structural Studies of rfhSP-D 62
1.5.2.1 Timeline 62
1.5.2.2 The ligand-free structure of rfhSP-D 63
1.5.2.3 Ca4 and the asymmetric Tyr228 64
1.5.2.4 Ligand-bound rfhSP-D structures 67
1.5.3 rfhSP-D mutants and the importance of residues 343 and 325 74
1.6 Conclusion 76
Chapter 2 - Experimental Techniques 80

2.1 Crystals 80
2.1.1 Unit Cell 80
2.1.2 Crystal Symmetry and Bravais Lattices 81
2.1.3 X-rays 82
2.1.4 Diffraction by Crystals 82
2.1.5 Data Collections and diffraction images 84
2.2 Fourier transform and Structure Factors 86
2.3 Data processing and Model Building 87
2.3.1 Data Integration 87
2.3.2 *TRUNCATE, UNIQUIFY* and *RE-INDEX* 88
2.3.3 Phase determination – Molecular replacement 89
2.3.4 Electron density maps 89
2.3.5 *REFMAC5* 91
2.3.6 Structure refinement and model building 92
2.3.7 Crystallographic parameters and indicators of data quality 92
2.3.7.1 Data collection parameters and indicators of data quality 93
2.3.7.2 Indicators of model quality 94
2.3.8 Fast data processing (fast_dp) 94
2.4 rfhSP-D protein preparation 95
2.4.1 rfhSP-D protein 95
2.4.1.1 Protein preparation – Expression 95
2.4.1.2 Protein preparation – Purification 96
2.4.1.2.1 Dialysis 96
2.4.1.2.2 Protein Concentration 97
2.5 Protein Crystallisation 97
2.5.1 Nucleation and Growth – Crystal formation 98
2.5.2 Crystallisation by Vapour Diffusion 100
2.5.2.1 The Sitting-drop Vapour Diffusion Method 101
2.5.2.2 Structure/industrial Screens 101
2.5.2.3 Co-crystallisation vs ligand-soaked crystals 102
2.5.3 Crystal Preparation and Data Collection 103
2.5.3.1 Cryo-protection, Flash Freezing and Radiation Damage 104
2.5.3.1.1 Cryo-protection 104
2.5.3.1.2 Flash freezing 105
2.5.3.1.3 Radiation damage 105
2.6 Lipopolysaccharides (LPS) 106

2.6.1 LPS in Crystallography 106
2.6.2 Hydrolysis of LPS 107
2.6.2.1 Hydrolysis Method 107
2.6.2.2 Acetic Mild Acid Hydrolysis - Impact on LPS structure 108
2.6.3 Triton X-100 and Absolute Ethanol 109
2.6.3.1 The solubilisation mechanism in polar solutions – like dissolves like 109
2.6.3.2 Triton X-100 110
2.6.3.3 Absolute Ethanol 111
2.6.3.4 Triton X-100 and absolute ethanol in combination with LPS 112
2.6.3.5 Impact of Triton X-100 and Ethanol on LPS solubility 113

Chapter 3 - Ligand binding structural studies of co-crystallised rfhSP-D and *H. influenzae* Eagan CA7 PS 117

3.1 Introduction 117
3.1.1 *H. influenzae* Eagan wild-type type b LPS 118
3.1.2 *H. influenzae* Eagan 4A and CA7 mutant LPS 119
3.2 Results 120
3.2.1 Protein, LPS and buffer constituents 120
3.2.2 Hydrolysis of CA7 LPS 121
3.2.3 Co-crystallisation of rfhSP-D and CA7 PS 121
3.2.3.1 Initial structure screens 121
3.2.3.1.1 SS1 condition 5: RS20 A5 122
3.2.3.1.2 SS1 condition 28: RS21 A4 123
3.2.3.1.3 SS1 condition 44: RS21 D2 123
3.2.3.1.4 SS1 condition 40: RS21 C4 123
3.2.3.2 Follow up screens 124
3.2.3.2.1 Follow up to SS1 RS20 A5: RS32 125
3.2.3.2.2 Follow up to SS1 RS21 D2: RS34 125
3.2.3.2.3 Follow up to SS RS21 C4: RS33 126
3.2.3.3 Cryo-protection of rfhSP-D and CA7 PS crystals 126
3.2.3.3.1 Cryo-protection strategies 126
3.2.3.3.2 RS20 A5 (SS1 - C5): Cryo-protection information 127
3.2.3.3.3 RS32 (follow up to RS20 A5, SS1 - C5): Cryo-protection information 128
3.2.3.3.4 RS34 (follow up to RS21 D2, SS1 - C44) – Cryo-protection information 129
3.2.4 X-ray diffraction and Data collection 130

3.2.4.1 RS32 A12 Crystal – Data collection 131
3.2.4.2 RS34 B21 Crystal - Data collection 132
3.2.5 Data Processing 132
3.2.5.1 fast_dp data processing 132
3.2.5.2 *iMOSFLM* / *CCP4* data processing 133
3.2.6 Rigid Body Refinement 134
3.2.7 Initial Electron Density maps 135
3.2.7.1 Chain A 136
3.2.7.2 Chain B 136
3.2.7.3 Chain C 137
3.2.7.4 Ca4 138
3.2.7.5 *BLEND* software – Merging RS32 A12 and RS34 B21 into one data set 138
3.2.7.5.1 Electron density map of *BLEND* merged data (RS32 A12 + RS34 B21) 139
3.2.8 Restrained refinement 141
3.2.9 Molecular Model Building - RS34 B21 141
3.2.10 Final Model 146
3.3 Discussion 154
3.3.1 Co-crystallisation of rfhSP-D in combination with CA7 PS 154
3.3.2 Follow up crystals 154
3.3.3 Data Analysis of promising crystals 155
3.3.4 Data processing and analysis 155
3.3.5 Electron Density Maps 156
3.3.6 The final model 156
3.3.6.1 Question 1 – why is the Eagan CA7 HepI rotated by 180° compared to the Eagan 4A HepI? 162
3.3.6.2 Question 2 – How does the novel orientation of HepI shown for Eagan CA7 PS relate to the decreasing affinity of SP-D and rfhSP-D for Eagan wild type, (3 Heps), CA7 (2 Heps) and 4A (1 Hep)? 168
3.4 Summary 172

Chapter 4 - Structural study of co-crystallised rfhSP-D and *S. enterica* minnesota LPS and PS strains. 174

4.1 Introduction 174
4.1.1 Lipopolysaccharide ligands 175
4.1.1.1 *Escherichia coli* O111:B4 smooth LPS 175
4.1.1.2 *Salmonella enterica* minnesota R7 (Rd_1P^-) rough mutant LPS 176

4.1.1.3 *Salmonella enterica* minnesota R5 (RcP^-) rough mutant LPS 177
4.2 Ligand preparation 178
4.2.1 Solubility of *E. coli* O111:B4 LPS in Triton X-100 and Ethanol 178
4.2.1.1 *E. coli* O111:B4 LPS solubilised in Triton X-100 178
4.2.1.2 O111:B4 LPS solubilised in absolute ethanol 179
4.2.1.3 O111:B4 LPS solubilised in triton X-100 and absolute ethanol 181
4.2.2 Solubility limit of *S. enterica* minnesota R7 LPS in Ethanol and Triton X-100 182
4.2.3 Hydrolysis of *S. enterica* minnesota R5 183
4.3 Co-crystallisation of rfhSP-D in combination with solubilised R7 LPS 183
4.3.1 RS44 A2 (SS1-20), RS45 A2 (SS1-38), RS46 B6 (SS2-12), RS47 A1 (SS2-19) 184
4.3.2 RS43 C3 (SS1- 15) 185
4.3.3 RS44 B4 (SS1- 28) 185
4.4 RS44 A2 (SS1-20), RS45 A2 (SS1-38), RS46 B6 (SS2-12), RS47 A1 (SS2-19) 186
4.4.1 Cryo-protection 186
4.4.2 Data collection 186
4.5 RS43 C3 (SS1–C5) 187
4.6 RS44 B4 (SS1-C28) 189
4.6.1 Cryo-protection 189
4.6.2 Data collection 190
4.6.3 Data Processing 191
4.6.4 Rigid Body Refinement 192
4.6.5 Initial Electron Density map 192
4.6.6 Restrained Refinement 194
4.6.7 Molecular Model Building 195
4.6.8 Final model 199
4.7 Summary and Conclusions 203
Chapter 5 - General discussion, Conclusions and Future Work 206
5.1 Co-crystallisation and crystal form 206
5.2 Calcium binding 206
5.3 Recognition and binding of LPS by rfhSP-D 209
5.3.1 Orientation of the bound Heptose 209
5.4 Implications and Future Work 210
References 213

List of Figures

Chapter 1 Page

- Figure 1.1 - A representation of both types of cell wall present in bacteria. 8
- Figure 1.2 - The classification of rough LPS mutants according to their core length into chemotypes Ra-Re. 10
- Figure 1.3 A representation of a typical *H. influenzae* rough LPS. 18
- Figure 1.4 - A 3D and 2D representation of the typical CTLD (CRD) fold. 22
- Figure 1.5 - A representation of a CRD trimer. 41
- Figure 1.6 - Collectin trimerisation and multimerisation 42
- Figure 1.7 - Sequence analysis for 16 collectins, displaying conserved residues. 45
- Figure 1.8 – A representation of the coordination bonds of Ca1 in the CRD. 51
- Figure 1.9 - C-terminal sequences of collectins displaying residues that coordinate the Ca1. 57
- Figure 1.10 - 3D structure of ligand-free rfhSP-D. 61
- Figure 1.11 – Ligand-free structure of rfhSP-D and the effect of the absence or presence of Ca4. 66
- Figure 1.12 – Ligand-bound rfhSP-D of various ligands tested and reported 72-73

Chapter 2

- Figure 2.1 - A representation of a typical unit cell. 80
- Figure 2.2 - A representation of Bragg's Law. 83
- Figure 2.3 - A representation of the typical data collection set up performed at DLS. 84
- Figure 2.4 - A typical image of an X-ray diffraction SNAP of a protein crystal. 86

- Figure 2.5 - Electron density map of the rfhSP-D binding pocket. 90
- Figure 2.6 - Phase diagram for the crystallisation of macromolecules. 99
- Figure 2.7 - A representation of the sitting-drop vapour diffusion method. 102
- Figure 2.8 - A representation of the Triton X-100 molecule. 110
- Figure 2.9 - A representation of how the surfactant Triton X-100 behaves in a polar solution. 111
- Figure 2.10 - A representation of an ethanol molecule. 112
- Figure 2.11 - A representation of how lipid A behaves in a polar solution. 114
- Figure 2.12 - The proposed mechanism of how Triton X-100 and ethanol can aid in the solubilisation of the LPS. 116

Chapter 3

- Figure 3.1 - An illustration of the *H. influenzae* Eagan Wild-type LPS. 119
- Figure 3.2 - An illustration of the *H. influenzae* Eagan CA7 and 4A LPS. 120
- Figure 3.3 - Typical crystal morphology produced in RS20 A5 well. 122
- Figure 3.4 - Typical crystal morphology produced in RS21 A4 well. 123
- Figure 3.5 - Typical crystal morphology produced in RS32 tray. 125
- Figure 3.6 - Typical crystal morphology produced in RS34 tray. 126
- Figure 3.7 - Images collected at DLS for the RS32 A12 crystal. 131
- Figure 3.8 - Images collected at DLS for the RS34 B21 crystal. 132
- Figure 3.9 - RS32 A12 and RS34 B21 electron density maps displayed side by side. 135
- Figure 3.10 – Electron density map of RS34 B21 data set displaying chain A binding pocket. 137

- Figure 3.11 – The fitting of a glucose and a heptose into the electron density map of RS34 B21 binding pocket chain B. 142
- Figure 3.12 - The fitting of a glucose and a heptose into the electron density map of RS34 B21 binding pocket chain B – refined map. 143
- Figure 3.13 – The fitting of both of heptose attempted and fitted into the density of RS34 B21 binding pocket chain B. 145
- Figure 3.14 – The final model following refinement, with the heptose sitting 180° rotated in relation to previously reported structures. 146
- Figure 3.15 - Electron density map of RS34 B21 binding pocket chain B, showing the novel orientation of HepI, stereo image. 149
- Figure 3.16 - An illustration of the novel orientation of flanking residue Arg343 and Glu347, stereo image – before refinement. 150
- Figure 3.17 - An illustration of the novel salt bridge interaction between Arg343 and Glu347, stereo image – after refinement. 151
- Figure 3.18 - The proposed novel orientation of HepI and the novel orientation of both Arg343 and Glu347. 153
- Figure 3.19 – An illustration of the *H. influenzae* CA7 PS molecule and possible sites of binding to rfhSP-D. 158
- Figure 3.20 - The fitting of the putative PEA off HepII into density in the vicinity of Arg349. 163
- Figure 3.21 - An illustration of how the full model of CA7 PS could be oriented vs the reported 4A PS. 164
- Figure 3.22 - An illustration of the superposition of models A, B and C. 165
- Figure 3.23 - An illustration of the superposition of models A and B, displaying the symmetry related molecule chain A and chain B. 167
- Figure 3.24- An illustration of the superposition of models A, C and D, and the CRD of chain B. 169
- Figure 3.25 - An illustration of the superposition of models A, C and D and the location of the loop 4 (L4). 171

Chapter 4

- Figure 4.1 - An illustration of the proposed structure of *E. coli* O111:B4 LPS. 176
- Figure 4.2 - An illustration of the *S. enterica* (minnesota) R5 and R7 LPS. 177
- Figure 4.3 - Images of *E. coli* O111:B4 LPS solubilised in Triton X-100 well. 179

- Figure 4.4 - Images of *E. coli* O111:B4 LPS solubilised in ethanol well. 181
- Figure 4.5 - Images of *E. coli* O111:B4 LPS solubilised in Triton X-100 and ethanol well. 182
- Figure 4.6 - Typical crystal morphology produced in RS44 A2, RS46 B6 and RS47 A1 wells. 185
- Figure 4.7 - Typical crystal morphology produced in RS43 C3 well. 185
- Figure 4.8 - Typical crystal morphology produced in RS44 B4 well. 185
- Figure 4.9 - Images collected at DLS for the RS44 B412 crystal. 190
- Figure 4.10 - Electron density map of RS44 B412 chain A with positive density for Ca1. 193
- Figure 4.11 - Electron density map of RS44 B412 chain B produced following the rigid body. 194
- Figure 4.12 - The fitting of a heptose into the density of RS44 B412 data set chain B, with evidence to model the linkages coming off the heptose – after refinement. 196
- Figure 4.13 - Illustration on why the model 1 was chosen as the best fit model. 198
- Figure 4.14 - Electron density map of RS44 B412 binding pocket chain B, showing the novel orientation the KdoI-HepI-HepII trisaccharide model 1. No protein residues modelled, stereo image. 202

List of Tables

Chapter 1 Page

- Table 1.1 - Site of expression and polypeptide length given for 11 collectins. 23
- Table 1.2 - Reported gram-negative, gram-positive bacteria and mycobacteria that SP-D has been reported to interact with. 32
- Table 1.3 - Viruses that SP-D has been reported to interact with. 34
- Table 1.4 - Fungi and parasite species that SP-D has been reported to interact with. 35
- Table 1.5 - Culmination of current knowledge in regards to N-terminal region, Collagen region, Neck region and CRD. 43
- Table 1.6 – The ranking of the binding affinities of collectins reported in the literature. 55
- Table 1.7 – Key structural studies reported for rMBP-A, rMBP-C, hMBP and rSP-A. 59
- Table 1.8 - Timeline into the rfhSP-D structures deposited in the protein data bank. 62
- Table 1.9 – Direct comparsion of the of reported ligand binding structural studies of rfhSP-D. 68

Chapter 2

- Table 2.1 - The 14 Bravais lattices. 81

Chapter 3

- Table 3.1 - Initial SS1 co-crystallisation trials that showed positive crystal growth. 122
- Table 3.2 - Follow ups trays produced from initial positive hits in SS1. 124
- Table 3.3 - Tray RS32 follow ups conditions. 125
- Table 3.4 - Tray RS34 follow up conditions. 126

- Table 3.5 - Cryo-protection details for RS20 A5 well. 128
- Table 3.6 - Cryo-protection details for RS32 A1 well. 129
- Table 3.7 - Cryo-protection details for RS34 B2 well. 129
- Table 3.8 - Summary of data collection for the crystals with the best resolution achieved. 130
- Table 3.9 - The parameters used for SNAP and data collection for RS32 A12 crystal. 131
- Table 3.10 - The parameters used for SNAP and data collection for RS34 B21 crystal. 132
- Table 3.11 - Data processing parameters for the RS32 A12 and RS34 B21. 134
- Table 3.12 - The data analysis crystallographic parameters for the *Blend* merged data. 140
- Table 3.13 – The refinement statistics of the best model. 147
- Table 3.14 - Ca1 and coordinating residues interatomic distances for RS34 B21 chain B and C. 153

Chapter 4

- Table 4.1 – The illustration of the three experiments performed to investigate the solubility potential of *E.coli* O111:B4. 180
- Table 4.2 - The six wells that produced positive crystal growth using SS1 and SS2. 184
- Table 4.3 - Cryo-protection details for RS44 A2, RS45 A2, RS46 B6 and RS47 A1 wells. 186
- Table 4.4 - The four crystals with the best resolution achieved. 187
- Table 4.5 - Cryo-protection details for RS43 C3 well. 188
- Table 4.6 - Cryo-protection details for RS44 B4 well. 189
- Table 4.7 - The parameters used for SNAP and data collection for RS44 B412 crystal. 190

- Table 4.8 - Data processing parameters for the RS44 B412 data set. 191
- Table 4.9 – The refinement statistics of the best model. 199
- Table 4.10 - Ca1 coordination distances present in chain A, B and C. 200
- Table 4.11 - Distances between the HepI, Ca1 and coordinating residues. 201

Chapter 5

- Table 5.1 - Direct comparison between best model of CA7, best model of *S. minnesota* R5/R7, other $P42_12$ and $P2_1$ crystals and the reported 4A PS, in terms of Ca1, Ca2, Ca3, Ca4 and ligand occupancy. 207

Abbreviations

ATP	Adenosine Triphosphate
BALF	Bronchoalveolar lavage fluid
CCP4	Collaborative Computational Project Number 4
CD-14	Cluster Differentiation 14
CD-91	Cluster Differentiation 91
CL-43	Collectin 43
CL-46	Collectin 46
CL-K1	Collectin Kidney 1
CL-L1	Collectin Liver 1
CL-P1	Collectin Placenta 1
CMC	Critical Micellar Concentration
COPD	Chronic Obstructive Pulmonary Disease
CRD	Carbohydrate Recognition Domain
CRP	C-reactive Protein
CTLD	C-type Lectin Domain
CTLD*	C-type Lectin Like Domain
DAMP	Danger Associated Molecular Pattern
DNA	Deoxyribonucleic Acid
DPPC	Dipalmitoylphosphatidylcholine
FucNAc	N-acetyl-fucosamine
Gal	Galactose
GalNAc	N-acetyl-galactosamine
Glc	Glucose
GlcN	Glucosamine

GlcNAc	N-acetyl-glucosamine
GP-340	Glycoprotein 340
Hep	L-D-Heptose
IAV	Influenza A Virus
IDL	Interstitial Lung Diseases
IPF	Idiopathic Pulmonary Fibrosis
KDO	3-deoxy-D-manno-oct-2-ulosonic acid
LAIR-1	Leukocyte-associated Immunoglobulin-like Receptor 1
LBP	Lipopolysaccharide Binding Protein
LPS	Lipopolysaccharide
LTA	Lipoteichoid Acid
Man	Mannose
ManM	Mannosamine
ManNAc	N-acetyl-mannosamine
MBP	Mannose Binding Protein
MPD	2-methyl-2,4-pentanediol
NLR	NOD-like Receptor
NTHi	Nontypeable *Haemophilus influenzae*
NTS	Non-thyphoidal *Salmonella*
OM	Outer Membrane
OSCAR	Osteoclast-associated Receptor
PAMP	Pathogen Associated Molecular Pattern
PAP	Pulmonary Alveolar Proteinosis
PC	Phosphatidylcholine
PCho	Phosphocholine

PEA	Phosphoethanolamine
PEG	Polyethylene Glycol
PG	Phosphatidyglycerol
PI	Phosphatidylinositol
PPEA	Pyrophosphate Ethanolamine
PRR	Pattern Recognition Domain
RDS	Respiratory Distress Syndrome
rfhSP-D	Recombinant Fragment of Human Surfactant Protein D
Rha	Rhamnose
RLIR	RIG-I-like Protein
RSV	Respiratory Syncytial Virus
SIRPα	Signal-regulatory Protein Alpha
SP-A	Surfactant Protein A
SP-B	Surfactant Protein B
SP-C	Surfactant Protein C
SP-D	Surfactant Protein D
TLR	Toll-Like-Receptor
TNFα	Tumour Necrosis Factor Alpha

Acknowledgments

I have completely enjoyed my time spent studying for my PhD under the supervision of Prof. Trevor Greenhough and Dr. Annette Shrive, where I developed a variety of technical and important life skills, resulting in my personal and professional development during my time at Keele.

I would like to thank Prof. Trevor Greenhough for his support and belief in my work and for sharing part of his immeasurable knowledge of protein X-ray crystallography. I would also like to thank Dr. Annette Shrive for her patience and knowledge of X-ray crystallography, where her door was always open to discuss work or any problems encountered. Thank you both for your patience and help throughout the years. I would also like to thank Prof. Paul Eggleston for his kind words during my first year, by giving me motivation to aim for a great thesis.

To the collaborators that made my work possible I would to thank them for the recombinant human SP-D expressed and purified by Howard W. Clark, Rose-Marie Mackay and Jens Madsen at the University of Southampton, Department of Child Health, Division of Clinical and Experimental Sciences, Faculty of Science, Faculty of Medicine, Sir Henry Wellcome Laboratories, Southampton General Hospital, Southampton, United Kingdom; I would also like to thank Mary E. Deadman and Derek W. Hood at Mammalian Genetics Unit, MRC Harwell, Harwell Science & Innovation Campus, Oxfordshire, United Kingdom for providing with me with the well characterised *H. influenzae* Eagan CA7 LPS.

I would like to acknowledge Ian Burns for his incredible knowledge and expertise in the subject of crystallisation of proteins and specific techniques, and for his help in performing a secondary purification of the recombinant fragment human SP-D. To all my colleagues and friends in our research group I would like to acknowledge your help in producing an astonishing environment in our office, where our daily discussions into science and politics gave me the motivation succeed, thank you Jenny Moran, Robert

Williams, Jamie Littlejohn, William Neale, Omar Alhamd, Sameer Mahmood and Anas Hamad.

Finally, I would like to thank my family for the amazing support and belief in me, thank you Dad, Mum, Sister and my nephew Tomas, for giving me the confidence and strength to complete this thesis. To my girlfriend Adriana I would like to thank you for being one of a kind, I could not have done it without you.

Chapter 1 - Introduction

1.1 The human lung and the immune system

The immune system in vertebrates is composed of two distinct mechanisms that work together, with the aim of eliminating pathogens and particles. The innate immune system is the first line of defence, with non-specific properties common to most species, leading to the response to pathogens and particles in a non-specific manner. The adaptive immune system is based on an immunological memory after an initial response to a specific pathogen, leading to an enhanced response to subsequent encounters with that specific pathogen.

The manner in which the host's immune system destroys and clears both commensal and pathogenic microbes in the lung is mediated by various different components of the innate and adaptive immune system. The clearance of foreign pathogens and particles occurs through an adequate immune response, with the aim of minimising inflammation and maintaining normal lung homeostasis.

1.1.1 Innate Immunity in the lung

The innate immune system acts as the first line of defence and includes a variety of secreted molecules and physical barriers, with the aim of protecting the host while maintaining normal lung physiology. Anatomically the lung can be divided into two sub-structures; the conducting airway, which includes the trachea, bronchi and bronchioles, and the peripheral airway, consisting of the alveoli. Both of these sub-structures operate distinct roles in the innate immune defence of the lung, with various different cells presented in the epithelium layer.

The conducting airway epithelium has been reported to be mostly decorated with ciliated cells and secretory cells such as serous, club cells (also known as clara or nonciliated bronchoalveolar cells) and globet cells. Underneath the variety of cells lining

the epithelium, submucosal glands line the inner layer of the epithelium. The various cells present in the conducting airway and the submucosal glands work together in secreting host-defence proteins onto the airway surface, in response to environmental stimuli. A variety of defence molecules can be secreted including β-defensins, lysozyme, lactoferrin, cathelicidin LL37 and surfactant protein A and D (SP-A and SP-D, respectively), leading to aggregation, trapping and destruction of microbes and particles (**Whitsett & Alenghat, 2015**).

In the peripheral airway, only two types of cell line the epithelium; squamous type I alveolar cells, covering 90% of the surface of the alveolar surface, and cuboidal type II epithelial cells, important in the production of surfactant lipids and proteins and crucial in preventing atelectasis (lung collapse) (**Whitsett & Alenghat, 2015**).

The innate immune system not only has the ability to recognise microbial infections and elicit immediate defence, but also has the ability to generate long lasting adaptive immunity. The innate immune system relies on recognising common structural and functional features, which are associated with different classes of microorganisms, and determining the type of infection encountered, impacting and instructing the lymphocytes into inducing an appropriate class of immune response (**Iwasaki & Medzhitov, 2015**).

1.1.2 Primary antimicrobial defences in the alveolar lining layer

The mucus and alveolar lining layer is composed of a variety of different immune molecules, including alveolar macrophages, lysozymes, defensins and cathelicidin. Defensins belong to a large family of antimicrobial peptides, divided into β and α defensins, with α-defensins expressed in neutrophils, while β-defensins are widely expressed in epithelia cells. Both of these contain a broad spectrum of activity, with the ability to recognise bacteria, fungi and viruses, and to recruit inflammatory cells, promoting innate and adaptive immune responses. Cathelicidin LL37 is an antimicrobial peptide

produced by neutrophils and respiratory epithelial cells, with a similar role to the defensins (**Tecle et al., 2010**). Mucins are glycoproteins with repeating amino acids rich in serine and threonine, which can form large oligomeric structures with antimicrobial and anti-inflammatory properties (**Voynow et al., 2009**). Surfactant proteins A and D are pattern recognition receptors (PRRs) present in the alveolar lining layer with antimicrobial properties, these will be discussed in more detail in 1.3 and 1.4. Surfactant proteins also impact on the normal physiology of the lung, and this is discussed in more detail below. If pathogens successfully evade the physical epithelial and mucus barriers, PRRs are activated by epithelial cells and secreted to the site of "invasion".

1.1.3 PRRs in the lung

A variety of PRRs have been reported in the human lung and these can be divided into two separate subgroups; (1) Signalling PRRs, which include membrane bound and cytoplasmic PRRs; (2) Endocytic PRRs, which promote the attachment, engulfment and destruction of invading microorganisms, by phagocytosis, with the ability to recognise carbohydrates. PRRs are primitive components of the immune system, with a broad variety of secreted, cytoplasmic and membrane bound PRRs reported to date. Secreted PRRs include collectins, ficolins, pentraxins and complement receptors; membrane bound PRRs include Toll-Like-Receptors (TLRs), C-type lectins Group I (Mannose Receptor) and Group II (asiaglycoprotein), while cytoplasmic PRRs include NOD-like receptors (NLR) and RIG-I-like protein (RLIR). (**Lambrecht & Hammad, 2012**; **Basset et al., 2003**; **Iwasaki & Medzhitov, 2015**).

The primary role of PRRs is the recognition of pathogen associated molecular patterns (PAMPs) and danger associated molecular patterns (DAMPs), present on the surface of microbial pathogens and viral nucleic acid (**Iwasaki & Medzhitov, 2015**; **Newton & Dixit, 2012**), and the activation of phagocytic cells, secretory cells and natural killer cells with the effect of the safe elimination of the unwanted cells and pathogens. PRRs have the ability to distinguish DAMPs and PAMPs from self-cells (host cells), due to

the molecular patterns associated with foreign microorganisms and particles differing from the molecular patterns presented by self-cells, in other words by means of cell-surface recognition (**Jiang & Chess, 2009**).

PAMPs include a variety of microbial structures such as lipopolysaccharides (LPS) from gram-negative bacteria, lipoproteins, lipoteichoic acid (LTA) from gram-positive bacteria, glucans from fungal cell walls and peptidoglycan found in the bacteria cell wall. DAMPs are endogenous molecules, normally found inside cells but which are released during necrosis. DAMPs include ATP, cytoplasmic proteins, uric acid, nuclear proteins and DNA (**Basset et al., 2003**; **Newton & Dixit, 2012**; **Iwasaki & Medzhitov, 2015**).

When the PRRs recognise the PAMPs or DAMPs, intracellular or extracellular signals inform the host of the presence of infection, triggering pro-inflammatory and antimicrobial responses towards the pathogen or particle. This results in either agglutination, opsonisation and/or activation of complement and subsequent activation of non-specific innate immune cells **(Mogensen, 2009**; **Newton & Dixit, 2012**; **Iwasaki & Medzhitov, 2015)**.

1.1.4 Pulmonary surfactant- important in the lung normal physiology

Surfactant proteins are important in ensuring a normal lung physiology, acting in combination with pulmonary surfactants. Pulmonary surfactants are a mixture of lipids (90%) and proteins (10%), where 80% of the lipids are phospholipids, produced by alveolar type II cells in the peripheral airway, that work together in lowering the surface tension at the air-liquid interface (**Ogasawara et al., 1992**; **Glasser & Mallampalli, 2012**). The elastic structure of the alveoli and the necessity to dynamically diminish the surface tension at the gas-liquid interface at the alveolar surface, presents a biophysical and host defence challenge. A failure in one of these dynamic processes, can lead to infection, tissue damage, alveolar capillary leakage and potential atelectasis (lung collapse), impairing ventilation and the pathogenesis of respiratory distress syndromes in infants and

adults. Surfactant deficiency can result in Respiratory Distress Syndrome (RDS), which occurs when infants are born prematurely, before their surfactant biosynthetic machinery has matured. Treatment with exogenous surfactants replacements has been reported to reduce morbidity and mortality (**Clark et al., 2010**).

1.1.4.1 Lamellar bodies and tubular myelin

Following biosynthesis of surfactant proteins and lipids in the alveolar type II cells, these are packaged within the lamellar bodies prior to secretion into the alveolar surface (**Goerke, 1998**; **O'Reilly et al., 1988**). Upon secretion, these lamellar bodies are transformed into tubular myelin, which is a mesh like structure, suggested to be a precursor of the surfactant monolayer that is present at the air-liquid interface. The tubular layer is then absorbed by alveolar surface and compressed, facilitating the reduction in surface tension (**Barreira et al., 2011**). The compression of the tubular layer has been suggested to cause the release of surfactant proteins (**Goerke, 1998**). Surfactant proteins and lipids are constantly being recycled, with a vast number of surfactants exchanged between in and out of alveolar surface, maintaining a constant surfactant pool size (**Barreira et al., 2011**).

1.1.4.2 Surfactant Lipids

Surfactant lipids are composed of approximately 85-90% phospholipids and 10-15% neutral lipids. Phosphatidylcholine (PC) is the most predominant phospholipid accounting for 70-80% of all phospholipids (**Haagsman et al., 2008**). The most common species of PC is termed dipalmitoylphosphatidylcholine (DPPC), which is a saturated phospholipid comprising approximately 50% of all phospholipids and with a key role in lowering the surface tension. Phosphatidyglycerol (PG) and phosphatidylinositol (PI) account for 8-15% of the phospholipid portion, and the remainder is comprised of minor phospholipids (**Blanco & Perez-Gil, 2007**; **Picardi et al., 2011**). Neutral lipids are composed of a variety of different lipids, including cholesterol (30-40%), free fatty acids

(25-30%), cholesterol ester (12%), monoglycerides (10-15%), diglycerides (14%) and triglycerides (8%) (**Markart et al., 2007**).

1.1.4.3 Surfactant Proteins

To date four surfactant proteins have been reported, SP-A, SP-B, SP-C and SP-D. SP-B and SP-C are small hydrophobic proteins, crucial in phospholipid packing, organisation of surfactant levels and important in lowering the surface tension in the air-liquid interface in the lung following expiration. SP-A and SP-D are surfactant proteins that belong to the collectin protein family and are important "players" in the host innate immunity response, with the ability to recognise a wide variety of pathogens in a calcium dependent binding mechanism (**Wright, 2005**). Pathogen recognition leads to agglutination, opsonisation and phagocytosis and complement activation through the lectin pathway, towards the clearance of the targets. Transgenic studies into surfactant homeostasis have showed that SP-D and SP-A are also important in regulating the levels of surfactant lipids and tubular myelin, the normal shape of type II cells and regulation of inflammation, highlighting a second role in ensuring and achieving a disease-free lung (**Korfhagen et al., 1998**; **Botas et al., 1998**). Therefore, it has been suggested that the collectins SP-A and SP-D possess the dual role of maintaining a disease-free lung, by stimulating the clearance of pathogens, which include a variety of microbes, allergens and dying cells, while maintaining an inflammation free mucosal environment (**Weis et al., 1998**; **Haczku, 2008**).

The experimental work performed in this thesis used a recombinant fragment of human SP-D (rfhSP-D) in combination with various ligands, investigated in structural studies by means of X-ray crystallography. Structural studies are of considerable importance in elucidating the binding mechanisms of SP-D and rfhSP-D towards natural ligands such as bacterial lipopolysaccharides (LPS).

1.2 Lipopolysaccharides (LPS)

Bacterial surface lipopolysaccharides (LPS) are an important constituent of the outer membrane of gram-negative bacteria and a target ligand for immune system defence molecules. LPS has the ability to cause pyrogenic effects on living organisms and was initially reported to be an endotoxin by Richard Pfeiter in 1892. Richard Pfeiter showed that *Vibrio cholerae,* which today is known to cause cholera, produced a heat stable toxic substance with insoluble properties. Richard Pfeiter postulated that the toxic particle was located inside the cell membrane and named it endo (inside- Classical Greek)-toxin (**Caroff et al., 2002**; **Alexander & Rietschel, 2001**).

The LPS of different species and phenotypes were used in the experimental work presented in this thesis. In this sub-chapter a general overview of lipopolysaccharides from *Haemophilus influenzae, Escherichia coli* and *Salmonella enterica* (minnesota) species, and the LPS type (smooth - long; rough - short), will be given.

1.2.1 Bacteria- gram-negative Vs gram-positive

Bacteria can be divided into gram-negative or gram-positive, which reflect structural differences in the cell envelope (**Fig. 1.1**). All bacteria species possess a cytoplasmic/plasma membrane composed of a phospholipid bilayer, surrounding the cytosol and providing a physical barrier for the movement of molecules in and out of the cell. The cell wall of bacteria is composed of peptidoglycan, which is a rigid layer that confers shape and osmotic strength to the bacterial cell, enclosing the cytoplasmic membrane. The major differences between the two types of cell wall are the difference in size and thickness of the peptidoglycan layer, with gram-positive bacteria containing around 95% of peptidoglycan in the cell wall, while gram negative contain approximately 10%; and the number of membranes present, with gram-positive bacteria containing one membrane (plasma membrane) and gram-negative containing two membranes (plasma membrane and outer membrane). Peptidoglycan in gram-positive bacteria constitutes the

outer layer and external portion of the cell wall, acting as an anchor for lipoteichoid acid (LTA) (an antigen similar to LPS), while gram-negative bacteria possess an outer membrane (OM) that surrounds the thin layer of peptidoglycans and acts as an anchor for LPS.

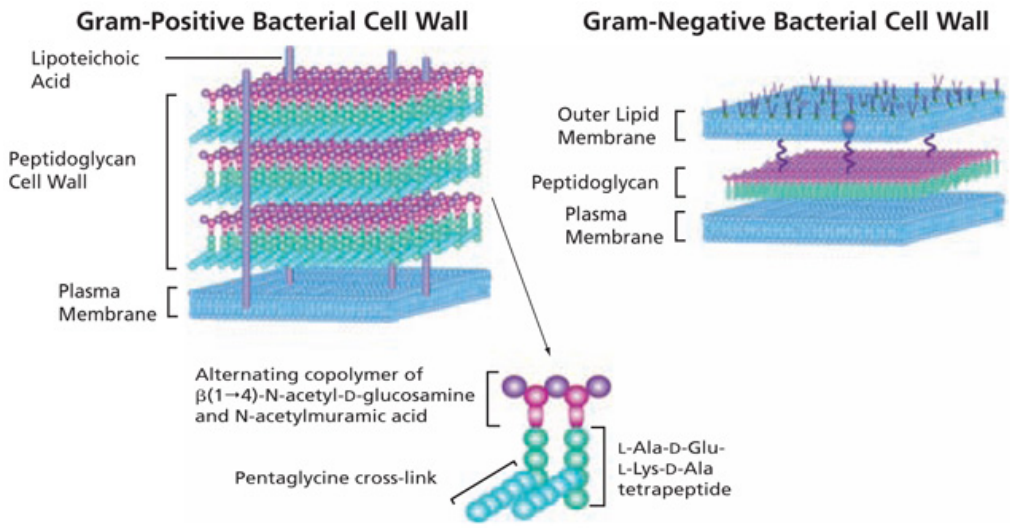


Fig. 1.1 A representation of both types of cell wall present in bacteria. The gram-positive bacteria cell wall is composed of a thick layer of peptidoglycan which the LTAs are anchored to. In the gram-negative bacteria cell wall this peptidoglycan layer is much thinner, with an outer membrane (OM) surrounding the thin peptidoglycan layer and acting as the site of anchoring for the LPS. Image taken from Sigma-Aldrich (**"Peptidoglycans"**).

1.2.2 Lipopolysaccharide (LPS) - Biological Role

Lipopolysaccharides (LPS) are major components of the outer membrane of gram-negative bacteria and contribute to the integrity and stability of the OM. The OM has the ability to act as permeability barrier to various different molecules such as detergents, antibiotics and toxic dyes and metals, and due to their external location can also participate in host-bacterium interactions such as adhesion, colonisation, virulence and symbiosis, making it indispensable for the viability of bacterial species. LPS is a strong endotoxin with the ability to elicit innate immune responses, playing a crucial role in the pathogenesis of gram-negative bacterial infections. LPS has the ability to cause septic shock if untreated, with very small quantities (one-nano-gram) of LPS necessary to induce

and cause septic shock (**Takeuchi & Akira, 2010**; **De Castro et al., 2010**; **Kucerka et al., 2008**). LPS is most often anchored into the OM of bacteria, where the Lipid A is embedded in the cell membrane, but when the bacteria multiply or die, LPS is released into the blood stream and interacts with a variety of immune receptors; lipid A is the main inducer of immunological reactions by the host caused by gram-negative bacteria (**Raetz et al., 1991**; **Kabanov & Prokhorenko, 2011**). Even though LPS has the ability to cause damage to the host, it is presented as a pathogen associated molecular pattern (PAMP) that can be recognised by the immune system, leading to the opportunity for elimination of the pathogen. Since PAMPs (make the bacteria more easily recognisable by host immune PRRs), bacteria have evolved different mechanisms that lead to LPS structure modification, which can lead to them becoming invisible and unrecognisable by the immune system and resistant to other defence components such as complement and antibiotics (**Wu et al., 2013**).

1.2.3 LPS common structure

The LPS common structure consists of a Lipid A composed of β1-6 di-glucosamine with an asymmetrical distribution of fatty acids, embedded in the cell membrane and acting as a hydrophobic membrane anchor. This is glycosidically linked to the core region by 3-deoxy-D-manno-oct-2-ulosonic acid (Kdo) (**Fig. 1.2**). The core is divided into two parts, the inner core region and the outer core region. The inner core region links the lipid A to the remainder of the LPS, through Kdo. The Kdo has been shown to be an essential sugar for the majority of LPS. The inner core also contains heptose (Hep) residues, which are normally substituted by phosphate (P) or phosphoethanolamine (PEA) or pyrophosphate ethanolamine (PPEA). The outer core usually consists of hexoses such as D-glucose, D-galactose, D-glucosamine, D-galactosamine or N-acteyl derivatives. Finally, O-specific antigen/polysaccharide structures, when present, confer serotype specificity to the strain and consist of up to 70 repeating oligosaccharide units formed by 2-8 monosaccharide moieties, with extremely high structural variability within the bacterial

species. LPS can be divided into two different phenotype categories; smooth, containing the O-antigen and rough, which lacks the O-antigen (**Brade et al., 1985**; **Alexander & Rietschel, 2001**). There is heterogeneity between strains of the same species and between different bacteria. Lipid A seems to be highly conserved throughout bacteria species. The number of Kdos present can also vary between species, with some different LPS containing either 1, 2, or 3 Kdos. The inner core is also very varied across the bacterial species, with different strains containing a varied number of heptoses followed by a variation in the substitution of those heptoses; some LPS have been reported to contain just Kdo (**Caroff et al., 2002**; **Gronow et al., 2009**; **Cox et al., 2002**).

Rough LPS or rough mutant LPS with varying inner core and outer core chain lengths can be classified according to their length into chemotypes Ra-Re, with the Ra showing a complete inner and outer core, while Re mutants consist only of Kdo groups. See **Fig. 1.2** for an illustration of the possible rough LPS variants using a *Salmonella enterica* Ra core LPS.

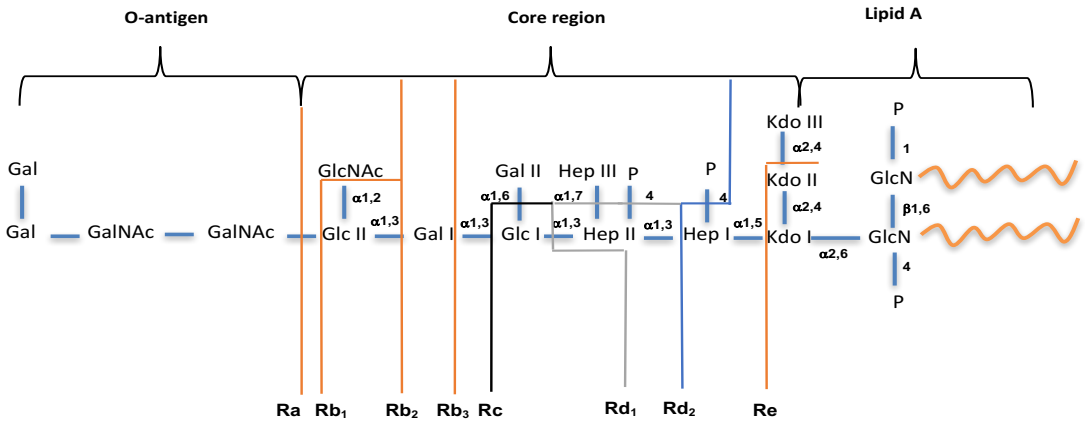


Fig. 1.2 The classification of rough LPS mutants according to their core length into chemotypes Ra-Re. Re contains only Kdo, while the Rd_1, Rd_2 and Rc mutants are progressively longer. Rb_3 divides the core region into inner core and outer core. The Ra mutant contains the longest core region but lacking the O-antigen and rendering it a rough LPS; smooth LPS contains the O-antigen. In the figure a typical structure of LPS is given (Ra core *Salmonella enterica*) displaying the linkage between the oligosaccharides in the core region and lipid A region. Glc- Glucose; Hep- L-D-Heptose, KDO- 3-deoxy-D-manno-2-octulosonic acid; Gal- Galactose; GlcNAc - N-acetyl-glucosamine; GalNAc - N-acetyl-galactosamine; P- Phosphate or Phosphoethanolamine (PEA). Image adapted from **Kato, 1993** and **Mansfield & Forsythe, 2001**.

Variations in LPS between species occur in the number of and links between the different carbohydrates, with mutations in the biosynthesis often suggested as one of the reasons for the diversity seen. In the structures presented in this thesis, three separate species of LPS from three different species of bacteria were investigated. Both smooth and rough LPS were investigated as potential natural ligands for the structural studies. An overview is given below of the three-separate species of bacteria used, highlighting each typical LPS structure and its heterogeneous nature.

1.2.4 *Escherichia coli, Salmonella enterica* (serotype minnesota) and *Haemophilus influenzae*

The LPS from three different bacteria have been investigated as natural ligands for rfhSP-D. Due to *E. coli and S. enterica* (serotype minnesota) belonging to the same family of gram-negative bacteria, denominated *Enterobacteriaceae*, these will be discussed together as some structural features are conserved in this family. *H. influenzae* belongs to the family of *Pasteurellaceae* bacteria.

1.2.4.1 *Enterobacteriaceae – Escherichia coli* and *Salmonella enterica* (serotype minnesota)

The family *Enterobacteriacea* is exclusively composed of gram negatives species such as *Salmonella enterica, Escherichia coli, Shigella dysenteriae, Yersinia pestis, Klebsiella pneumoniae* and *Serratia marcescens*. All of the *Enterobacteriaceae* contain LPS, as this gives the bacteria enhanced survival, by providing membrane stability and acting as a protection mechanism against antibiotics and bactericidal components of the hosts (**Holst, 2007**).

The LPS of the *Enterobactereacease* family share a common LPS architecture with limited structural diversity for the core region, but a high structural variation for the O-antigen (**Brade et al., 1985**). In the work described in this thesis, both rough and smooth

LPS from *Escherichia coli* and a rough mutant LPS of *Salmonella enterica* (minnesota) were used.

Escherichia coli is a facultatively anaerobic, rod shaped, gram-negative bacterium. *E. coli* contains the typical gram-negative bacteria LPS present in the outer membrane and can cause sepsis/meningitis, urinary tract infections and diarrhoea (food poisoning), depending on the pathogenic strain. *E. coli* is the predominant facultative anaerobe of the human colonic flora with mutual benefits for host and coloniser. Some strains have acquired specific virulence factors, increasing their ability to adapt to new niches and causing a broad range of diseases (**Stenutz et al., 2006**; **Wu et al., 2013**).

Salmonella enterica is a facultatively anaerobic, rod shape, flagellated gram-negative bacteria, which can be further divided into six subspecies and over 2500 serovars (**Chessa et al., 2014**). *Salmonella enterica* can be divided into thyphoidal and non-thyphoidal *Salmonella* (NTS). The subtype enterica serovar used for the experimental work presented here was the NTS serovar minnesota. This type of serovar has been reported to be a major cause of gastrointestinal illness, which includes intestinal infection, diarrhoea, fever and abdominal cramps, after the consumption of infected food or water (**Steinebrunner et al., 2013**; **Eckerle et al., 2010**). *S. enterica* minnesota has been reported to have the ability to act as a severe pathogen in immunocompromised hosts, by causing septic urinary tract infections (**Steinebrunner et al., 2013**).

1.2.4.2 *E. coli* and *S. enterica* (minnesota) LPS

The LPS from *E. coli* and *S. enterica* (minnesota) contain a common structure of an acylated 1, 4 diphosphorylated β (1→6) linked glucosamine disaccharide (lipid A). The lipid A is attached to a fairly conserved Kdo structure, which in turn is attached to a L-glycero-α-D-manno-heptopyranose (Hep) (**Fig. 1.2**). *S. enterica* (minnesota) LPS contains three Kdos, while in *E. coli* two or three Kdos can be present, depending on the core type. Phosphate residues are also present, highlighting its role in the outer membrane stability.

In *E. coli* LPS the oligosaccharide core can be subdivided into the inner and outer core, with five different core types being reported (R1 - 69.4% of all *E. coli* species tested, R2- 11.1%, R3- 11.1%, R4-2.8% and K-12 -5.6%) (**Amor et al., 2000**; **Muller-Loennies et al., 2003**; **Muller-Loennies et al., 2002**; **Vinogradov et al., 1999**), that differ in the outer core structure. For *S. enterica* (minnesota) two cores have been reported, the Ra-core type and non-Ra core type (**Tsang et al., 1991**; **Olsthoorn et al., 1998**). Both species of LPS can be divided into smooth, which contains the O-polysaccharides attached to a sugar in the outer core, or rough that contains no O-polysaccharide. The rough mutants are normally the result of a mutation in the biosynthesis. As shown in **Fig. 1.2,** rough LPS can be divided into Ra, Rb, Rc, Rd and Re, with the Ra containing a complete core, while the Re contains a short core, in this case just Kdo groups (**Klein et al., 2013**; **Mansfield & Forsythe, 2001**; **Nnalue, 1999**).

1.2.4.2.1 O-antigen/O-polysaccharide

The O-antigen is typical of smooth LPS in isolates from *Enterobacteriaceae* and other gram-negative bacteria. These possess a high structural diversity of O-antigens and these have been reviewed extensively in the literature. In *E. coli* over 180 O-antigens/serotypes forms have been reported, while in *S. enterica* (minnesota) over 46 antigens/serotypes have been discovered (**Stenutz et al., 2006**; **Heinrichs et al., 1998**; **Liu et al., 2013**).

Upon the completion of the outer core (Ra mutant), pre-assembled O-antigen can ligate the pre-assembled O-antigen to the Lipid A – Core region, through the action of various different enzymes that ensure transport and ligation into one complex (**Han et al., 2014**). The O-antigen reveals a large number of possible glycoconjugates making up the O-antigen in the *Enterobacteriaceae* family. The O-antigen is made up of a repeating unit composed of 2 to 8 sugar residues, with an n number of possible repeats and specific to each sub-species and serotype. *S. enterica* (minnesota) was shown to have on average 70 repeat units and *E. coli* to have on average 27-28 repeat units in the O-antigen

(**Peterson & McGroarty, 1985**), with the ability to present itself as linear, branched or double structure, with each sugar residue in either an α or β configuration, **Fig.1.2** displays a typical O-antigen repeat unit of *S. enterica*. The O-antigens can be divided into two major groups: N-acteylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc) or Gal as the first sugar in the O unit (**Liu et al., 2013**; **Stenutz et al., 2006**). Common sugars found in the *E. coli* O-antigen are D-Glc, D-GlcNAc, D-Gal, D-GalNAc, D-Man and L-Rha, in both anomeric configurations, while L-FucNAc has only been found in one anomeric configuration (**Stenutz et al., 2006**). For *Salmonella,* the common sugars found in the O-polysaccharide are D-GlcNac, D-GalNac, D-Glc and D-Gal (**Fig. 1.2**) (**Liu et al., 2013**).

1.2.4.2.2 Biosynthesis of *E. coli* and *S. enterica* LPS structure

E. coli and *S. enterica* share common mutations in the biosynthesis of their LPS structures. The main difference between both species is that in *E. coli* there is a higher variability of possible cores, which in turn means that more genes are expressing for more enzymes. This variability can be seen in the variation of the outer core and O-antigen region; the inner core region remains fairly conserved within and between both species. The fact that *Salmonella* and *E. coli* contain similar variations can be attributed to both organisms sharing a common ancestor, but also the fact that the core region of the LPS is known to be important in the stability of the outer membrane in bacteria cells, so it is not expected to contain a high degree of difference across gram-negative bacteria. LPS biosynthesis in *E. coli* and *S. enterica* have been reviewed and reported in a variety of studies (**Raetz & Whitefield, 2002**; **Gronow & Brade, 2001**; **Brade et al., 1984**; **Luderitz et al., 1967**; **Klein et al., 2013**; **Liu et al., 2013**; **Stenutz et al., 2006**).

1.2.4.2.3 Rough, rough mutant and smooth LPS

Rough mutant LPS occur when mutations affect the biosynthesis of the LPS, inhibiting enzyme activation or addition of glycosyl/heptosyl sugar precursors, leading to an incomplete core region without the addition of O-antigen (see **Fig. 1.2**). Attention is

needed when referring to rough mutant LPS and rough LPS. Rough mutant LPS strains lack the O-antigen due to a genetic defect, while rough LPS indicates the absence of an O-antigen independent of whether this due to a genetic error or not (**Gronow & Brade, 2001**; **De Castro et al., 2010**). This is the case for some non-enteric bacteria such as *Neisseria, Haemophilus* and *Chlamydia*, which produce only rough type LPS, naturally lacking the O-antigen (**Cox et al., 2002**; **Gronow et al., 2009**; **Schweda et al., 2007**). In the case of *E. coli* and *Salmonella,* rough mutant phenotypes are the result of defects in biosynthesis. Both species of bacteria contain genetic similarities in regards to the production of LPS mutant cores and have been reported to contain defects in the same genes (**Parker et al., 1992**; **Parker et al., 1992a**).

The ability of rough mutants to multiply *in vivo* suggests that the O-chain is dispensable for bacterial survival. However, there are indications that pathogenic enterobacteria can only persist and survive if the O-antigen is present, as it protects the bacteria against uptake for phagocytosis and the attack of serum complement (**Vukajlovich, 1986**) and provides enhanced bacterial virulence and survival opportunities. The O-antigen also offers a vast diversity of glycoconjungates, contributing to a selective advantage to adjust to a particular environment by the bacteria (**Liu et al., 2013**).

Due to the high heterogeneity, smooth and rough LPS interact with molecules of the immune system differently. Rough LPS has been shown to activate a broader spectrum of cells with higher potency than smooth LPS. Toll-like receptor 4 (TLR4) present in mast cells was reported to be readily activated by rough LPS, while the smooth LPS interactions required the activation of the co-receptors lipopolysaccharide binding protein (LBP) and cluster of differentiation 14 (CD-14). Tumour necrosis factor alpha (TNF-α) has also been shown to have increased expression when rough LPS is detected rather than smooth (**Huber et al., 2006**), with rough mutant LPS being more susceptible than smooth to growth inhibition by collectins (**Wu et al., 2003**).

LPS has been suggested to exist in a heterogeneous form where smooth and rough are mixed together, with the aim of ensuring rapid adaptation to specific environments (**Freudenberg et al., 2008**). This ability to change between smooth (S) and rough (R) has been reported to occur in *Enterobacteriaceae* species (**McCallum et al., 1989**; **Walsh & Moran, 1997**). The variation of the O-antigen has been suggested to offer bacteria selective advantage to specific environments. One interesting fact to consider is that some species of bacteria have been reported to lose the O-antigen when cultured in laboratories, by "self-inflicting" or controlling mutations in the biosynthesis or transfers. This suggests that the bacteria have the ability to adapt to different environments, by modulation of LPS composition and that the O-antigen is only required to be present in a natural environment (**Reeves, 1995**). This mechanism remains mostly unexplored.

1.2.4.3 *Pasteurellaceae – Haemophilus influenzae*

Haemophilus influenzae is a coccobacillus and facultatively anaerobic gram-negative bacteria, belonging to the Pasteurellaceae family. *H. influenzae* is divided into typeable and nontypeable strains, depending on the presence of a polysaccharide capsule. The typeable strains containing capsules are classified into six serotypes (a to f). Nonencapsulated forms are referred as nontypeable *H. influenzae* (NTHi) (**King, 2012**).

NTHi strains are known to be the main pathogens in otitis media and acute and chronic lower respiratory tract infections (**Schweda et al., 2007**). In terms of typeable strains, *H. influenzae* type b is the most prominent typeable form. Type b capsular strains are associated with invasive bacteraemia diseases, such as meningitis, epiglottitis, cellulitis and pneumonia (**King, 2012**).

The capsule of the typeable form of *H. influenzae* has been the focus of extensive research in order to produce vaccines against it. Glycoconjugates based on the capsular polysaccharide structure of *H. influenzae* type b have been developed and are available commercially (Hib vaccine), and these have proved to be successful against *H. influenzae*

type b by acting as prophylactic treatment and reducing morbidity in children ("***Haemophilus Influenzae* Type B (Hib) - NHS Choices**"). The Hib vaccine uses a capsular polysaccharide antigen, which incorporates a synthetic bactericidal carbohydrate antigen mimicking the effect of the type b *H. influenzae* (**Verez-Bencomo et al., 2004**). These vaccines are only effective against the type b capsular *H. influenzae* due to their specificity, having no effect on the other typeable serotypes or NTHi strains (**Schweda et al., 2007**).

1.2.4.3.1 LPS from *Haemophilus influenzae*

Haemophilus influenzae LPS contains a natural rough phenotype, independent of any mutations of the biosynthesis, consisting of a short LPS lacking the O-antigen. Studies of *H. influenzae* LPS have identified a conserved glucose-substituted triheptosyl inner core moiety; L-α-D-Hep-(1->2)-[PEA ->6]-L- α-D-Hep-(1->3)-[β-D-Glc-(1->4)]-L- α-D-Hep linked to a lipid A via a Kdo 4-phosphate, where variations can occur from R1, R2, R3, R4 and Y as shown in **Fig. 1.3** (**Schweda et al., 2007**), which shows a representation of the typical LPS from *H. influenzae* and its variations. *H. influenzae* LPS contains only one KDO, phosphorylated at position 4, whereas a second Kdo is present for *E. coli* and *S. enterica* (minnesota).

LPS from *H. influenzae* has the characteristic feature of having extensive intra- and inter-strain heterogeneity in the glycoform populations, in other words the LPS produced by *H. influenzae* is extremely heterogeneous (**Mannerstedt et al., 2008**). Many oligosaccharide structures have been identified as extensions from the three conserved heptoses. One important fact about these varied oligosaccharide structures is the resemblance to some human glycolipids (**Li et al., 2001**). *H. influenzae* has the ability to express epitopes containing the Gal-Gal glycoconjugate, which is elongated from Hep II and is believed to be behind the masking of the LPS in the host. This Gal-Gal glycocojungate is found in human cells, such as in human glycolipids from the blood group P^k antigen, where it acts as a receptor for bacterial ligands (**Massoud et al., 2003**;

Massoud et al.,1997). *H. influenzae* LPS seems to have the ability to mimic human glycolipids and the ability to reversibly switch off its terminal epitopes in the oligosaccharide portion (**Li et al., 2001**). It is not just the position of the carbohydrates that is variable, the location and pattern of non-carbohydrate components, such as phosphate substituents, O-acetyl groups and ester linked glycine moieties is also highly variable (**Schweda et al., 2007**).

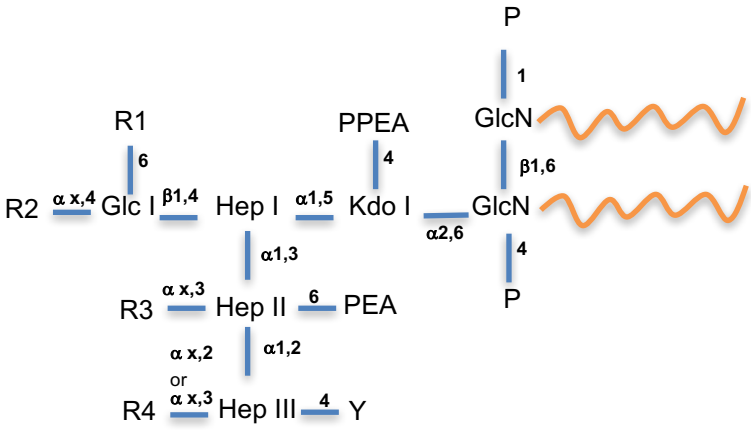


Fig. 1.3 A representation of a typical rough LPS from *H. influenzae*. The R1 position can be further substituted by hydrogen, phosphocholine, D,D-Hep or L,D-Hep; the R2 position can be further substituted by Glc, Gal or acetyl group; the R3 position can be further substituted by Glc; the R4 position can be further substituted by Glc, Gal or acetyl group and the Y position can be further substituted by an acetyl group, glycine, phosphate (P) or phosphoethanolamine (PEA). Kdo I is substituted at position 4 by pyrophosphate ethanolamine (PPEA). Image adapted from **Schweda et al., 2007**.

1.2.5 Phosphorylation

Substitution of phosphate (P), phosphocholine (PCho) or phosphoethanolamine (PEA) to either of HepII, HepIII, GlcI and KdoI inner core residues add to the heterogeneity of the LPS (**Fig. 1.3**). The addition of PCho increases the ability of the LPS to colonise and persist within the human respiratory tract. PCho mediates bacterial adherence to and invasion of hosts epithelia (**Weiser & Pan, 1998**). However, expression of the PCho also makes it easily targeted by human C-reactive protein (CRP). When CRP binds PCho it mediates killing via the activation of the complement system (**Schweda et al., 2007**). The biosynthesis of PCho, requires that a choline be acquired from the hosts lipids turnover during bacterial colonisation (**Clark et al., 2012**). The addition of the PCho

during colonisation has been shown to reduce binding of antibodies and increase survival in the presence of antibody-dependent complement mediated killing. This reduction in antibody binding is only present in LPS containing Hep III and phosphorylation of the Kdo conserved structures (**Clark et al., 2013**). In the conserved structure, the Kdo has been shown to be substituted at O-4 by pyrophosphate ethanolamine (PPEA), while Hep II is substituted at the O-6 position by phosphoethanolamine (PEA). PCho substitution has also been shown at O-6 of GlcI in Rd LPS strains, while PCho substitution at O-6 of a terminal Gal residue at Hep III has been noted in RM7004 and Eagan strains (**Schweda et al., 2007**).

1.2.6 Sialic Acids

Sialic acids are a common structural feature of *H. influenza*e LPS, that create further heterogeneity in the LPS species. A survey of *H. influenzae* LPS strains reported that one third of the typeable strains and over two thirds of the non-typeable strains contained sialic acid in their LPS. In the *Neisseria* gram-negative bacteria, sialylation has been linked with resistance to killing by complement, but reduced ability to enter mucosal cells. In *H. influenzae* LPS the sialylation is believed to help the bacteria resist the killing by normal human serum anti-bacterial components, but has no effect on attachment or invasion of cultured human cells or neutrophils (**Hood et al., 1999**). In terms of NTHi strains the presence of sialic acid in *H. influenzae* LPS has been linked with infection in a study of experimental otitis media, while in strains deficient in sialic acid the infections were reduced or absent (**Bouchet et al., 2003**). Sialylation of the LPS seems to present the bacteria with an advantage through masking with host-like epitopes that could potentially interfere with the activation of the classical or alternative pathways of the complement mediated response to pathogens (**Schweda et al., 2007**). The O-acetylation of LPS in *H. influenzae* has been shown to be linked with resistance to killing of the bacteria by normal human serum cells. Inactivation of the *oaf A* gene, which is responsible for the addition of the O-acetyl groups in the inner core of the LPS, caused an increased

killing of the bacteria when comparing with the wild type containing an active *oaf A* gene (**Fox et al., 2005**).

1.3 C-type lectins

The C-type lectin super family are Ca^{2+} dependent carbohydrate binding proteins that share primary and secondary homology in their carbohydrate recognition domains (CRDs). This super family of proteins contains a conserved C-type lectin fold, which is a structural reference for all members of the family, independent of their ability to bind carbohydrates. C-type lectin members have varied *in vivo* roles and include collectins, selectins and endocytic receptors among others (**Cummings & McEver, 1999**). In 1993 Drickamer, introduced a classification of C-type lectins with 7 subgroups (I to VII) proposed by using differences in the overall domain architecture (**Drickamer et al., 1993**). Recently, ten extra subgroups have been added (VIII to XVII), and again these were based on differences in the structure of the proteins, without any consideration of function as in many cases it remains unclear (**Zelensky & Gready, 2005**). The 17 subgroups are; I- Lecticans, II- Asialoglycoprotein and DC receptors, III- collectins, IV- selectins, V- NK-cell receptors, VI- Multi-CTLD endocytic receptors, VII- Reg group, VIII- Chondrolectin, Layilin, IX- Tetranectin, X- Polycystin, XI- Attractin, XII- Eosinophil major basic protein (EMBP), XIII- DGCR2, XIV- Thrombomodulin, XV- Bimlec, XVI- SEEC and XVII CBCP/Frem1/QBRICK.

1.3.1 C-type lectin domain (CTLD) and the C-type lectin-like domain (CTLD*)

The term C-type lectin was first introduced to distinguish a group of Ca^{2+} dependent lectin animal proteins from the other Ca^{2+} independent types of animal lectins. The binding activity of proteins containing a C-type lectin domain (CTLD) towards carbohydrates was initially reported to occur by means of a compact module, the CTLD or carbohydrate recognition domain (CRD). When new protein sequences were compared

with the CTLD, if they showed a conserved motif characteristic of the CTLD, these proteins were grouped into the CTLD subgroups. However, as the number of known sequences grew, it became clear that not all proteins containing the CTLD were able to bind carbohydrates or even Ca^{2+}, and these CTLD domains were termed C-type lectin-like domain (CTLD*). The conserved CTLD fold seen across many proteins is independent of functionality and has been reported to have occurred due to convergent evolution, where the 17 subgroups of C-type lectins share structural homology (CTLD fold) but not functional homology (**Zelensky & Gready, 2005**).

1.3.2 CTLD / CRD FOLD

The CTLD fold (universally referred to in collectins as the CRD) is well conserved and is present in all C-type lectins super family members, consisting of approximately 110-140 residues. The first CRD structure solved was that of rat Mannose Binding Protein (rMBP), using MAD phasing (**Weis et al., 1991**). The reported structure of the rMBP CRD showed antiparallel β-sheets (β1- β7), two α-helices and four loops, stabilised by four cysteine residues (C1-C4), where C1 and C4 linked β7 to α1, while C2 and C3 linked β5 to β6, forming two disulphide bridges, shown in **Fig.1.4**. Structurally the CRD can be divided into two lobes and a loop region; the lower lobe containing the β1, β2, β3, β7, α1 and α2, the upper lobe β4, β5, β6 and the long loop region (LLR). The C1-C4 disulphide bond helps the stabilisation of the main domain loop, while the C2-C3 disulphide bond stabilises the LLR region. Structurally the CRD follows the following pattern; N-terminal - β1-β2-α1-β3-α2-β4-L1-L2-L3-L4-β5-β6-β7-C-terminal. The LLR is of special interest for this experimental work, as it is involved in the Ca^{2+} dependent lectin activity of the collectins and SP-D (**Veldhuizen et al., 2011**; **Zelensky & Gready, 2003**; **Hakansson & Reid, 2000**).

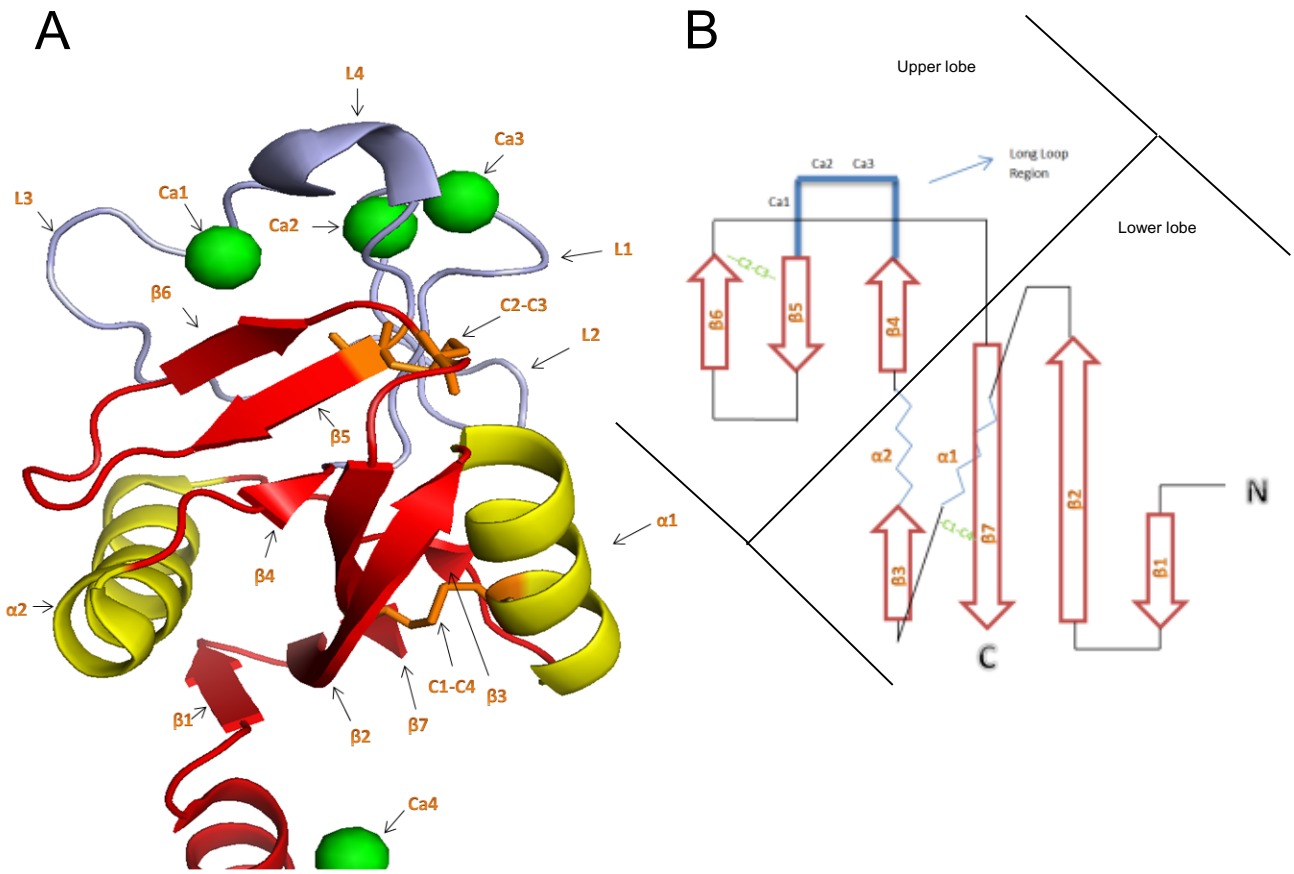


Fig. 1.4 - **A**- three-dimensional representation of the typical CTLD fold. In this case a recombinant fragment of hSP-D, showing the β-strands and α-helices, numbered β1-β7 in red and α1-α2, in yellow, with the four reported Ca^{2+} in SP-D in green. The long loop region (LLR) is shown in blue, numbered L1-L4 and finally the disulphide bridges in orange C1-C4 and C2-C3. Image generated using *PYMOL* from PBD code: 1PW9. **B**- Two-dimensional representation of typical CTLD fold, where it shows the typical pattern of N-terminus - β1-β2-α1-β3-α2-β4-LLR-β5-β6-β7-C-terminus, present in all C-type lectins, with the upper lobe and lower lobe displayed. Image adapted from **Hakansson & Reid, 2000.**

1.4 Collectins

Collectins are a family of Ca^{2+} dependent binding C-type lectins, and to date nine members have been reported, shown in **Table 1.1.** The family includes mannose binding protein (MBP), pulmonary surfactant protein A and D (SP- A and SP-D respectively), collectin kidney 1 (CL-K1), collectin liver 1 (CL-L1), collectin placenta 1 (CL-P1) and three bovine serum collectins, CL-43, CL-46 and conglutinin (**Kishore et al., 2006**; **van de Wetering et al., 2004**; **Ohtani et al., 2012**).

Collectin	Site of Expression	Polypeptide Length	Literature reported in:
Human SP-D	Lungs, Mucosal surfaces	375aa (signalling peptide 20aa). Mature protein 355aa	**Hartshorn et al., 2002 Kishore et al., 2006**
Human SP-A1 and SP-A2	Lungs, Mucosal Surfaces	248aa (signalling peptide 20aa). Mature protein 228aa.	**Hartshorn et al., 2002 Silveyra & Floros, 2012**
Human MBP	Liver	248aa (signalling peptide 20aa). Mature protein 225aa	**Weis & Drickamer, 1994**
Rat MBP-A	Liver, Kidney, Small Intestine	238aa (signalling 17aa). Mature protein 221aa	**Weis et al., 1991 Van de wetering et al., 2004**
Rat MBP-C	Liver, Kidney, Small Intestine	244aa (Signalling peptide 18aa). Mature protein 226aa	**Van de wetering et al., 2004 Ng et al., 1996 Ng et al., 1998**
Human CL-K1 or COLEC 11	Adrenal Glands, Kidneys, Liver, Small intestines, thymus, spinal cord, placenta and pancreas	271aa (signalling peptide 25aa) – Mature protein 246aa	**Hansen et al., 2010 Keshi et al., 2006 Motomura et al., 2008 Ohtani et al., 2012**
Human CL-L1 or COLEC 10	Liver	277aa (signalling peptide 27aa). Mature protein 250aa	**Ohtani et al., 1999 Axelgaard et al., 2013**
Human CL-P1	Expressed in most tissues, especially vascular endothelial cells in all murine and human heart	742aa (signalling unknown at present, so assumed 20aa)- Mature protein 722aa	**Ohtani et al., 2001**
Bovine CL-46	Thymus and Liver	371aa (signalling peptide 20aa). Mature protein 351aa	**Dec & Wernicki, 2006 Hansen et al., 2002**
Bovine CL-43	Liver	321aa (signalling peptide 20aa). Mature protein 301aa	**Hartshorn et al., 2002 Holmskov et al., 1995 Dec & Wernicki, 2006 Hansen et al., 2003**
Bovine Conglutinin	Liver	371aa (signalling peptide 20aa). Mature protein 351aa	**Holmskov et al., 1995 Dec & Wernicki, 2006 Lim & Holmskov, 1996**

Table 1.1 In total 11 collectins are shown in the above table. The site of expression and polypeptide length of each monomer (including signalling peptide) is given. Different species often produce more than one single isoform of each collectin, such as in Human SP-A and Rat MBP. Human SP-A has been reported to exist in two isoforms, SP-A1 and SP-A2, while Rat MBP has also been reported to exist as two isoforms, with MBP-A and MBP-C reported.

The name collectin derives from the words collagen and lectin (collagen-containing C-type lectins), two structural features present in all collectins with all collectins reported to date displaying both features. Collectins are known to act as PRRs against PAMPs and DAMPs and are a crucial part of the innate immune system. Collectins have the ability to recognise different pathogens through the PAMPs present on the surface of bacteria, fungi, viruses and allergens. Upon recognition the collectins have the ability to aggregate and agglutinate, neutralise, opsonise, activate macrophages and complement receptors, effect inflammatory responses and activate the adaptive immune system (**Alcorn et al., 2004**; **Veldhuizen et al., 2011**).

1.4.1 Collectins: Function

1.4.1.1 Mannose-Binding Protein (MBP)

MBP is a liver derived serum protein, consisting of two forms in rats, rMBP-A (serum) and rMBP-C (liver), while in Humans only one type of MBP has been found, often referred to as MBL (mannose binding lectin). MBP has been reported to have the ability to interact with a variety of pathogens, including gram-positive and gram-negative bacteria, yeast, parasites, mycobacteria and viruses. MBP binding preferences have been extensively reviewed in **Dommett et al., 2006**.

MBP has a variety of immunological roles, such as the activation of the lectin pathway, through the activation of MBL-associated serine proteases (MASPs) (**Thiel et al., 1997**). MBP has also been shown to hold a conflicting relationship in ischemic heart disease conditions. Studies in Humans have shown that MBP deficiency in ischemic diseases could hold beneficial effects such as reduced mortality for patients who suffered an acute myocardial infarction and long lasting protection in cerebral ischemia. Patients with unstable angina were shown to contain higher levels of MBP, with MBP transfusion to patients who suffered a myocardial infarction shown to be fatal (**Pagowska-Klimek & Cedzynski, 2014**). The above information has been disputed by evidence highlighting

that high levels of MBP in serum seemed to reduce the risk for myocardial infarction (**Saevarsdottir et al., 2005**). The complete role of MBP in cardiovascular diseases remains undefined. MBP protein deficiency has also been reported to increase the risk for infections from a neonatal age to adulthood. In neonatal age, MBP has been regarded as key for specific protection, until the development of the adaptive immune system (**Dommett et al., 2006**; **Strunk et al., 2011**). Septic events have also been linked with low serum of MBP levels (**Auriti et al., 2010**), with the suggestion that low serum MBP is a susceptible factor for gram-negative neonatal sepsis (**Schlapbach et al., 2010**).

1.4.1.2 Surfactant Protein A (SP-A)

SP-A is a surfactant protein, similar to SP-D (see 1.4.1.5), which has the dual roles of aiding in the reduction of surface tension in the lung and acting as a component of the innate immune system. SP-A has been reported to interact with a variety of pathogens, including gram-positive/negative bacteria, viruses, yeast and fungi and this has been reviewed in **Kishore et al., 2006**. SP-A is often reported to be of clinical importance in the pathogenesis of interstitial lung diseases (IDL). IDL are a group of diseases characterised by interstitial changes, such as pulmonary alveolar proteinosis (PAP), idiopathic pulmonary fibrosis (IPF), asthma and allergy, cystic fibrosis and chronic obstructive pulmonary disease (COPD). The IDLs are characterised by raised or lowered SP-A levels, suggesting that varied SP-A levels could be cause or consequence of the specific disease states (**Doyle et al., 1998**; **Vazquez de lara et al., 2000**; **McCormack et al., 1991**; **Greene et al., 2002**; **Wang et al., 2011**).

The way in which SP-A assembles into the biological unit (a trimer) has been the focus of extensive discussion, as SP-A has two isoforms SP-A1 and SP-A2 with the ratio of the two reported to vary in the trimer. An initial report into to the ratio suggested two SP-A1 and one SP-A2 molecules, in a 2:1 ratio (**Voss et al., 1991**), while recent studies into mRNA levels do not support the idea of a 2:1 ratio (**Wang et al., 2007**). Understanding the ratio of SP-A1 and SP-A2 in the trimer is of biological importance, with

reports suggesting a link between SP-A1 levels and age and disease state (**Tagaram et al., 2007**); and the fact that SP-A2 has been reported to be more efficient in enhancing phagocytosis and in stimulating TNF-α than SP-A1 (**Mikerov et al., 2004**; **Mikerov et al., 2007**; **Wang et al., 2000**). It is interesting to note that SP-A1 and SP-A2 have been shown to hold different affinities for a variety of sugars (**Table 1.6**) (**Oberley & Sneyder, 2003**).

1.4.1.3 Bovine/Cattle Collectins- CL-43, CL-46 and Conglutinin

Bovine collectins share structural and sequence homology and were initially reported to be exclusively present in cattle, with later studies suggesting that this may not be the case (**Wernicki & Urban-Chmiel, 2011**). The bovine collectins CL-43, CL-46 and conglutinin share high homology with bovine SP-D, with conglutinin and CL-43 reported to have arisen from a duplication of an ancestral SP-D-like gene (**Thiel & Takahashi, 2013**; **Lim et al., 1994**). Conglutinin was the first reported collectin, initially reported in 1902 by Elrich and Sachs (**Sliwa-Dominiak et al., 2010**); in 1993 CL-43 was reported in cattle (**Holmskov et al., 1993**) and in 2002 CL-46 was reported (**Hansen et al., 2002**). The bovine collectins differ in their site of expression and their ability to form higher order multimers, with the main site of expression for CL-43, CL-46 and conglutinin being the liver, but for CL-46 also the thymus, mammary glands and digestive system (**Hansen et al., 2002**; **Holmskov et al., 1995**).

The complete biological role of the bovine collectins remains undefined at present, as changes in serum concentrations of the collectins can occur due to season, diet and infections, making it hard to investigate the specific collectins. Bovine collectins have been suggested to be key components of the innate immune system. Conglutinin has been reported to interact with a variety of cells and pathogens, including erythrocytes, the complement component Cb3, bacterial microbes, such as *Listeria monocytogenes, Streptococcus pyogenes* and *Salmonella typhimirium*, and IAV (**Sliwa-Dominiak et al., 2010**). CL-43 has been reported to interact with IAV, gram-negative bacteria, rotavirus Nebraska calf diarrhoea virus and yeast *Cryptococcus neoformans* (**Hartshorn et al.,**

2002; **Reading et al., 1998**; **Schelenz et al., 1995**). At present, no studies showcasing the CL-46 biological role and the ability to bind to pathogens have been reported, however it has been suggested that CL-46 will hold similar properties to the reported CL-43 and conglutinin.

1.4.1.4 Novel Collectins - CL-K1, CL-L1 and CL-P1

Novel collectins were identified by screening short single read transcript sequences with C-terminal sequence homology, with three new collectins identified. CL-L1 was first reported in 1999 (**Ohtani et al., 1999**), while CL-P1 was reported in 2001 as a new scavenger receptor membrane bound type collectin (type II transmembrane protein) present in vascular rich tissues, such as heart and umbilical veins/arterial endothelial cells (**Ohtani et al., 2001**; **Ohtani et al., 2012**). CL-K1 is the latest reported collectin, present in a variety of organs tested (**Keshi et al., 2006**; **Motomura et al., 2008**).

The novel collectins have been suggested to share a biological role with classical collectins, by possessing the ability to interact and recognise glycoconjugates. CL-P1 has been reported to bind and enhance phagocytosis of gram-positive/negative bacteria and yeast. CL-P1 holds similar structural characteristics to the scavenger receptor AI (SR-AI), which is known to be expressed in macrophages and to bind to low density lipoproteins (LDL), with CL-P1suggested to play a role in innate immunity and the ability to control bacteria numbers, yeast and the amount of modified LDLs in vascular spaces (**Ohtani et al., 2001**). It is interesting to note that recent research has suggested that CL-P1 has the ability to bind antigens present in human carcinoma cells, preventing tumour metastasis (**Yoshida et al., 2003**); and to be important in Alzheimer disease (AD), with increased expression of CL-P1 in AD patients, suggesting that the CL-P1 might be involved in the clearance of amyloid molecules (**Nakamura et al., 2006**). CL-K1 has been reported to have similar anti-microbial properties to CL-P1, by binding to both gram-positive/negative bacteria and yeast, including LPS from *Klebsiella pneumonia*, lipoteichoic acid (LTA) of *Staphylococcus aureus* and *Streptococcus pyogenes/sanguis* and mannan from

Saccharomyces cerevisiae (**Keshi et al., 2006**). The biological role and binding ability of CL-L1 remains to be investigated, however it is expected to hold similar properties to the novel collectins and the whole collectin family, with the ability to recognise a variety of microbial ligands and act as a component of the innate immune system (**Axelgaard et al., 2013**; **Ohtani et al., 1999**).

It is worth mentioning that a CL-L1:CL-K1 heterotrimer complex has been reported, containing one chain of CL-L1 and two polypeptide chains of CL-K1 chains. This complex has been shown to interact with mannan-binding lectin (MBL)-associated serine proteases (MASPs) 1 and 2, with the ability to activate the complement system through C4b (**Henriksen et al., 2013**). The significance of this complex and if other mixed complexes can occur, remains to be investigated.

1.4.1.5 Surfactant protein D

Surfactant protein D (SP-D) is a member of the collectin subfamily of C-type lectins and is involved in immunologic functions as part of the innate immune system. SP-D is a pattern recognition receptor (PRR) and has the ability to interact with carbohydrates structures present on the surface of a wide variety of microorganisms targets, such as viruses, bacteria, yeasts and fungi (**Singh et al., 2003**).

SP-D has been reported to have the ability to interact and increase the clearance of pathogens, by mediating different immunological events and interacting with different components of host immune system (**McCormack & Whitsett, 2002**). The way that SP-D interacts and recognises pathogens is dependent on the repetitive PAMPs and DAMPs displayed on the surface of the microbes or particles, with SP-D holding varying affinities for different carbohydrates. The binding of SP-D onto repetitive arrays of carbohydrates found on the surfaces of bacteria, viruses, fungi, allergens and nucleic acids, can lead to agglutination, enhanced phagocytosis and an oxidative response, with the aim of

enhanced killing and clearance by macrophages, neutrophils or the adaptive immune system.

SP-D was initially reported in 1988 in freshly prepared rat alveolar type II cells. These type II cells were shown to synthesise various collagenous proteins and one of these proteins was initially termed CP4 (SP-D), the second reported soluble surfactant protein, after SP-A (**Persson et al., 1988**). A second site of production in the lung was reported for SP-D in 1992 in non-ciliated bronchial epithelial cells (or clara cells) (**Crouch et al., 1992**). SP-D mRNA and protein have also been detected in other non-pulmonary surfaces such as pancreatic ducts, intrahepatic biliary ducts and gall bladder in the gastrointestinal tract, the sweat glands and epithelial cells of the skin, parotid and lachrymal glands (**McCormack & Whitsett, 2002**).

1.4.1.5.1 Agglutination

The binding of microorganisms by SP-D can lead to direct agglutination or neutralisation of the microorganisms, with various reported cases where SP-D agglutinates foreign pathogens; SP-D has been reported to have the ability to agglutinate Group B streptococcus (GBS) a gram-positive bacterium and *H. influenzae* gram-negative bacterium. This agglutination was calcium dependent (**LeVine et al., 2000**). Influenza A virus (IAV) has also been shown to be agglutinated by SP-D (**LeVine et al., 2001**).

The ability of SP-D in instigating agglutination of the intact microbes or particles is due to its multimeric structure and the ability to recognise a wide variety of carbohydrates present on the surface of the various microbes. The aggregation of the intact microbes has been suggested to aid in the mucociliary removal in the respiratory tract, while preventing adhesion of pathogens to cell surfaces, microbial colonisation and increased phagocytosis (**Van de Wetering et al., 2004**). It is interesting to note that agglutination can also lead to reduced phagocytosis. *Candida albicans* is a diploid fungus that was shown to be recognised and agglutinated by SP-D, causing a reduction of fungal growth.

This study also reported that SP-D inhibited the phagocytosis of *Candida albicans* by alveolar macrophages, suggesting that agglutination and limitation of growth is sufficient to impair the *Candida albicans* fungus (**van Rozendaal et al., 2000**). A second study also reported that aggregation of *Pneumocystic carini* by SP-D impaired phagocytosis by alveolar macrophages, with the authors suggesting that SP-D mediated aggregation could aid the organisms in avoiding elimination (**Yong et al., 2003**).

1.4.1.5.2 Opsonisation

SP-D has also been reported to have the ability to act as an opsonin, where it decorates the surface of pathogens and necrotic cells and presents them directly to alveolar macrophages. SP-D has been reported to bind nucleic acids from disrupted cells, both liberated and cell-surface DNA. SP-D was able to bind free DNA and DNA present on the surface of apoptotic cells, by forming large complexes with DNA and with the CRD being suggested as the main region where the interaction takes place. SP-D (-/-) mice showed a high concentration of alveolar macrophages containing degraded DNA. The intranasal delivery of a recombinant fragment of SP-D including the CRD resulted in the clearance of apoptotic macrophages from the lungs, suggesting that the CRD could be the domain necessary to recognise and bind to nucleic acids. SP-D acts as an opsonin when bound to the DNA of apoptotic cells, leading to aggregation and phagocytosis (**Palaniyar et al., 2003**; **Palaniyar et al., 2004**).

1.4.1.5.3 Phagocytosis

Even though phagocytosis was showed to be reduced upon agglutination of *Candida albicans* by SP-D (**van Rozendaal et al., 2000**), various other studies have shown the ability of SP-D to promote phagocytosis of apoptotic cells and foreign pathogens. SP-D has been described to increase the calcium-dependent neutrophil uptake of *Escherichia coli, Streptococcus pneumoniae* and *Staphylococcus aureus*. The increased uptake was a consequence of mechanisms involving bacterial aggregation and the direct action of SP-D on neutrophils (**Hartshorn et al., 1998**). SP-D has also been

reported to enhance the uptake of influenza A virus by neutrophils, which was dependent on the viral aggregation and multerisation of the SP-D (**Hartshorn et al., 1997**). The ability to enhance phagocytosis seems to be dependent on the aggregation of the bacteria and viruses (**Hartshorn et al., 1998**), suggesting that SP-D has the ability to alter the uptake of foreign pathogens by phagocytes and act as a bridge between neutrophils and pathogens, with the aim of promoting clearance.

It is interesting to note that a previously reported study demonstrated that SP-D was an insufficient agonist and did not alter the uptake of gram-negative bacteria by rat alveolar macrophages (**Pikaar et al., 1995**). The contradictory results seen in different studies have been suggested to be the consequence of different interactions between SP-D with macrophages or neutrophils, and the differences in the SP-D multimerisation used in the different studies (**Hartshorn et al., 1998**).

1.4.1.5.4 Direct Microbial Activity

SP-D (and SP-A) has been reported to inhibit the proliferation of gram-negative bacteria. SP-D was shown to inhibit the growth *of E. coli, Klebseilla pneumoniae* and *Enterobacter aerogenes*, independent of host defence macrophage action, opsonisation and agglutination. This inhibition has been reported to impact on the bacteria cell membrane, with an increase of microbial membrane permeability reported. The authors suggested that this novel mechanism plays a role in macrophage-independent control of gram-negative bacteria in the airspace (**Wu et al., 2003**). This direct microbial activity has also been reported in *Candida albicans*, where SP-D inhibited phagocytosis of the fungi, but inhibited the fungal growth (**van Rozendaal et al., 2000**).

1.4.1.5.5 Bacteria and mycobacteria

SP-D has been reported to be able to interact with both gram-positive and gram-negative bacteria, and mycobacteria, via its CRD in a calcium dependent manner (**Haagsman et al., 2008**), see **Table 1.2.** The interaction of SP-D with gram-

positive/negative bacterial cell walls occurs by means of interaction with the specific antigens presented. For gram-positive bacteria, SP-D binds peptidoglycan and lipoteichoic acid in a calcium dependent manner via the CRD (**van de Wetering et al., 2001**). In the case of gram-negative bacteria, SP-D binds to the core oligosaccharides or the polysaccharides in the O-antigen, both characteristic regions of LPS, with no reported interactions with lipid A; it is interesting to note that SP-A has been reported to bind to lipid A (**Lim et al., 1994**; **Alcorn et al., 2004**).

Gram Type	Species	Literature reported in:
Gram Negative	*Escherichia coli*	**Lim et al., 1994** **Hartshorn et al., 1998** **Wu et al., 2003**
	Haemophilus influenzae	**Restrepo et al., 1999**
	Klebsiella pneumoniae	**Lim et al., 1994** **Sahly et al., 2002**
	Pseudomonas aeruginosa	**Lim et al., 1994** **Kishore et al., 1996**
	Helicobacter pylori	**Moran et al., 2005**
	Chlamydia trachomatis	**Oberley et al., 2004** **Oberley et al., 2004a**
	Chlamydia pneumoniae	**Oberley et al., 2004a**
Gram Positive	*Staphylococcus aureus*	**Hartshorn et al., 1998**
	Streptococcus pneumoniae	**Hartshorn et al., 1998**
	Bacillus subtilis	**van de Wetering et al., 2001**
Mycobacteria	*Mycobacterium tuberculosis*	**Ferguson et al., 1999**
	Mycobacterium avium	**Kudo et al., 2004**

Table 1.2 The reported gram-negative, gram-positive and mycobacteria species that SP-D has been reported to interact with.

The ability of SP-D to interact with smooth and rough LPS has been reported to be different. SP-D preferentially binds to rough LPS, by recognising the core oligosaccharide domain, which is rich in glucose and heptose residues, with an enhanced cellular response towards rough LPS reported (**Wang et al., 2008**; **Yamazoe et al., 2008**; **Kuan et al., 1992**). In terms of smooth LPS, there is contradictory information in the literature regarding the ability of SP-D to directly bind to smooth LPS. The ability to bind smooth

LPS has been suggested to be dependent on the extent and variation of the O-antigen. Following a challenge by smooth or rough LPS, SP-D has the potential to promote a pro-inflammatory cell response, with the aim of clearing the invading bacterium (**Bufler et al., 2003**; **Yamazoe et al., 2008**; **Sahly et al., 2002**). One study demonstrated the ability of SP-D to interact with rough LPS mutants; SP-D has been shown to have the ability to bind rough LPS mutants such as Rc and Rd mutants, but not the Re (shortest rough mutant) or Ra (longest rough mutant). This implies that LPS core sugars such as glucose and heptose are recognised by SP-D. Affinity binding studies reported that SP-D has a preference for heptosyl and glucosyl saccharides, with the affinity towards glucosyl disrupted by the addition of heptoses, suggesting a higher affinity for heptosyl saccharides (**Kuan et al., 1992**). Smooth LPS and rough Ra mutant LPS were reported to interact weakly with SP-D, with limited agglutination by SP-D towards the LPS. The poor interaction between SP-D and smooth and Ra LPS was suggested to be due to the longer chains in both types of LPS, where the additional outer core oligosaccharides (for Ra mutant + smooth LPS) and the additional polysaccharides (for smooth LPS), mask the binding of the inner sites associated with the key oligosaccharides (**Kuan et al., 1992**).

1.4.1.5.6 Viruses

Viruses differ in their infectivity from other pathogens, as they have the requirement of needing entry into cells for subsequent replication and survival, leading to devastating effects in host cells. SP-D has also been reported to be important in the recognition of viruses (**Table 1.3)**. Respiratory syncytial virus (RSV) is a common pathology in children with severe morbidity. SP-D has the ability to bind to the glycosylated G protein of the RSV, through the CRD, with the loss of viral infectivity (**Hickling et al., 1999**). A case control study reported significant differences between infants with RSV and controls, in terms of the allele for the SP-D residue located at position 11. In this position two alternative residues were noted, methionine (Met) and threonine (Thr), with the Met residue being hydrophobic and the Thr polar. Position 11 in

SP-D is situated near the Cys15 and Cys20, crucial in inter-chain disulphide binding and the formation of higher order multimers, with enhanced antiviral properties. The authors suggested that the prevalence of Met at position 11 in children infected with RSV could impact on the formation of disulphide bonds thus halting the formation of higher order multimers (**Lahti et al., 2002**). IAV is normally highly glycosylated for both the neuraminidase (NA) and hemagglutinin (HA) protein components of the virus, which SP-D can recognise and bind to (**Tate et al., 2014**).

Virus	Notes	Literature reported in:
Influenza A Virus (IAV)	N/A	**Hartshorn et al., 1994** **LeVine et al., 2001**
Respiratory syncytial virus (RSV)	Binds to RSV G Protein.	**Hickling et al., 1999**
Human immunodeficiency virus (HIV)	Binds to envelope protein gp120, inhibiting HIV replication	**Meschi et al., 2005**

Table 1.3 Virus species that SP-D has been reported to interact with.

1.4.1.5.7 Fungi and Parasites

SP-D has been reported to interact and recognise a variety of yeast and fungus (**Table 1.4**). *Pneumocystis carinii,* an opportunistic pathogen with the potential of creating pathogenesis in immunosuppressed patients, has been reported to be a ligand for SP-D, leading to enhanced binding of *Pneumocystis carinii* to macrophages (**O'Riordan et al., 1995**). *Candida albicans* fungus, a commensal and normal constituent of gut flora, which can impact on the health of immunosuppressed patients, has been reported to be recognised by SP-D leading to increased agglutination of the microorganisms and inhibition of their growth (**van Rozendaal et al., 2000**). *Schistosoma mansoni* is a multicellular parasite with significant infection in humans. SP-D was reported to bind to the *Schistosoma mansoni* by recognising fucose repeats on the surface of the organism (**van de Wetering et al., 2004a**).

Fungi and Parasite	Notes:	Reference:
Pneumocystis carinii	N/A	**O'Riordan et al., 1995 Yong et al., 2003**
Candida albicans	N/A	**van Rozendaal et al., 2000**
Aspergillus fumigatus	Can cause an allergic reaction	**Madan et al., 1997**
Schistosoma mansoni	First study showing that SP-D can interact with multicellular lung pathogens.	**van de Wetering et al., 2004a**

Table 1.4 The fungi and parasite species reported to interact with SP-D.

1.4.1.5.8 Allergens

Allergens have the ability to induce an allergic reaction and airway inflammation, causing the activation of various different components of the immune system such as dendritic cells, T helper lymphocytes, IgE – producing B lymphocytes (plasma cells), mast cells and eosinophils.

In the lung, allergens that can induce inflammation are initially dissolved in the airway lining fluid before any contact with immune cells. This airway lining is packed with surfactant components, including SP-D and SP-A. SP-D has been shown to bind to whole mite extracts (*Dermatophagoides pteronyssinus*) and the purified allergen, leading to the inhibition of IgE allergen specific, suggesting a role in modulating allergen sensitisation and allergic reactions (**Wang et al., 1996**). SP-D has also been reported to bind to *Aspergillus fumigatus,* an allergy inducer in humans, enhancing agglutination and binding of the microbe to alveolar macrophages and neutrophils, with increased killing of *Aspergillus fumigatus* by phagocytosis and oxidative burst (**Madan et al., 1997**).

Upon recognition of allergens and other pathogens, SP-D has the ability to activate or inhibit the adaptive immune system. SP-D has been reported to interact with dendritic cells and T-lymphocytes, by presenting *E. coli* to the adaptive immune cells, leading to the increased association of dendritic cells around sites of presentation (**Brinker et al., 2001**). SP-D has also been reported to inhibit the proliferation of T-lymphocytes, IL-2 and histamine release, with the aim of protecting the lung against collateral immune-mediated damage (**Borron et al., 1998**; **Wang et al., 1998**).

1.4.1.5.9 Necrotic/Apoptotic Cells

SP-D has been shown to be important in clearing apoptotic cells, with SP-D (-/-) mice reported to contain more apoptotic and necrotic cells, with the alveolar macrophages reported to contain accumulated and degraded DNA on the surface. Using labelling experiments, the authors reported that SP-D bound to normal alveolar macrophages, but showed preferential interaction with phagocytes that contained degraded DNA (**Palaniyar et al., 2003**).

1.4.1.5.10 SP-D and the immune system

SP-D has been shown to interact with various immune system cells, such as macrophages and neutrophils, by means of membrane bound receptors (**Miyamura et al., 1994**; **Harshorn et al., 1998**) and with soluble receptors such as gp-340. Glycoprotein 340 (gp-340) was the first reported receptor for SP-D and was identified in BALF (Bronchoalveolar lavage fluid) of patients with PAP (**Holmskov et al., 1997**). Gp-340 is a soluble opsonin with the ability to stimulate alveolar macrophages (**Tino & Wright, 1999**).

Cluster of differentiation 14 (CD-14) is an innate immune protein that can exist in membrane bound form (mCD-14) on monocytes, macrophages and neutrophils and as a soluble form (sCD-14) released/exocytosed by these cells (**Wall et al., 2009**). CD-14 has the ability to directly bind to smooth and rough LPS. SP-D has been reported to bind to CD-14 in a calcium dependent manner, with EDTA (calcium chelating) and mannose inhibiting the binding to CD-14. SP-D has been reported to be able to reduce the binding of CD-14 to smooth and rough LPS, suggesting the ability to alter the interactions of CD-14 and LPS (**Sano et al., 2000**). The ability of SP-D to impact on CD-14 was shown in transgenic mice studies, where SP-D (-/-) knockout mice resulted in reduced CD-14 presentation on alveolar macrophages, causing a reduced uptake of LPS and decreased production of TNF- α after LPS stimulation (**Senft et al., 2005**).

SP-D has been reported to interact with calreticulin and CD-91 through the

collagen region. This interaction generated a pro-inflammatory signalling pathway and clearance of PAMPs by phagocytosis (**Gardai et al., 2003**). When not bound to a PAMP or DAMP, SP-D is reported to bind to Signal regulatory protein α SIRPα through the CRD region, initiating anti-inflammatory cytokine production (**Gardai et al., 2003).**

Leukocyte-associated immunoglobulin-like receptor 1 (LAIR-1), an inhibitory immune Ig-superfamily receptor, has been shown to bind to SP-D through the collagen region. The binding resulted in inhibition of reactive oxygen species production through LAIR-1, by inhibiting the FcαR-induced oxidative burst (**Nordkamp et al., 2014**).

SP-D has been reported to bind to the immunoreceptor OSCAR. OSCAR has been reported to be present on monocytes, macrophages, neutrophils and dendritic cells and has the ability to enhance the pro-inflammatory response in monocytes and promote activation of inflammatory cell signalling proteins such as TNF-α. OSCAR was reported to bind to SP-D through the long collagen region attached to the CRD, with no binding seen when using a shorter recombinant fragment of human SP-D **(Barrow et al., 2015**).

The ability of SP-D in enhancing pro or anti-inflammatory signals has been suggested to be dependent on the presence or absence of pathogens and the interaction with specific receptors. This suggestion seems to be supported by growing evidence in the literature; dodecamers of CRDs occur as the structural component of SP-D in the normal non-inflammatory lung, ensuring a reduced inflammatory signalling as access to the collagen regions is more restricted for pro-inflammatory molecules such as calreticulin, CD91 and OSCAR, and where the CRD is available to interact with SIRPα (anti-inflammatory). When SP-D recognises PAMPs, DAMPs or CD14, the multimers have been suggested to dissociate into trimers, with NO suggested as the dissociating factor, leading to S-glycosylation of the N-terminal, enhancing the macrophages chemotaxis and leaving the collagen region to interact with pro-inflammatory molecules. Once the infection has been destroyed, LAIR-1 is suggested to reduce the NO burst, avoiding secondary

damage and uncontrolled inflammation (**Guo et al., 2008**; **Leith-Larsen et al., 1999**; **Atochina-Vasserman, 2011**).

1.4.1.5.11 SP-D - Transgenic Studies and Pathogenesis

Lung collectins have the ability to recognise a variety of pathogens and self-cells by binding to various motifs, leading to agglutination/aggregation, opsonisation, phagocytosis and the regulated release of inflammatory mediators. Knockout SP-D gene studies or transgenic studies of SP-D (-/-) mice have allowed for the understanding of the physiological role of SP-D in the normal lung and the impact of reduced SP-D levels in disease states. The loss of SP-D in mice has resulted in increased inflammatory response and impaired clearance of bacteria, virus, fungal pathogens and apoptotic cell clearance, suggesting a crucial role of SP-D in the maintenance of a healthy lung.

Various studies into the role of SP-D in surfactant homeostasis have been performed, where SP-D (-/-) mice demonstrated abnormal type II cells, with increased pool sizes of alveolar macrophages and tissue phosphatidylcholine (PC), in comparison to the controls. The SP-D (-/-) mouse studies also reported increased numbers of large foamy alveolar macrophages, enlarged alveoli, abnormal morphology increased levels of surfactant lipids, and decreased tubular myelin, SP-A mRNA and SP-A (**Korfhagen et al., 1998**; **Botas et al.,1998**). The levels of SP-D in the host have been suggested to be the crucial element, with a study reporting that mutant mice containing a heterozygous SP-D gene, with concentrations of 50% of normal wild-type, reported no signs of phenotype abnormality (**Botas et al., 1998**). The above studies suggest that SP-D is important in surfactant homeostasis, impacting on the respiratory process and surface tension reduction, if the levels of SP-D are greatly reduced.

Transgenic studies of SP-D (-/-) have also been important in showcasing the important role that SP-D holds in pulmonary health, its contribution towards the normal physiology and in the clearance of microbes. SP-D (-/-) mice have been suggested to

have impaired clearance of microbes and necrotic own cells. SP-D (-/-) mice infected with group B Streptococcus and *H. influenzae* through intratracheal instillation, showed increased levels of oxidant production (NO and hydrogen peroxide), pulmonary inflammation and reduced uptake of both organisms by alveolar macrophages. There was however no difference in the killing of both organisms. The authors suggested that the lack of SP-D could impact on the opsonisation and phagocytosis ability towards the microorganisms; however, the increased oxidant production contributed towards the killing of both organisms (**LeVine et al., 2000**). SP-D (-/-) mice infected with two strains of IAV, showed reduced clearance of one of the strains, while towards the second strain there was no difference from the controls. The infection of both strains of IAV in SP-D (-/-) mice, showed similar reactions in the mice, with an increased production of pro-inflammatory cytokines reported (**LeVine et al., 2001**). SP-D (-/-) mice have also been shown to have reduced clearance of cells with degraded DNA. In the SP-D (-/-) mice the alveolar macrophages showed accumulated DNA, suggesting that SP-D has a role in the clearance of apoptotic and necrotic own cells (**Palaniyar et al., 2003**). One interesting study using overexpressed SP-D (+/+) mice, reported increased clearance of *E. coli* K-12 bacteria (**Wu et al., 2003**), suggesting that overexpressed SP-D could be beneficial in the clearance of microbes.

As mentioned above, SP-D (-/-) mice have been reported to acquire chronic inflammation, emphysema (due to activated macrophages), fibrosis and increased oxidative response. SP-D (-/-) mice have been reported to show signs of Pulmonary Alveolar Proteinosis (PAP) at three weeks old, with the advancing of the age of the mice further progressing the disease state. PAP is a rare lung disease that leads to the accumulation of surfactants and larger and foamy alveolar macrophages within the alveoli, causing distal air space dilation that interferes with gas exchange (**Wert et al., 2000**).

1.4.1.5.12 Disease - Cause or Consequence of SP-D levels?

SP-D levels have been investigated in a variety of lung diseases. These studies report variations in the level of SP-D, with all clinical conditions studied showing reduced levels of SP-D, with the suggestion that SP-D level variation is either the cause or consequence of specific clinical conditions. The reduction of SP-D levels to above 50%, is suggested to lead to lung morphology changes, with the reduction of the levels to below 50% potentially halting pulmonary abnormalities (**Wert et al., 2000**). Reduced levels of SP-D have been reported in cystic fibrosis (CF), Chronic Obstructive Pulmonary Disease (COPD), smokers, Idiopathic pulmonary Fibrosis (IDF), Interstitial Pneumonia Associated with collagen vascular disease (IPCD), PAP and RSV virus infected patients BALF (**Postle et al., 1999**; **Honda et al., 1996**; **Nishikiori et al., 2014**; **Kerr & Paton et al., 1999**). It remains to be investigated if the reduced levels of SP-D are either a cause or a consequence of the diverse pathologies.

1.4.2 Collectin structure and sequence

The basic structural unit for all collectins is the trimer (**Fig. 1.5**), with most collectins sharing similar structural features, including a cysteine rich N-terminus, which allows for higher order multimers, by means of the formation of disulphide bonds; a collagen rich region domain of varying length and important in interacting with other immune receptors; an α-helical neck region, crucial in the trimerisation and orientation of the CRDs; and finally the CRD involved in the Ca^{2+} dependent binding to PAMPs and DAMPs. Some collectins have the ability to multimerise into larger structures, composed of various trimers attached together. Although collectins share similar structural features, differences in tissue expression, the number of cysteines in the N-terminal region, the length of the collagen region, the ability to multimerise and the preference of the CRD for specific glycoconjungates have been reported (**Hartshorn et al., 2002**).

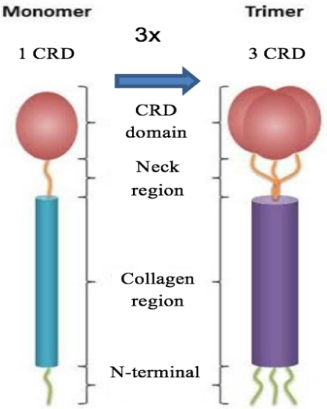


Fig. 1.5 - A schematic representation of the collectins structural unit, the trimer. Each trimer is composed of three polypeptide monomers grouped together (trimerised). Adapted from **Kishore et al., 2006.**

1.4.2.1 Collectin trimerisation and multimerisation

Collectins differ in the ability to multimerise, with impact on the binding affinities and recognition preferences for specific glycoconjugates. As a monomer, collectins have a low affinity for the carbohydrate targets, with the trimerisation into a basic trimer, with multiple recognition sites, yielding a stronger binding avidity to target molecules. Higher order multimers have been reported to have a higher affinity than lower multimers (**Holmskov et al., 2003**). The collectins ability to multimerise has been reported to vary and this is displayed in **Fig. 1.6**.

- **SP-D** has been reported to produce most often tetrameric cruciform-like structures (**Fig. 1.6B**) (**Shrive et al., 2003**) where 4 trimers come together, with further oligomerization, possible where 8 cruciform like structures combine, yielding over 96 individual CRDs (12 CRDs x 8 cruciform like structures), denominated “fuzzy balls”. Trimers have also been reported (**Fig. 1.6D**) (**Kishore et al., 2006**; **Crouch, 2000**).
- **SP-A** has been shown to form a hexamer of trimers, in which 6 trimers associate together to yield a molecule resembling a bouquet of flowers (**Fig. 1.6C**) (**Head et al., 2003**; **Kishore et al., 2006**; **Shrive et al., 2003**).
- **MBP** is reported to be very similar to SP-A, producing a hexamer of trimers, resembling a bouquet of flowers (**Sheriff et al., 1994**; **Weis et al., 1991**).
- **CL-43, CL-K1** and **CL-L1** have been reported to produce only trimers (**Fig. 1.6A**). CL-K1 has been reported to produce higher order oligomers in complex formation with other plasma protein such as MASP-1/3. (**Holmskov et al., 1993**; **Hansen et al., 2010**; **Ohtani et al., 1999**; **Axelgaard et al., 2013**).

- **CL-46** and **Conglutinin** have been reported to share similar structure to SP-D, by forming tetramers of four trimers (**Hansen et al., 2002**; **Hansen et al., 2002a**).
- **CL-P1** produces membrane bound trimers (**Ohtani et al., 2001**; **Ohtani et al., 2012**).

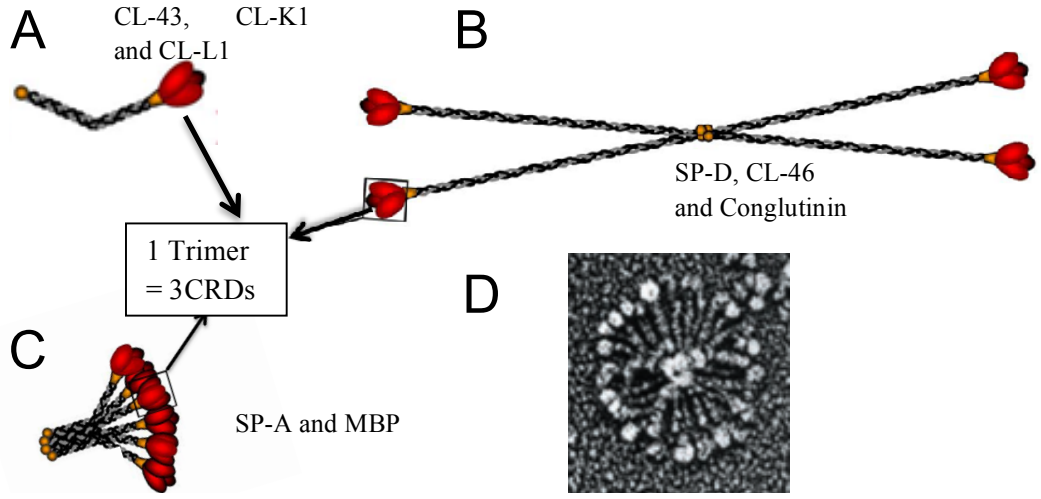


Fig. 1.6 A- The CL-43, CL-K1 and CL-L1 secreted collectins, with no higher multimers reported. CL-P1 is also a trimer, however it is membrane bound. **B-** CL-46, Conglutinin and SP-D produce a cruciform like structure, composed of a tetramer of trimers. **C-** SP-A and MBP, resembling a bouquet of flowers, similar to C1q. **D-** "fuzzy balls" produced by 8 SP-D cruciform like structures coming together. **A, B, C** images adapted from **Kishore et al., 2006**. **D** image taken from **Crouch et al., 2000**.

1.4.2.2 Sequence and structure

The number of residues in each of the structural regions of selected collectins is shown in **Table 1.5** and a sequence analysis of selected collectins is given in **Fig 1.7**. All sequences used in the production of the sequence analysis were sourced from the National Centre of Biotechnology Information (NCBI - http://www.ncbi.nlm.nih.gov/) with the following collectins from *Homo sapiens, Rattus norvegicus* and *Bos taurus* included: hSP-D (AAB59450.1), rSP-D (AAH70507.1) , bSP-D (CAD69922.1), hSP-A1 (NP_001158119.1), hSP-A2 (AAI57891.1), bSP-A (AAI23547.1), rSP-A (2202163A), rMBP-A (NP_036731.2), rMBP-C (NP_073195.2), bConglutinin (NP_783630.2), hMBP (CAA34079.1), bCL-46 (NP_001001856.1), bCL-43 (AAI16148.), hCL-K1 (BAF43301.1), hCL-L1(NP_006429.2) and hCL-P1 (BAB72147.1).

Collectin	Number of Cysteines in N-terminal region	Number of Triplets in the collagen region	Number of residues in the Neck region	Length of CRD (number of residues)	Literature reported in:
Human SP-D	2	59 Gly-X-Y	33	119	**Hakansson & Reid, 2000** **Zhang et al., 2001** **Hansen & Holmskov, 2002a** **Weis & Drickamer, 1994** **Shrive et al., 2003** **Kishore et al., 2006** **Allen et al., 2001**
Human SP-A1 and SP-A2	1-2	23 Gly-X-Y	34	123	**Elhalwagi et al., 1997** **Hakansson & Reid, 2000** **Uemura et al., 2006** **Kishore et al., 2006**
Human MBP	3	19 Gly-X-Y	34	113	**Hakansson & Reid, 2000** **Sheriff et al., 1994**
Rat MBP-A	3	17 Gly-X-Y	34	114	**Wallis & Drickamer, 1999** **Hakansson & Reid, 2000** **Weis & Drickamer, 1994** **Ng et al., 1996**
Rat MBP-C	2	19 Gly-X-Y	36	114	**Van de wetering et al., 2004** **Ng et al., 1996** **Ng et al., 1998** **Hakansson & Reid, 2000**
Human CL-K1 or COLEC 11	1	24 Gly-X-Y	29 or 34 (exact number unknown)	122	**Ohtani et al., 2012** **Keshi et al., 2006** **Selman & Hansen, 2012**
Human CL-L1 or COLEC 10	1	24 Gly-X-Y	37	122	**Ohtani et al., 1999** **Hakansson & Reid, 2000** **Axelgaard et al., 2013** **Selman & Hansen, 2012**
Human CL-P1	2	49 Gly-X-Y	17	153	**Ohtani et al., 2001**
Bovine CL-46	2	57 Gly-X-Y	28	127	**Holmskov et al., 1995** **Hansen & Holmskov, 2002a** **Dec & Wernicki, 2006** **Lim & Holmskov, 1996**
Bovine CL-43	2	38 Gly-X-Y	31	128	**Holmskov et al., 1995** **Hakansson & Reid, 2000** **Hansen & Holmskov, 2002a** **Dec & Wernicki, 2006** **Hansen et al., 2003**
Bovine Conglutinin	2	57 Gly-X-Y	28	127	**Sliwa-Dominak et al., 2010** **Hakansson & Reid, 2000** **Hansen & Holmskov, 2002a** **Dec & Wernicki, 2006** **Lim & Holmskov, 1996**

Table. 1.5 The data presented above is the culmination of all the current knowledge reported in literature, in regards to the composition of the N-terminal, Collagen Region, Neck Region and CRD for each of the collectin monomers reported.

A total of 18 conserved residues have been previously reported to exist in the CRDs of 22 different extracellular collectins (hSP-D numbering); Lys243 (K), Cys261, 331, 345 and 353 (C), Pro271 and 322 (P), Asn 277, 316 and 341 (N), Tyr314 (Y), Trp317 and 340 (W), Glu321 and 329 (E), Gly338 (G), Asp342 (D) and Phe355 (F) (**Hansen et al., 2002**; **Drickamer,1993**). A sequence alignment produced for the purpose of this thesis

used 15 (out of the 16 - 1 membrane bound collectin) extracellular secreted collectins displaying 19 conserved residues (**Fig. 1.7**). This differs from the reported 18 conserved residues, perhaps due to sequence updates or the discovery of new collectins (**Hansen et al., 2002**; **Drickamer, 1993**). The results presented here differ from the reported sequence analysis, in part due to the authors use of the Human SP-A last three residues as CDF, while in the above sequence, the final three residues for either SP-A1 or SP-A2 are CEF, which is highly conserved in all the extracellular secreted collectins investigated. The extra conserved residue reported here is Glu354 (E).

The only extracellular secreted collectin that does not have the 19 conserved residues is CL-K1, which shows variation and only 18 conserved residues, with a substitution of Tyr314 (Y) to Phe314 (F) occurring. This substitution does not seem to affect its structural features, as Phe contains an ortho hydrogen and Tyr a hydroxyl group on the aromatic ring. hCL-P1, the only non-extracellular secreted collectin investigated, contains 14 out of the 19 conserved residues; this could be a consequence of CL-P1 being a membrane bound collectin (type II transmembrane) (**Ohtani et al., 2001**).

Collectins can be divided into classical collectins and novel collectins. Novel collectins which include hCL-K1, hCL-L1 and hCL-P1 have been reported to have evolved separately from the classical collectins, with a weak lectin activity reported, in comparison with the classical collectins (**Keshi et al., 2006**; **Ohtani et al., 2012**). Classical collectins, due to similar evolutionary selection, contain residues that are conserved solely in the classical collectins; Gly241, Phe245 Arg272, Glu276, Ala279, Phe304 and Gly309. Phe304 is part of the hydrophobic cluster important in holding the CRD region between β4, β5, β6 (**Fig.1.5**) (**Hakansson & Reid, 2000**). In humans, two SP-A isoforms have been reported, SP-A1 and SP-A2, consisting of up to 10 amino acid differences in both sequences, two in the signal sequence, six in the collagen region and two in the neck and CRD region (**Garcia-Verdugo et al., 2002**).


```
hSP-D   181 PGLKGDKGIP GDKGAKGESG LPDVASLRQQ V-EALQGQVQH LQAAFSQYK-K VELFPNGQSV 240
rSP-D   182 PGLKGDRGAP GDRGIKGESG LPDSAALRQQ M-EALNGKLQR LEAAFSRYK-K AALFPDGQSV 241
bSP-D   175 PGLKGDRGTP GERGAKGESG LAEVNALRQR V-GILEGQLQR LQNAFSQYK-K AMLFPNGRSV 234
hSP-A1   54 -GLPGAPGIP GECGEKGEPG ERGPPGLPAH LDEELQATLHD FRHQILQTRGA LSLQGSIMTV 114
hSP-A2   54 -GLPGAPGVP GERGEKGEAG ERGPPGLPAH LDEELQATLHD FRHQILQTRGA LSLQGSIMTV 114
bSP-A    59 PGL------P GERGEKGEPG ERGPPGFPAY LDEELQGTLHE IRHQVLQSQGV LRLQGSVLAV 114
rSP-A    56 PGAPGA---P GERGDKGEPG ERGLPGFPAY LDEELQTELYE IKHQILQTMGV LSLQGSMLSV 114
rMBP-A   54 -------GAP GSQGPKGQKG DRGDSRAIEV KLANMEAEINT LKSKLELTN-K LHAFSMGKKS 107
rMBP-C   66 ------KGAT GPKGDRGESV EFDTTNIDLE IA-ALRSELRA M--------RK WVLLSMSENV 112
hMBP     74 ------KGDP G-KSPDGDSS L--AASERK- ----------A LQTEMARIK-K WLTFSLGKQV 114
bCL-46  175 PGLKGDRGDP GERGAKGESG LADVNALKQR VTI-LEGQLQR LQNAFSRYK-K AVLFPDGQAV 234
bCL-43  124 PGLKGDRGDP GEKGARGETS VLEVDTLRQR MRN-LEGEVQR LQNIVTQYR-K AVLFPDGQAV 183
bCongl  175 PGLKGDRGDP GETGAKGESG LAEVNALKQR VTI-LDGHLRR FQNAFSQYK-K AVLFPDGQAV 234
hCL-K1   60 -GRHGKIGPI GSKGEKGDSG 78-98-LRKA IGEMDN-QVSQ LTSELKFIK-N A--VAGVRET 129
hCL-L1   80 -------GIP GEKGKAGT-- VCDCGRYRKF VGQ-LDISIAR LKTSMKFVK-N --VIAGIRET 128
hCL-P1       ---------- ---------- ---------- ---------- ---------- ----------

hSP-D   241 GEKIFKTAGF VKPFTEAQLL CTQAGGQLAS PRSAAENAAL Q-QLVVAKNEA AFLSMTDSKT 300
rSP-D   242 GDKIFRAANS EEPFEDAKEM CRQAGGQLAS PRFATENAAV Q-QLVTAHSKA AFLSMTDVGT 301
bSP-D   235 GEKIFKTVGS EKTFQDAQQI CTQAGGQLPS PRSGAENEAL T-QLATAQNKA AFLSMSDTRK 294
hSP-A1  115 GEKVFSSNGQ SITFDAIQEA CARAGGRIAV PRNPEENEAI A-SFVKKYNTY AYVGLTEGPS 174
hSP-A2  115 GEKVFSSNGQ SITFDAIQEA CARAGGRIAV PRNPEENEAI A-SFVKKYNTY AYVGLTEGPS 174
bSP-A   115 GEKVFSTNGQ SVNFDAIKEL CARVGGHIAA PRSPEENEAI V-SIVKKYNTY AYLGLVEGPT 174
rSP-A   115 GDKVFSTNGQ SVNFDTIKEM CTRAGGNIAV PRTPEENEAI A-SIAKKYNNY VYLGMIEDQT 174
rMBP-A  108 GKKFFVTNHE RMPFSKVKAL CSELRGTVAI PRNAEENKAI Q-E--VAKT-S AFLGITDEVT 164
rMBP-C  113 GKKYFMSSVR RMPLNRAKAL CSELQGTVAT PRNAEENRAI Q-N--VAK-DV AFLGITDQRT 169
hMBP    115 GNKFFLTNGE IMTFEKVKAL CVKFQASVAT PRNAAENGAI Q-NLI---KEE AFLGITDEKT 171
bCL-46  235 GKKIFKTAGA VKSYSDAQQL CREAKGQLAS PRSAAENEAV A-QLVRAKNND AFLSMNDIST 294
bCL-43  184 GEKIFKTAGA VKSYSDAEQL CREAKGQLAS PRSSAENEAV T-QLVRAKNKH AYLSMNDISK 243
bCongl  235 GEKIFKTAGA VKSYSDAEQL CREAKGQLAS PRSSAENEAV T-QMVRAQEKN AYLSMNDIST 294
hCL-K1  130 ESKIYLLVKE EKRYADAQLS CQGRGGTLSM PKDEAANGLM AAYLAQAGLAR VFIGINDLEK 190
hCL-L1  129 EEKFYYIVQE EKNYRESLTH CRIRGGMLAM PKDEAANTLI ADYVAKSGFFR VFIGVNDLER 189
hCL-P1  608 ---------- ---FEDAKLF CEDKSSHLVF INTREEQQWI KKQMVGR--ES HWIGLTDSER 653

hSP-D   301 EGKFTYPTGE SL-VYSN-WAP-G EPNDDGGS----ED CVEIFTNGKW NDRACGEKRL VVCEF 355
rSP-D   302 EGKFTYPTGE AL-VYSN-WAP-G EPNNNGGA----EN CVEIFTNGQW NDKACGEQRL VICEF 356
bSP-D   295 EGTFIYPTGE PL-VYSN-WAP-Q EPNNDGGS----EN CVEIFPNGKW NDKVCGEQRL VICEF 349
hSP-A1  175 PGDFRYSDGT PV-NYTN-WYR-G EPAGR-GK----EQ CVEMYTDGQW NDRNCLYSRL TICEF 228
hSP-A2  175 PGDFRYSDGT PV-NYTN-WYR-G EPAGR-GK----EQ CVEMYTDGQW NDRNCLYSRL TICEF 228
bSP-A   175 AGDFYYLDGA PV-NYTN-WYP-G EPRGR-GK----EK CVEIYTDGQW NDKNCLQYRL AICEF 228
rSP-A   175 PGDFHYLDGA SV-NYTN-WYP-G EPRGQ-GK----EK CVEMYTDGTW NDRGCLQYRL AVCEF 228
rMBP-A  165 EGQFMYVTGG RL-TYSN-WKK-D EPNDHGSG----ED CVTIVDNGLW NDISCQASHT AVCEF 219
rMBP-C  170 ENVFEDLTGN RV-RYTN-WNE-G EPNNVGSG----EN CVVLLTNGKW NDVPCSDSFL VVCEF 224
hMBP    172 EGQFVDLTGN RL-TYTN-WNE-G EPNNAGSD----ED CVLLLKNGQW NDVPCSTSHL AVCEF 226
bCL-46  295 EGKFTYPTGE SL-VYSN-WAS-G EPNNNNAGQP--EN CVQIYREGKW NDVPCSEPLL VICEF 351
bCL-43  244 EGKFTYPTGG SL-DYSN-WAP-G EPNNRAKDEGP-EN CLEIYSDGNW NDIECREERL VICEF 301
bCongl  295 EGRFTYPTGE IL-VYSN-WAD-G EPNNSDEGQP--EN CVEIFPDGKW NDVPCSKQLL VICEF 351
hCL-K1  191 EGAFVYSDHS PMRTF-NKWRS-G EPNNAYDE----ED CVEMVASGGW NDVACHTTMY FMCEF 246
hCL-L1  190 EGQYMFTDNT PLQNYSN-WNE-G EPSDPYGH----ED CVEMLSSGRW NDTECHLTMY FVCEF 245
hCL-P1  654 ENEWKWLDGT SPD-YKN-WKA-G QPDNWGHGHGPGED CAGLIYAGQW NDFQCEDVNN FICE- 711
```

Fig. 1.7 - Sequence comparison of 16 collectins, 15 extracellular soluble and 1 membrane bound (hCL-P1). The sequences were sourced from National Centre of Biotechnology Information (NCBI - http://www.ncbi.nlm.nih.gov/) and the sequence analysis was performed using blast (NCBI - http://blast.ncbi.nlm.nih.gov/Blast.cgi). Each collectin was analysed against other collectins, with 14 out of the 15 extracellular soluble collectins containing 19 conserved residues (highlighted in blue), with hCL-K1 containing 18 conserved residues. The amino acids important in coordinating the Ca1 are shown in bold. Highlighted in yellow are conserved amino acids present in the classical collectins.

1.4.2.3 N-terminal region

The N-terminal region of collectins shows little homology within the collectin family. Although there is limited similarity, the conservation of cysteine residues in all collectins seems to be a structural feature, with different numbers of cysteines present in each collectin N-terminal. Collectins typically contain an N-terminal region of 7-28 residues in length and containing 1-3 cysteine residues, see **Table 1.5** (**Velduizen et al., 2011**; **Crouch et al., 1994**; **Gupta & Surolia, 2007**). Of the 9 collectins reported to date, 8 contain a typical N-terminus of variable length, with the exception of Human CL-P1, which is a type II membrane protein containing an atypical N-terminus. CL-P1 includes an N-terminal cytoplasmic region (37 residues long), followed by a transmembrane domain (21 residues long) (**Yoshida et al., 2003**).

The number of cysteines present in the N-terminal has been suggested in the literature to be important in the ability of collectins to form higher order multimers. One of the studies showed that mutations in one of the two cysteine residues within the N-terminus in hSP-D, to a serine, failed to produce tetramers of trimers (natural assembly of SP-D), with only trimers reported. This study also reported that the mutation of cysteine to serine caused a phenotype similar to the one reported in SP-D gene knockout mice (SP-D (-/-), where emphysema and foamy macrophages developed in models of SP-D (-/-) mice, highlighting the importance of the multimerisation of SP-D in lung homeostasis (**Zhang et al., 2001**).

Two cysteines have been suggested as necessary for the multimerisation of the trimers into higher order multimers. This suggestion is supported by the fact that CL-K1 and CL-L1 contain only one cysteine residue in the N-terminal and seem to occur solely as a trimer, with no higher multimers reported to date (**Ohtani et al., 1999**; **Axelgaard et al., 2013**; **Henriksen et al., 2013**; **Hansen et al., 2010**). This does not seem to apply to all collectins; CL-43 is secreted as a trimer, despite containing two cysteine residues located in the N-terminus in similar positions to those in SP-D, known to produce higher order

multimers. The difference between SP-D and CL-43 lies in the collagen region, with CL-43 containing a smaller collagen region than SP-D, which could impair additional interaction between trimers and inhibit the formation of higher order multimers. An example of how the collagen region can impact on the ability to produce high order multimers can be seen in human SP-A. SP-A exists as two isoforms, SP-A1 and SP-A2, with SP-A1 containing an additional cysteine at position 85, allowing for intermolecular disulphide bond formation and the creation of larger multimers than SP-A2 (**Wang et al., 2007**). Therefore, both the N-terminal and collagen regions appear to play crucial roles in allowing high order multimers to be formed.

1.4.2.4 Collagen Region

Collagenous regions are widely distributed in the animal kingdom, with current structural information on collagen regions derived from fibre diffraction studies and X-ray crystallographic studies (**Beck & Brodsky, 1998**; **Gingras et al., 2011**). Collagen structures can be recognised by the repetitive triplet of Gly-Xaa-Yaa sequences, where Xaa and Yaa can be any amino acid, but frequently include proline (**Kishore et al., 2006**; **Haagsman et al., 2008**). Collagen is formed by three left-handed helices which form a right handed collagen triple helix, which in the collectins is connected to the neck of three right handed helices which form a left handed coiled coil (**Shrive et al., 2003**). The collagen triple helix is held by interchain hydrogen bonds, between the amine groups of Glycine and carboxylic groups of the X amino acid. The three chains composing the collagen region are highly exposed to solvent, making the triple-helical collagen region a prime target for interacting with other molecules and for self-association (**Kramer et al., 1999**).

The length of the collagen region varies across the nine reported collectins (see **Table 1.5**), with hSP-D containing the largest collagen domain reported at 177 residues long, while the shortest collagen chain reported is 53 residues long in rMBP-A. Interruptions in the Gly-X-Y repeat pattern have been reported within the collagen domain

of SP-A and MBP, introducing a kink or region of flexibility in the SP-A and MBP trimeric stems, which then angle away from the central core, producing the bouquet-like structure, similar to C1q (**Wallis & Drickamer, 1999**; **Voss et al., 1988**).

The biological importance of the kink seen in SP-A and MBP remains unknown, however mutational studies where two SP-A proteins were constructed without the interruption of the Gly-X-Y sequence, showed that the kink seen in SP-A and MBP is critical in the formation of the bouquet-like structure. The two SP-A proteins constructed were composed of the straight shaped collagenous domains under microscopic observation, with no resemblance to the usual bouquet like structure (**Uemura et al., 2006**).

The collagen region holds various biological functionalities, including the ability to produce higher order multimers and the ability to interact with immune receptors, yielding a pro or anti-inflammatory response. In MBP, the collagen region has been reported to be important in the binding of MBP-associated serine proteases (MASPs) and activation of the lectin pathway of complement (**Thiel et al., 1997**; **Matsushita et al., 1995**). SP-D has been reported to interact with various soluble receptors through the collagen region and this is discussed in more detail in 1.4.1.5.10.

1.4.2.5 α-helical neck region

The α-helical neck region is present in all collectins, consisting of short stretch 17-37 residues long (see **Table 1.5**). The neck region has been proposed to be the main site where the trimerisation in collectins occurs, leading to the formation of a triple-helical parallel-coiled coil, which acts as a linker between the collagen region and the CRD of all three monomers (**Shrive et al., 2003**; **Kovacs et al., 2002**; **Hoppe et al., 1994**).

The crucial role played by the α-helical neck region in the trimerisation of the three separate monomers into a trimer has been widely investigated in the literature. A study using NMR spectra reported that a fragment of human SP-D, containing 64 residues

(seven Gly–X-Y repeats of the collagen (7x3=21) + 1 amino acid, the α-helical neck domain of 35 amino acids and 7 residues of the CRD), was used in collagenase digestion, removing all collagen triplets. The resulting fragment was able to undergo trimerisation, suggesting that trimerisation is independent of the collagen region and complete CRD (**Hoppe et al., 1994**). In contrast, studies into rat MBP-A neck + fragment of CRD was reported to be unable to form trimers, with the complete CRD region necessary for a stable trimer to be formed. Isolation of the CRD (without the neck region) has been reported to only produce dimers, with no trimers reported (**Weis & Drickamer, 1994**). It is interesting to note the differences seen in both studies, where the human SP-D isolated neck region can form stable trimers, while for rat MBP-A to form stable trimers, the complete CRD and neck region is necessary (**Hoppe et al., 1994**; **Weis & Drickamer, 1994**).

The reasons for these differences in terms of trimerisation properties has been suggested to occur due to differences in the length of the complete heptad repeat pattern present in the α-helical neck region of each collectin. Heptad repeats are blocks of seven residues, termed abcdefg, where a and d are often hydrophobic (non-polar) amino acids (often Leu or Ile), that can act as a linker to other protein domains (**Thormahlen et al., 1998**). Human SP-D contains four complete heptad repeats (**Zhang et al., 2001a**; **Hoppe et al., 1994**), while rat MBP-A contains three complete heptads (**Weis & Drickamer, 1994**; **Kovacs et al., 2002**). In SP-D and MBP, the heptad repeat pattern is composed of hydrophobic amino acids at a and d, which help to stabilise the neck domain; however, the fact that SP-D contains a fourth heptad repeat, means there are additional hydrophobic residues in the neck region that interact with the CRD. In SP-D, valine is present at position 204, 211 and 218, with Leucine present at position 207, 214 and 221, Phenylalanine at position 225 and Tyrosine at position 228, providing the α-helical turns prior to the helix termination (**Fig. 1.5** and **Fig. 1.7)** (**Zhang et al., 2001a**; **Kovacs et al., 2002**; **Weis & Drickamer, 1994**). The α-helical neck region therefore acts as a trimerising

linker holding the three CRDs in place, by means of strong hydrophobic forces. In human SP-D, the last residue of the neck region is a cis Proline at position 235, which causes a change to the fold from a α-helix to a β-strand **(Shrive et al., 2003**).

1.4.2.6 Carbohydrate Recognition Domain (CRD)

All collectins reported to date share a highly conserved carbohydrate recognition domain (CRD) and secondary structure homology (**Cummings & McEver, 1999**). The CRDs range between 113-153 residues long (**Table 1.5** and **Fig. 1.7)**, with noticeable similarities in the conserved binding pocket motifs. The CRD domain is the site where the collectins bind to non-self glycoconjugates on the surface of microorganisms in a calcium dependent manner, leading to neutralization, aggregation and opsonisation by macrophages, in combination with different collectin receptors (**Hansen et al., 2002**; **LeVine et al., 2000**; **Palaniyar et al., 2003**; **Hartshorn et al., 1997**).

The trimerisation of the neck and CRDs is important in determining the relative orientation of the CRDs, thus impacting on antigen/ligand binding and recognition for all the collectin CRDs (**Kovacs et al., 2002**). The structural organisation of some CRDs has been reported to vary in different species and two examples of this are; the rat SP-A CRD domain produces a “T-shaped” structure, while MBP and SP-D produce a “Y-shaped” structure, with larger interdomain angles between the neck and the CRD (**Head et al., 2003**). In addition, different species contain different spatial orientation in the CRDs. The rat MBP CRDs are separated by 45 Å from each other, while in human MBP the separation is 53 Å (**Sheriff et al., 1994; Weis & Drickamer, 1994**). Sheriff and colleagues postulated that, similar to antibodies, the difference in spatial orientation of the CRDs across different species of collectins, could be the cause of differences in binding affinity for different antigenic regions (**Sheriff et al., 1994**). The ranking of collectins binding affinities are discussed in 1.4.5.

The CRD contains a long loop region (LLR) (**Fig. 1.5)** of structural and biological significance, as it is involved in the Ca^{2+} dependent carbohydrate binding reported in all collectins to date (**Veldhuizen et al., 2011**). In the LLR region three calcium ions have been reported; Ca1, Ca2 and Ca3. Calcium has the ability to make 6, 7 or 8 coordination bonds to moieties with partial or complete negative charge (see **Fig. 1.8**) (**Katz et al., 1996**).

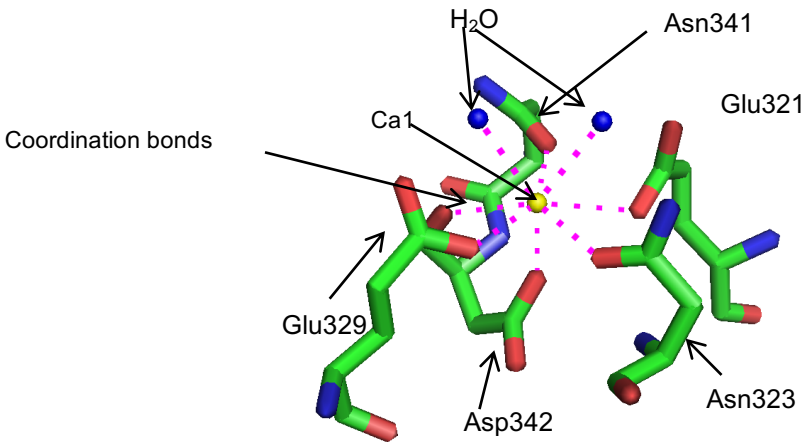


Fig. 1.8 An illustration of the coordination bonds of Ca1 with 5 coordinating residues of the rfhSP-D, including Glu321 (E), Asn323 (N), Glu329 (E,) Asn341(N) and Asp342 (2x) (D) and 2 coordination bonds made with 2 water molecules in the ligand-free structure or with 2 hydroxyls in the ligand-bound structure. The coordination bonds are displayed in pink dashed lines. Figure produced using *PYMOL*.

In the collectin protein family, 8 coordination interactions have been reported; six highly conserved coordination bonds to the binding site residues and two coordination bonds with either two water molecules in the ligand-free structure or two hydroxyl groups of bound ligand (**Shrive et al., 2003**; **Shrive et al., 2009**; **Hakansson et al., 1999**; **Wang et al., 2008**; **Clark et al., 2016).** The Ca^{2+} is important in the recognition of glycoconjungates present in PAMPs and DAMPs and was initially denominated in the literature as Ca2 or Ca^{2+} binding site 1. In order to avoid confusion, throughout this thesis Ca^{2+} binding site 1 will be denominated Ca1 (**Haagsman et al., 1990**; **Veldhuizen et al., 2011**).

Ca1 is directly involved in the binding of carbohydrates and is responsible for the

Ca^{2+} binding dependence seen in collectins. In the binding pocket, Ca1 is coordinated by five conserved residues present in the binding pocket of the CRDs of the collectins, and using hSP-D numbering these are; Glu321 (E), Asn 323 (N), Glu329 (E), Asn341 (N) and Asp342 (D) which makes two interactions with Ca1 (**Fig. 1.8**) (**Shrive et al., 2003**; **Shrive et al., 2009**). This produces a highly conserved motif, the **E**P**N**/**E**/W**ND** motif, conserved in other collectins and species, such as rSP-D, bSP-D, rMBP-A, rMBP-C, hMBP, bCL-43, bCL-46, bConglutinin, hCL-K1, see also 1.4.5.2.

Three other Ca^{2+} ions have been reported in the collectins (Ca2, Ca3 and Ca4) (**Shrive et al., 2003**; **Sheriff et al., 1994**; **Weis et al., 1992**). Ca1, Ca2 and Ca3 (see **Fig. 1.5**) have been suggested to be important in the stability of the CRD. A study by Haagsman et al., 1990, using a collagenase-resistant fragment of SP-A, reported a link between resistance to proteolysis by both trypsin and chymotrypsin and Ca^{2+} concentrations, with a higher concentration of calcium ions producing a higher percentage of SP-A that remained undigested (**Haagsman et al., 1990**). A second study also highlighted the importance of the role of calcium ions in the CRD, where it was reported that the removal of Ca^{2+} affected the core structure of the CRD, with dramatic conformational changes in the loops. The absence of Ca2, which mediates the interaction between L1 and L4, produced a rigid body movement of L1 and some changes in L4. Ca1 interacts with residues present in L3 and L4, with the Pro235 residue at the junction of the L3 and L4 loops in a cis conformation, locking L3 and L4 into a conformation that surrounds Ca1. The removal of Ca1 in MBP leads to the Pro235 residue adopting both the cis and trans conformations, with no preference reported, impacting on the structure of L3 and L4. The trans conformation induces an extension of the loop away from the core of the protein, culminating in loss of carbohydrate binding affinity and increased sensitivity to proteolysis **(Ng et al., 1998**).

Ca4 has been reported to be present in the structural studies of a recombinant fragment of human SP-D, where it was observed to occur at the bottom of the funnel

formed by the branching of the neck into the three CRDs, with the suggested role of perhaps acting as a secondary binding site for extended natural ligands. Ca4 is not present in all the structural studies reported, with the number of Ca^{2+} bound to each CRD under physiological conditions unclear at present. Three Ca^{2+} (Ca1, Ca2 and Ca3) are commonly found in X-ray crystal structures, with Ca4 suggested to be present when the concentration of Ca^{2+}, in the crystallisation and cryo-protection stages, is 2 mM or above (**Shrive et al., 2003**).

1.4.3 Distinction between Self and Non-self

Collectins have the ability to distinguish between self-cells and non-self-cells. Even though collectins have been reported to possess a broad selectivity for various sugars (**Table 1.6**), the presentation of the sugars in microbes and the geometrical arrangement of the CRDs have been suggested as the two mechanisms by which the discrimination between host cells glycoconjugates and the PAMPs and DAMPs occurs (**Weis et al., 1998**). Collectin CRDs selectively bind to carbohydrates such as glucose, mannose, heptose and fucose, prevalent on glycolipids and glycoproteins decorating the surface of microbes, but not to sialic acid and other common carbohydrates commonly present as terminal sugars in host cell surface molecules. Self-cells normally carry widely spaced glycoconjungates, to which the CRDs of collectins cannot bind efficiently. The distance between each individual CRDs makes self-cells poor ligands for collectins, as these distances are often too far apart, making it impossible for more than one CRD within the trimer to interact with the self-cells. On the other hand, microbes often present narrowly spaced glycoconjungates on their surfaces, ideally positioned for the binding of multiple CRDs in the trimer (**McCormack & Whitsett, 2002**; **Weis & Drickamer, 1994**).

1.4.4 Glycosylation of collectins

The glycosylation of collectins has been suggested to be an important post-translational modification, impacting on signalling potential. Glycosylation is a chemical

process whereby carbohydrates are attached to proteins, as a consequence of co-translational and post-translational modifications. Glycosylation can occur by means of different attachment mechanisms; N-linked glycans are attached to the nitrogen atom on Asn side chains, O-linked glycans are attached to the oxygen of the hydroxyl groups in Ser, Thr and Tyr; S-linked glycans are attached to the Sulphur groups of Met and Cys. The CRD of rSP-A (rat) and pSP-D (porcine) have been reported to be N-glycosylated (**Head et al., 2003**; **van Eijk et al., 2002**). CL-46 has been suggested to contain N-glycosylation in the collagen region, similar to Human SP-D (**Hansen et al., 2002**; **Leith-Larsen et al., 1999**). Human SP-D has been reported to be glycosylated in the collagen region and the N-terminal with the collagen region reported to contain both O- and N-linked carbohydrates (**Leith-Larsen et al., 1999**). The N-terminal of hSP-D has also been reported, both in *in vivo* and *in vitro,* to be S-glycosylated by pulmonary nitric oxide (NO), leading to the disruption of the multimer. The NO modifies the cysteine residues in the N-terminal leading to dissociation into trimers, whereby the N-terminal-NO tail binds to calreticulin/CD91 initiating a pro-inflammatory response (**Guo et al., 2008**; **Atochina-Vasserman, 2011**).

1.4.5 Sequence, structure and function

1.4.5.1 Collectin binding affinities

Collectins possess a broad range of monosaccharide specificity and affinity, with structural variations impacting on the ability of each collectin to specifically recognise and interact with different glycoconjugates, the ranking of binding affinities for each of the collectins are given in **Table 1.6.**

The binding affinity of the diverse collectins varies, with the different residues present in and around the binding pockets contributing to specificity and affinity towards specific carbohydrates, with single and double mutational studies supporting this view. Classical collectins have a higher affinity for GlcNAc, ManNAc and Mannose, with the

novel collectins showing a higher affinity for L-fucose, GalNac and Mannose. The binding affinities of novel collectins have been demonstrated to be reduced in comparison with the classical collectins (**Hakansson & Reid**; **2000**; **Keshi et al., 2006**).

Collectin	Class	Ranking of relative binding affinities	Binding affinity established by:	Literature reported in:
MBP	Classical	Mannose > GlcNAc > L-Fucose > ManNAc > Glucose	ELBA	**Haurum et al., 1993**
SP-A	Classical	ManNAc >>> L-Fucose, Maltose > Glucose > Mannose **SP-A2 higher affinity than SP-A1 for:** Mannose, L-Fucose, GlcNAc and GalNAc	ELBA. Carbohydrate binding assay.	**Haurum et al., 1993 Hansen et al., 2002 Oberley & Sneyder, 2003**
Conglutinin	Classical	GlcNAc >>> Mannose, ManN, L-Fucose > Glucose > ManNAc, GlcN, Maltose	ELBA	**Haurum et al., 1993 Hansen et al., 2002**
CL-46	Classical	GlcNAc >>> ManNAc, Mannose > αMe-Man > Maltose and GlcN	Competition assay GlcNAc-BSA	**Hansen et al., 2002**
CL-43	Classical	Mannose >>> ManNAc > L-Fucose> GlcNAc > Glucose and Maltose >>> Galactose >Lactose >>GalNAC	ELBA	**Holmskov et al., 1993 Hartshorn et al., 2002 Hansen et al., 2002**
SP-D	Classical	L-D-heptose> Maltose and Posphatidylinositol (PI) > Glucose > ManNac, ManN and GluN> D-Fucose and GalN > Mannose and D,D-Heptose >>>Galactose, L-Fucose and Lactose > GluNAc >GalNAc	Competion assay maltosyl-BSA, mannose and LPS. ELISA	**Persson et al., 1990 Ogasawara et al., 1992 Lim et al., 1994; Haczku et al., 2008 Wang et al., 2008 Hansen et al., 2002**
CL-K1	Novel	L-Fucose > weakly to mannose	Sugar-blot method and western blot.	**Keshi et al., 2006**
CL-P1	Novel	GalNAc > L,D-Fucose and Galactose >>> Glucose > GlcNAc and Mannose	Competion assay GalNac, fucose, galactose and mannose.	**Yoshida et al., 2003**
CL-L1	Novel	Mannose, L-Fucose >> ManNAc >>> Glucose and GlcNAc	Sugar-blot method and western blot.	**Axelgaard et al., 2013 Ohtani et al., 1999**

Table 1.6 The ranking of binding affinities for both the classical and novel collectins, >, >> and >>> more than 10, 50 and 100% respectively, higher binding than previous saccharide. ELBA - Enzyme-linked lectin binding assay, BSA – Bovine serum albumin.

Carbohydrate binding by SP-D seems to preferentially recognise the α-anomeric configuration of carbohydrates that consist of a six-membered ring (five carbons and one oxygen) of either glucose or mannose type. The binding specificities and affinities of SP-D towards a variety of carbohydrates has been investigated by using ligand binding assays and binding competition assays of maltosyl-BSA, mannose and LPS (**Table 1.6**). SP-D

shows preferential binding affinity for L-D-heptose> Maltose and Phosphatidylinositol (PI) > Glucose > ManNac, ManN and GluN> D-Fucose and GalN > Mannose and D,D-Heptose >>> Galactose, L-Fucose and Lactose > GlcNAc > GalNAc (**Persson et al., 1990**; **Ogasawara et al., 1992**; **Lim et al., 1994**; **Haczku et al., 2008**; **Wang et al., 2008**).

1.4.5.2 Ligand binding pocket

Collectins contain a fairly conserved calcium and ligand binding pocket motif. The amino acids important in coordinating the Ca1 are shown in **Fig 1.9** with some of these residues, in addition to Ca1, also involved in ligand binding.

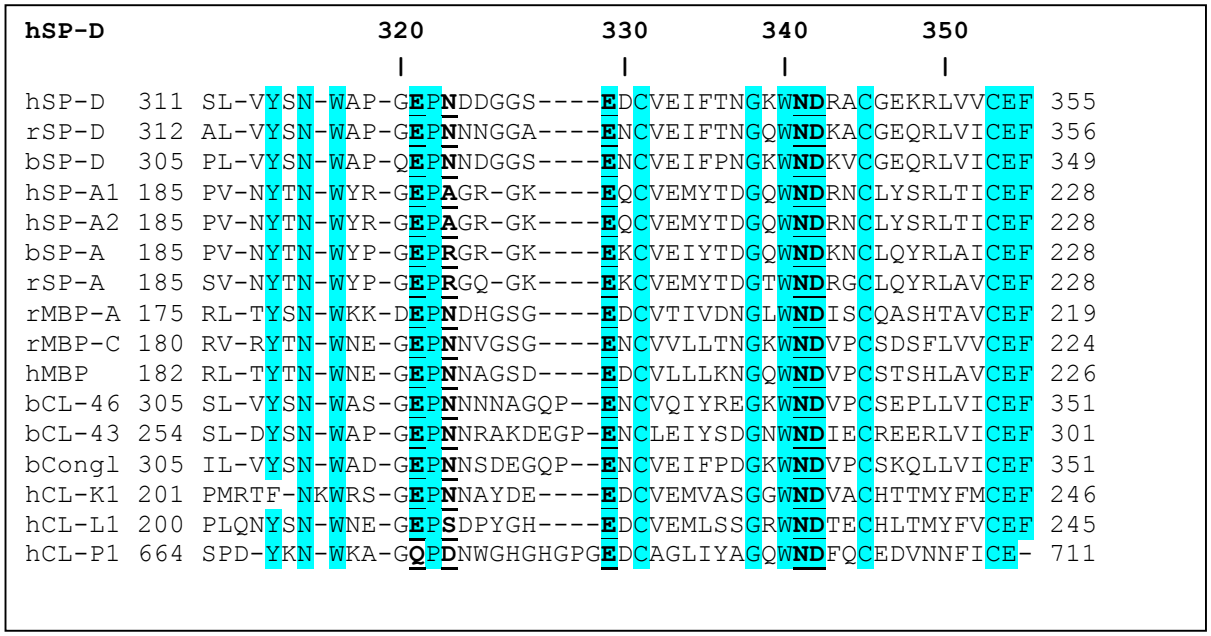


Fig. 1.9 C-terminal sequences of selected collectins. Residues in the binding pocket which also coordinate to Ca1 are in bold and underlined. In light blue, the conserved residues of the various collectins analysed is shown. hSP-D, rSP-A, rMBP-A, rMBP-C and hMBP represent the only collectins whose structure has been solved.

The CRD of the collectins contains some structural conservation, with the variations reported leading to minimal structural diversity in the binding pocket of the CRD region. Even though the variations are often single site substitutions, these can impact on the specificity and affinity for different carbohydrates. The impact of these variations, in combination with variations in the residues flanking the binding pocket (positions hSP-D

325 and 343) is the main reason why some collectins exhibit a higher affinity for specific carbohydrates present on the variety of pathogens.

Each collectin possesses a broad monosaccharide specificity and in general terms collectins can be divided into mannose/glucose type or galactose type (see also 1.3). Two motifs have been reported; the **E**P**N**/**E**/W**ND** motif in hSP-D, rSP-D, bSP-D, rMBP-A, rMBP-C, hMBP, bCL-43, bCL-46, bConglutinin and hCL-K1 has been shown to have a high specificity for mannose, glucose (mannose is a C-2 epimer of glucose) and heptose (**Zelensky & Gready, 2005**; **Shrive et al., 2003**; **Shrive et al., 2009**; **Wang et al., 2009**; **Clark et al., 2016**). In other collectins variation occurs in the **E**P**N** tri-residue, especially in the Asn (N) residue. hSP-A1 and hSP-A2 contains **E**P**A,** bSP-A and rSP-A contains **E**P**R** and hCL-L1 contains **E**P**S.** Understanding the **E**P**X**/**E**/W**ND motif** and its variations is crucial in understanding differences in binding affinities. hCL-P1, the only membrane bound collectin reported, possesses a **Q**P**D**/**E**/W**ND** motif (**Table 1.6** and **Fig. 1.9**).

SP-A in Humans, rats and bovine show some variation in the **E**P**N** tri-residue, where the third residue of the motif is substituted by an arginine (**E**P**R**) for rat and bovine SP-A and an alanine (**E**P**A**) for human SP-A (both A1 and A2). The change from **E**P**N** to either **E**P**R** or **E**P**A** does not seem to affect the binding affinity for mannose/glucose type carbohydrates (**Haurum et al., 1993**). Single mutational studies have shown that **E**P**N**, **E**P**R** and **E**P**A** have similar carbohydrate binding properties; rSP-A Arg (R) was substituted by Asn (N) and Ala (A) with no changes in the activity of rSP-A carbohydrate binding, type II cell binding, inhibition of surfactant secretion, lipid binding, lipid aggregation and lipid uptake by type II cells (**Pattanajitvilai et al., 1998**). This suggests that Asn (N) is not completely essential for the carbohydrate binding specificity towards mannose/glucose type carbohydrates, with Ala (A) and Arg (R) efficiently mimicking binding properties towards mannose/glucose carbohydrates.

Human CL-L1 has also been shown to contain a variation in the **E**P**N** motif, with

the third residue changed from an Asn (N) to a Ser (S), producing an **E**P**S** motif. Contrary to SP-A (**E**P**A** and **E**P**R** motifs), this modification has been reported to allow hCL-L1 to recognise both mannose/glucose and galactose type sugars. hCL-L1 has been described as a "hybrid" collectin, placed between the normal mannose/glucose specific collectins and the collectins specific to galactose (**Ohtani et al., 1999**). It is interesting to note that this substitution of Asn (N) to a Ser (S) has also been reported in chicken SP-A (**Hogenkamp et al., 2006**).

Human CL-P1 is the only collectin reported to date to have high specificity for galactose type carbohydrates. This has been suggested to be due to the **Q**P**D**/**E**/W**ND** **(Fig. 1.9**) motif and this has been shown in mutational studies, where a mutant of MBP containing **Q**P**D** instead of the typical **E**P**N**, exhibited binding specificity changes, with a higher affinity towards galactose than mannose type carbohydrates (**Drickamer, 1992**).

1.4.6 Structural studies

Structural studies of rat MBP, human MBP and rat SP-A have been carried out in addition to the large number of structural studies of hSP-D which are detailed elsewhere in this thesis. MBP was the first collectin to have its structure solved, with the rat MBP-A structure solved using MAD phasing (**Weis et al., 1991**); and a year later the first structure of rat MBP-A CRD complexed with ligand, oligomannose asparaginyl-oligosaccharide ($Man_6GlcNAc_2Asn$), was reported (**Weis et al., 1992**). The trimeric human MBP structure was reported in 1994 by Sheriff and colleagues, with MBP reported to organise itself into a hexamer of trimers (bouquet of flowers) (**Sheriff et al., 1994**). rSP-A has been investigated structurally in two different studies (**Head et al., 2003**; **Shang et al., 2011**), both in the ligand-free state and in ligand binding studies, see **Table 1.7** for key structural studies reported to date for rMBP-A, rMBP-C, hMBP and rSP-A and the key findings reported.

Collectin	Ligand-free or Ligand-bound	Key findings and extra information.	Literature reported in:
rMBP-A	Ligand-free	The first reported structural study of collectins. Ligand-free dimeric fragment of rMBP-A used.	**Weis et al., 1991**
	Ligand-bound oligomannose asparaginyl-oligosaccharide ($Man_6GlcNAc_2Asn$)	The first reported ligand binding structural study of collectins. Ligand-free dimeric fragment of rMBP-A used. rMBP-A recognises the ligand through the CRD and by binding to the equatorial hydroxyls of mannose O3' and O4'.	**Weis et al., 1992**
	Ligand-free	Ligand-free trimeric fragment of rMBP-A containing the neck-CRD displayed the same structure as the dimeric rMBP-A.	**Weis & Drickamer, 1994**
	Ligand-bound galactose and GalNAc	Trimeric fragment of rMBP-A containing neck-CRD used. rMBP-A CRD bound to galactose and GalNAc through the O3' and O4' equatorial hydroxyls.	**Kolatkar & Weis, 1996**
	Ligands-bound - α1-2, α1-3 and α1-6 di-mannose.	Trimeric fragment of rMBP-A containing neck-CRD used. rMBP-A CRD bound to the various di-mannose through the O3' and O4' equatorial hydroxyls. The orientation of the α1-6 di-mannose was reported to differ from the orientation of the α1-2 and α1-3 di-mannoses, by binding in an orientation rotated 180° in regards to the CRD (O3' and O4' interchanged by 180°).	**Ng et al., 2002**
rMBP-C	Ligand-bound – mannose, fucose and GalNAc	Ligand-free dimeric fragment of rMBP-C used. rMBP-C CRD bound to the various ligands through the O3' and O4' equatorial hydroxyls. The binding mechanism of rMBP-C CRD towards mannose, fucose and GalNAc differed from the binding mechanism of rMPB-A CRD, with the hydroxyls rotated by 180° in regards to the CRD (O3' and O4' interchanged by 180°).	**Ng et al., 1996**
hMBP	Ligand-free	Trimeric fragment of hMBP containing neck-CRD used and displayed a similar structure to the trimeric fragment of rMBP-A	**Sheriff et al., 1994**
rSP-A	Ligand-free	Trimeric fragment of rSP-A containing neck-CRD used. The trimeric fragment of rSP-A was reported to be similar to the CRD fold reported for SP-D and MBP, with the only difference of the trimeric fragment of rSP-A containing a nearly perpendicular neck-CRD orientation, with lower interdomain angles.	**Head et al., 2003**
	Ligand-bound – mannose, α-methylmannose and glycerol	Trimeric fragment of rSP-A containing neck-CRD used. rSP-A CRD bound to mannose and α-methylmannose through the O3' and O4' equatorial hydroxyls. Cryo-protectant glycerol molecule was also shown to bind at Ca1.	**Shang et al., 2011.**

Table 1.7 Key structural studies reported to date for rMBP-A, rMBP-C, hMBP and rSP-A and the key findings reported for each study. Individual PDB entries given for each study. GalNAc; N-acetylglucosamine.

1.5 Structural studies of Surfactant Protein D (SP-D)

1.5.1 Recombinant fragment of human SP-D (rfhSP-D)

The work presented in this thesis is focused on the structural studies of a recombinant trimeric fragment of hSP-D complexed with lipopolysaccharides of gram-negative bacteria. The trimeric recombinant fragment rfhSP-D has been expressed in *E. coli* and is composed of a short collagen domain, the α-helical coiled-coil neck region (residues 203-235) and three CRDs (residues 236-355). rfhSP-D constitutes only a small fragment (≈60kDa) of the entire natural tetrameric SP-D, but appears to exhibit similar therapeutic and biological activity both *in vitro* and *in vivo* as the full hSP-D (**Shrive et al., 2009**).

rfhSP-D *in vitro* was reported to have the ability to distinguish between human eosinophils from healthy donors and allergic patients, leading to increased apoptosis, oxidative burst and CD69 expression in allergic eosinophils. An increase of apoptosis of allergic eosinophils, without affecting normal eosinophils, was reported, suggesting an SP-D mediated resolution of allergic-induced eosinophils inflammation (**Mahajan et al., 2008**). Various i*n vivo* studies have investigated the potential of rfhSP-D as a therapeutic agent. rfhSP-D has been reported to reduce serum IgE and peripheral blood eosinophilia following a challenge by *D. pteronyssinus allergen* (**Strong et al., 2003**; **Singh et al., 2003**). rfhSP-D was also shown to bind to the RSV virus, inhibiting RSV entry into epithelial cells and replication (**Hickling et al., 1999**). The challenge of *Asperigillus fumigatus* in mice led to an increase of IgE and IgG antibodies, peripheral and pulmonary eosinophilia, central bronchiectasis and expectoration in affected hosts, with mice dying after 7 days. The intranasal administration of rfhSP-D following the challenge resulted in an 80% survival rate, with similar results to AmB, an antifungal drug nephrotoxic (**Madan et al., 2001**). rfhSP-D has also been reported to have prevented endotoxin shock in new-born lambs challenged with *E. coli* LPS (**Ikegami, 2006**).

The *in vivo* and *in vitro* studies, combined with structural studies, offer evidence for the potential of rfhSP-D as a therapy for various lung conditions. Structural studies have the advantage of providing information at the molecular level of the recognition and binding mechanisms of SP-D and rfhSP-D towards a variety of ligands. A variety of structural studies using rfhSP-D have been reported, with simple mono/di/tri saccharides as the ligands of choice, as these are characteristic components of SP-Ds' natural ligands. Further structural studies of rfhSP-D with natural ligands (LPS) could provide additional knowledge into the recognition of larger and naturally occurring ligands, while providing further evidence of the medical potential of rfhSP-D.

All the structural studies of rfhSP-D (see **Fig 1.10** for the representation of rfhSP-D) in combination with a variety of mono/di/tri saccharides have enhanced the knowledge of the molecular mechanisms of pathogen recognition and binding. Residues which coordinate to Ca1 and ligand in known SP-D structures are included in **Fig 1.9**.

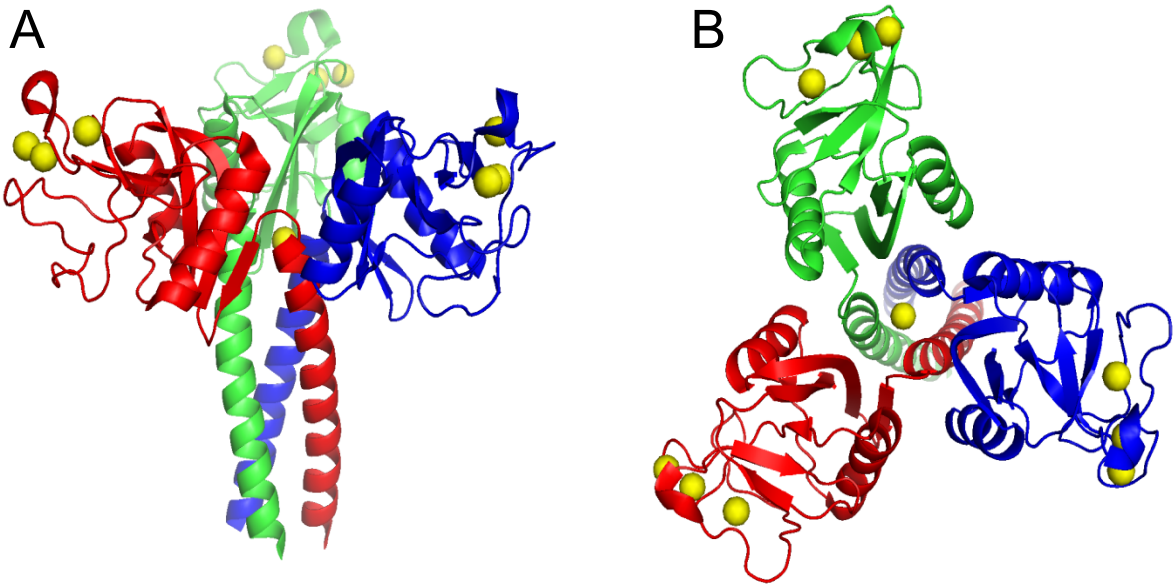


Fig. 1.10 – The three-dimensional structure of ligand-free rfhSP-D. **A**- a front-view of the rfhSP-D trimer with the three subunits forming the trimer, highlighted with different colours, with the Ca1, Ca2 and Ca3 shown for each subunit and with the Ca4 sitting in the neck-CRD interface. **B**- A top view of the rfhSP-D trimer with the three subunits forming the trimer highlighted with different colours, with the Ca1, Ca2 and Ca3 shown for each subunit, and with the Ca4 sitting in the neck-CRD interface. Image generated using *PYMOL* from PDB: 1PW9.

The work presented in this thesis is aimed at elucidating the binding mechanisms of longer natural ligands such as LPS, to characterise bacterial recognition and binding by rfhSP-D following on from previous work performed in our research group. As more structural data becomes available, showcasing the characterisation of the binding of longer ligands by rfhSP-D, the more evidence is presented for the medical potential of rfhSP-D, which may one day be exploited for the benefit of human health

1.5.2 Structural Studies of rfhSP-D

1.5.2.1 Timeline

Year - PDB Accession	Findings and references
1999 - 1B08	1st High resolution X-ray Crystallographic structural study of ligand-free rfhSP-D - **Hakansson et al., 1999**
2003 - 1PW9 1PWB	2nd structural study of ligand-free rfhSP-D with additional interactions reported. 1st reported rfhSP-D ligand-bound structural work with maltose. Binding occurs through equatorial hydroxyls O3' and O4' - **Shrive et al., 2003**
2006 - 2GGU 2GGX	rfhSP-D ligand-bound structural work with Maltotriose and P-Nitrophenyl Maltoside. Suggestion of an extension of ligand binding site to Phe335 - **Crouch et al., 2006**
2007 - 2OS9 2ORK 2ORJ	rfhSP-D ligand-bound structural work with Myoinositol, Inositol-1-phosphate and N-acetylmannosamine. Mutant rfhSP-D Arg343Lys yielded stronger binding for PI, a major surfactant lipid - **Crouch et al., 2007**
2008 - 2RIA 2RIB 2RIC	rfhSP-D ligand-bound structural work with a variety of L-glycero-D-manno-heptoses, highly conserved in LPS. New binding mechanism reported, O6' and O7' hydroxyls groups - **Wang et al., 2008**
2009 - 3IKR 3IKQ 3IKP 3IKN	rfhSP-D ligand-bound structural work with mannobiose α1,4 and α1,2; Inositol Phosphate and Galactose. - **Shrive et al., 2009**
2009 - 3G83 3G81 3G84	2009- rfhSP-D ligand-bound structural work with mannobiose α1,2 and α- Methyl mannoside. Mutant rfhSP-D Arg343Val increased affinity for Mannobiose - **Crouch et al., 2009**
2013 - 4M18	rfhSP-D double mutant Asp325Ala and Arg343Val - 1st rfhSP-D (mutant) structural study inhibiting viral glycoprotein target - **Goh et al., 2013**
2016 - 4E52	1st rfhSP-D ligand bound structural work with natural ligand, *Haemophilus influenzae* Eagan 4A PS, through L-glycero-D-manno-heptose, preferential binding over available glucose - **Clark et al., 2016**

Table 1.8 A timeline of the all the rfhSP-D structural studies using X-ray diffraction and deposited in the protein data bank. The different structural studies are organised by date of report with key information for each structure given. PDB accession codes are given for each structural study published.

1.5.2.2 The ligand-free structure of rfhSP-D

The ligand-free structure of rfhSP-D (**Fig 1.11**) has been investigated in two separate X-ray crystallographic studies. The overall structure of rfhSP-D reveals a trimeric assembly with three terminal CRD domains (236-355), spatially separated and linked to a α-helical coiled-coil neck region (203-235), where the transition of neck-CRD domain is made by the cis Pro235 (**Hakansson et al.,1999**; **Shrive et al., 2003**). Three calcium ions (Ca1, Ca2 and Ca3) were first reported (**Hakansson et al., 1999**), with an additional calcium ion (Ca4) located at the trimeric axis of the neck-CRD interface and reported later (**Shrive et al., 2003**). Ca1 is the main site of ligand interaction and is coordinated by five fairly conserved residues. These coordinations are the result of six interactions between the Ca1 and the side chains of Glu321 (OE1), Asn323 (OD1), Glu329 (OE1), Asn341 (OD1), Asp342 (OD1) and the main chain carbonyl (O) of Asp342 (using SP-D numbering). The protein-Ca1 interaction makes up six of the total of eight interactions of Ca1. The final two interactions are completed by two water molecules in the ligand-free structure of rfhSP-D (**Fig. 1.11B**) (**Hakansson et al.,1999**; **Shrive et al., 2003**).

Ca2 is coordinated by the acidic side-chains of Asp297 (OD1), Glu301 (OE1), Asp324 (OD1) and Asp330 (OD1), and by a water molecule and the main chain carbonyl (O) of Glu329. Ca3 is coordinated by three acidic side chains Glu301 (OE1), Asp330 (OD1 and OD2) and four water molecules (**Fig. 1.11C**). Ca2 has been suggested to be important in the stability of the CRD domain, as it stabilises the L4 through the interactions with the carbonyl (O) of Glu329, reducing any potential flexibility of the side chain and maintaining contact with Ca1 (**Shrive et al., 2003**). Ca2 and Ca3 are located near each other and share a coordinating residue, Glu301. Ca3 has been suggested to play a role in surface receptor interactions. Docking studies of SP-D suggested that flanking saccharide residues in trisaccharides form additional hydrogen bonds with residues outside the CRD binding pocket, contributing to the overall binding energy (**Allen et al., 2001**). At present the full functions of Ca2 and Ca3 remain unknown, however it is interesting to note that

the proximity of Ca2 and Ca3 to the binding site Ca1 may allow for secondary ligand-receptor interactions. Ca4 was reported to be located at the top of the funnel at the intersection of the three monomers and is coordinated by three Glu232 and three water molecules. The presence of the Ca4 seems to be dependent on the Ca^{2+} concentration present during the crystallisation stages (**Shrive et al., 2003**).

1.5.2.3 Ca4 and the asymmetric Tyr228

In the first reported structural study of rfhSP-D no Ca4 was reported. In this initial structure Tyr228 chain C was shown to be asymmetric. Tyr228 is part of the neck region, which is made up of eight helical turns before it is terminated by the helix breaker Pro235. The neck (coiled-coil) contains some buried residues; Val204, Leu207, Val211, Leu214, Val218, Leu221, Tyr228 and Glu232, all hydrophobic except for the Glu232. Glu232 and Lys246 (CRD) form long salt bridges in all subunits A (3.5Å), B (3.7Å) and C (2.8Å) offering the only intramolecular contact between neck (coiled coil) and CRD (**Hakansson et al., 1999**).

When the Ca4 is not present in the neck-CRD interface (**Fig 1.11D**), this causes Glu232 in chain B to assume a new conformation by moving towards the centre of the funnel and the vicinity of Tyr228 in chain C. Tyr228 in chain C in turn moves deeper into the neck, leading to the formation of hydrogen bonds between both residues. Tyr228 chain C is buried in the coiled-coil neck, while in chains A and B the Tyr is exposed and able to hydrogen bond with water molecules. The conformation of Glu232 in chain B in the absence of Ca4 differs from the orientation of Glu232 in chains A and C, as these remain in place and interact with Lys246 in chains A and C respectively. Lys246 in chain B maintains a similar orientation to Lys246 in chains A and C, although it no longer interacts with Glu232 in chain B.

When Ca4 is present at the neck-CRD interface (**Fig 1.11C)**, all three Glu232 (subunit A, B, C) are in a similar conformation and all interact with Lys246 (subunit A, B,

C, respectively) and Ca4, with no interactions with the asymmetric Tyr228 in chain C. The conformation of Glu232-Lys246 in the presence of Ca4 is considerably more symmetric in comparison to the Ca4 depleted structure (**Shrive et al., 2003**). Therefore, two alternative conformations for Glu232 in chain B are possible and this can be seen in the structures represented in **Fig 1.11C** with Ca4 present and **Fig 1.11D** without Ca4 present. This conformational change does not seem to perturb the geometry of the neck-CRD interface (**Hakansson et al., 1999**; **Shrive et al., 2003)**.

The central pore where Ca4 sits has been suggested to be a secondary site for the binding of long extended natural ligands, such as lipopolysaccharides and phospholipids. A study using a rat SP-D neck-region homotrimer (no CRD, just neck coiled-coil) reported strong binding towards PI and DPPC (surfactant lipids), with similar binding affinities as hSP-D. The authors suggested that the SP-D neck homotrimer (no CRD) hydrophobic residues, which are normally buried within the SP-D neck/CRD interface region, but are exposed in the SP-D neck homotrimer (no CRD) have the ability to readily bind to surfactant lipids. Clearly the exposed neck of the SP-D neck homotrimer (no CRD) region does not exist in the biological molecule and therefore no firm conclusions can be drawn regarding the importance of these residues in binding surfactant lipids (**Kishore et al., 1996**; **Shrive et al., 2003**).

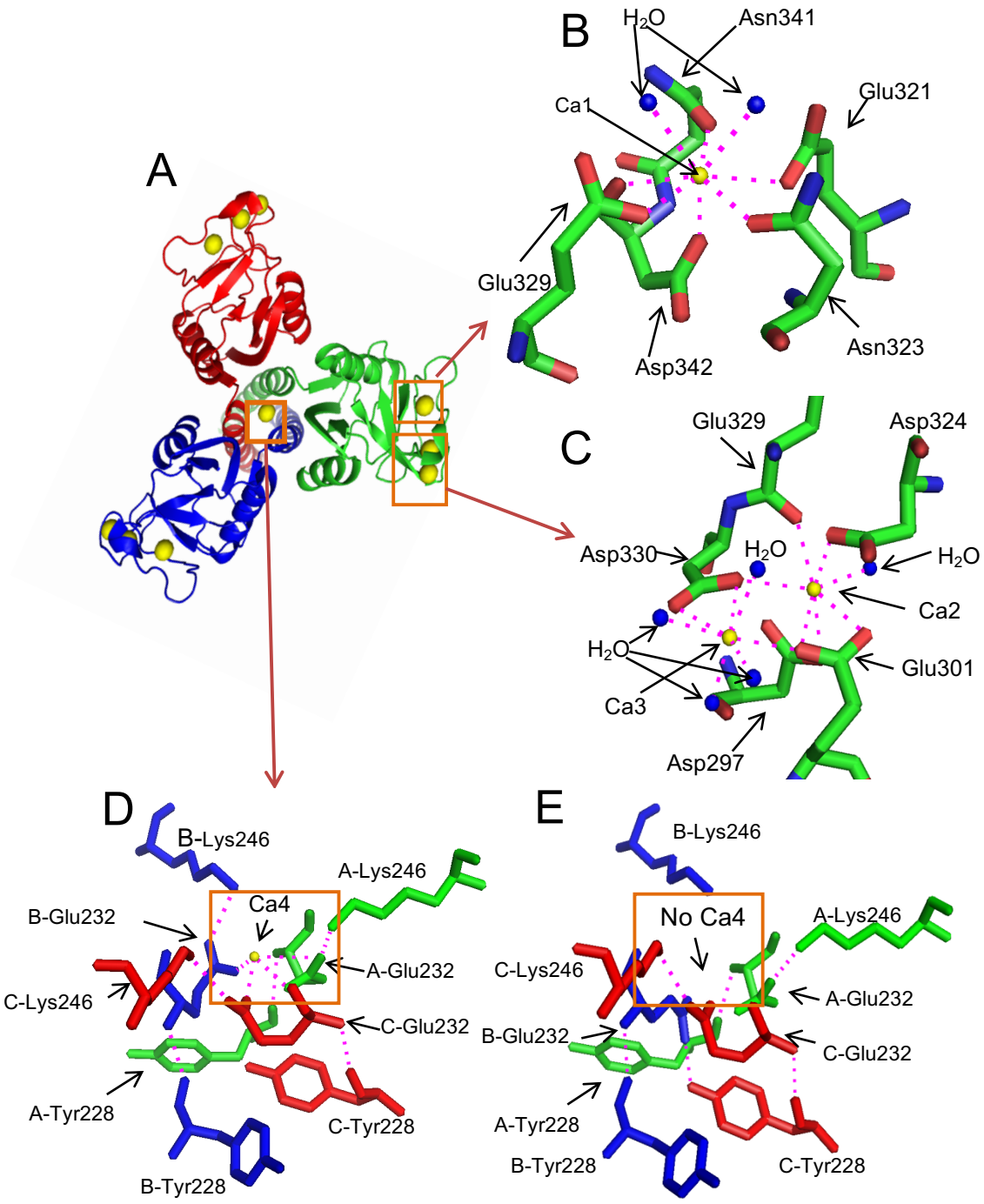


Fig 1.11 A- Ligand-free structure of rfhSP-D shown from the top with Ca1, Ca2, Ca3, Ca4 and the binding pocket shown. **B-** close-up of the binding pocket showing the organisation of the five key residues Glu321, Asn323, Glu329, Asn341 and Asp342, showing the six coordination bonds between Ca1 (orange circle) and the protein binding pocket residues. In the ligand-free state the rfhSP-D Ca1 last two coordination bonds are completed by two water molecules (blue spheres). **C-** Close-up Ca2 and Ca3 and coordinating residues and water (H_2O) molecules. **D-** Close-up of Ca4 at the neck-CRD interface. Ca4 (orange circle) is coordinated by three symmetrical Glu232 and three water molecules (not present in the image). The orange box shows the orientation of the Glu232 in subunit B (in blue). **E-** a representation of the loss of coordination between Ca4 (not present) and Glu232 in subunit B (in blue). This causes the Glu232 in subunit B to move deeper into the neck, forming a hydrogen bond with the asymmetric Tyr228 in subunit C. Images **A, B** and **C** were generated by *PYMOL* using PDB: 1PW9, while image **D** was generated using *PYMOL* using PDB: 1PWB.

1.5.2.4 Ligand-bound rfhSP-D structures

The various ligand binding studies reported to date have demonstrated the variety of structural and functional differences in how rfhSP-D recognises the different ligands studied. Short mono-, di- and trisaccharide residues representative of longer natural ligands were used to highlight differences in specific ligand interactions and how key residues in the CRD domain contribute to the ability to recognise the variety of ligands.

rfhSP-D has been reported to bind a variety of ligands including maltose (Glc-α1,4-Glc), maltotriose (Glc-α1,4-Glc-α1,4-Glc), P-nitrophenyl-α-D-maltoside, N-acetyl-mannosamine, mannobiose α1,2 and α1,4, α-Methyl-mannobiose, galactose, inositol phosphate (IP), L-glycero-D-manno-heptose, di-L-glycero-D-manno-heptose (α1,3), D-glycero-D-manno-heptose and to *Haemophilus influenzae* Eagan 4A PS (through the highly conserved L-glycero-D-manno-heptose), see **Table 1.9** for a direct comparison of how (which hydroxyls) rfhSP-D has been reported to bind to the various ligands.

The first ligand-bound rfhSP-D structure used maltose (α1-4 glycosidic bond between Glc1-Glc2) with the maltose bound structure suggesting minimal conformational change upon binding, with the CRD site ready for the ligand binding. Ca1, Ca2 and Ca3 were reported to be present in all three subunits of the homotrimer. The residues coordinating the Ca1 are Glu321 (OE1), Asn323 (OD1), Glu329 (OE1), Asn341 (OD1), Asp342 (OD1) and the main chain carbonyl (O) of Asp342, forming six coordination bonds with Ca1. The final coordination was achieved through the hydroxyl groups (OH), positioned equatorially on carbon 3 (O3') and carbon 4 (O4') of the terminal sugar (glc1), replacing the two water molecules reported in the ligand-free rfhSP-D structure (**Fig. 1.12A**). Four residues that are important in coordinating Ca1 were also shown to be important in coordinating the glc1 of the disaccharide; the O3' hydroxyl interacts with Glu321 OE2 and with ND2 of Asn323, while the O4' hydroxyl interacts with OE2 of Glu329 and with ND2 of Asn341 (**Shrive et al., 2003**).

Ligand and **PDB accession code**.	Binding mechanism of rfhSP-D to reported ligands (which hydroxyls does rfhSP-D Ca1 bind to)	Displayed in figure:	Literature reported in:
- Maltose (Glc1-α1,4-Glc2) - **1PWB**	**Binding**- Ca1 binds to O3' and O4' equatorial hydroxyls of Glc1 + the O3' interacts with Glu321 OE2 and Asn323 ND2 and the O4' interacts with Glu321 OE2 and Asn341 ND2.	**1.12A**	**Shrive et al., 2003**
- Maltotriose (Glc-α1,4-Glc-α1,4-Glc = Glc1-Glc2-Glc3) - **2GGU**	**Binding-** Ca1 binds to O3' and O4' equatorial hydroxyls of Glc1 + additional interactions similar to maltose (in terms of O3' and O4' and Ca1 coordinating residues). Glc3 and Phe335 stacked together and suggested to provide additional stability.	**1.12B**	**Crouch et al., 2006**
- P-nitrophenyl-α-D-maltoside - **2GGX**	**Binding-** Ca1 binds to O3' and O4' equatorial hydroxyls of Glc1 + additional interactions similar to maltose.	Similar to **1.12A**	**Crouch et al., 2006**
- N-acetyl-mannosamine - **2ORJ**	**Binding-** Ca1 binds to O3' and O4' equatorial hydroxyls of ManNAc + additional interactions similar to maltose.	Similar to **1.12C**	**Crouch et al., 2007**
Mannobiose α1,2 (Man1-Man2) - **3G83** - **3IKQ**	**Binding-** Ca1 binds to O3'and O4' equatorial hydroxyls of Man1 + additional interactions similar to maltose	**1.12C**	**Shrive et al., 2009; Crouch et al., 2009**
- Mannobiose α1,4 (Man1-Man2) - **3IKR**	**Binding-** O3'and O4' equatorial hydroxyls of Man1 + additional interactions similar to maltose. Chain A and chain C orientation of the O3' and O4' hydroxyls of man1 interchanged by 180°. Chain B displays both orientations.	**1.12D**	**Shrive et al., 2009**
- α-Methyl-mannobiose - **3G81**	**Binding-** O3' and O4' equatorial hydroxyls of mannose. + additional interactions similar to maltose.	Similar to **1.12C**	**Crouch et al., 2009**
- Galactose - **3IKN**	**Binding**- Ca1 binds to O1' and O2' hydroxyls of galactose + the O2' interacts with Glu321 OE2 and Asn323 ND2 and the O1' interacts with Glu321 OE2 and Asn341 ND2.	**1.12E**	**Shrive et al., 2009**
- Inositol Phosphate (I-1-P) - **3IKP**	**Binding**- Ca1 binds to O4' and O5' hydroxyls of IP	**1.12F**	**Shrive et al., 2009**
- L-glycero-D-manno-heptose - **2RIB**	**Binding**- Ca1 binds to O6' and O7' hydroxyls of the glycero group.	Similar to **1.12G** but only with one Hep.	**Wang et al., 2008**
- di-L-glycero-D-manno-heptose (Hep1-α1,3-Hep2) - **2RIC**	**Binding**- Ca1 binds to O6' and O7' hydroxyls of the glycero group of HepI.	**1.12G**	**Wang et al., 2008**
- D-glycero-D-manno-heptose - **2RIA**	**Binding**- Ca1 binds to O3' and O4' hydroxyls of the glycero group of HepI.	**1.12H**	**Wang et al., 2008**
- *Haemophilus influenzae* Eagan 4A PS (L-glycero-D-manno-heptose) - **4E52**	**Binding**- Ca1 binds to O6' and O7' hydroxyls of the glycero group of Hep, even though a Glc is available to bind through O3' and O4'.	**1.12I**	**Clark et al., 2016**

Table 1.9 A direct comparison of the reported ligand binding structural studies of rfhSP-D displaying the differences in how (which hydroxyls) does rfhSP-D Ca1 achieve the final coordination. In bold on the left hand-side the PDB accession code for each of the displayed structures is given.

Other ligands tested have also been shown to be recognised by rfhSP-D through the O3' and O4' equatorial hydroxyls; maltotriose (α1-4 Glc1-Glc2-Glc3) (**Fig. 1.12B**) (**Crouch et al., 2006**), P-nitrophenyl-α-D-maltoside (**Crouch et al., 2006**), N-acetyl mannosamine (**Crouch et al., 2007**), mannobiose α1-2 (α1-2 Man1-Man2) (**Fig. 1.12C**) (**Shrive et al., 2009**; **Crouch et al., 2009**), mannobiose α1-4 (α1-4 Man1-Man2) (**Fig. 1.12D**) (**Shrive et al., 2009**) and α-Methyl-mannobiose (**Crouch et al., 2009**).

In the maltotriose structural study a secondary interaction was suggested, whereby the aromatic ring of Phe335 can act as an extended carbohydrate binding site, by stabilising the binding of CRDs near surfaces through stacking with carbohydrate rings or hydrophobic moieties in the same molecules (**Fig. 1.12B**). The authors reported that the maltotriose had stronger binding than maltose, with the suggested reason for the greater stability being the stacking of Phe335 and glc3 of the maltotriose. This secondary interaction appears to stabilise the ligand in such way, that all three chains were occupied by maltotriose and all showed evidence of orientation for all Glc1-Glc2-Glc3 (**Crouch et al., 2006**).

Mannobioses α1-2 and α1-4 (Man1-Man2) were also used for ligand binding structural studies in complex with rfhSP-D. Mannobiose α1-2 is a highly typical repeating unit of mannan, normally displayed in N-linked glycans in viral targets (**Tate et al., 2014**), while mannobiose α1-4 offered the opportunity to compare how mannobiose and maltose binding compares, as both are α1-4 linked. Both ligands bound to Ca1 in a similar fashion to maltose, through the hydroxyls O3' and O4' of the terminal sugar (man1), with no reported interactions involving the non-terminal sugar (man2) (**Fig. 1.12C** for mannobiose α1-2) (**Shrive et al., 2009**; **Crouch et al., 2009**). In the mannobiose α1-4 structural study two different orientations of the ligand binding were observed. Chain A displayed a similar orientation to mannobiose α1-2 and maltose however, chain C shows a different orientation with the terminal man1 rotated by 180° (**Fig. 1.12D**), interchanging the residues with which the hydroxyls interact; O4' interacts with Glu321 OE2 and Asn323

ND2, and O3' with Glu329 OE2 and Asn341 ND2; in chain B there is evidence for both orientations of the man1 (**Shrive et al., 2009**).

Different recognition mechanisms (how rfhSP-D Ca1 binds the carbohydrates hydroxyls) have been reported for other ligands tested, including galactose, inositol phosphate (I-1-P) and heptose derivatives that differ from those reported for maltose, maltotriose and mannobiose. Galactose has been reported to be recognised through a novel orientation with the O1' and O2' hydroxyls (**Fig. 1.12E**) of the ring structure binding to Ca1. O1' was by coordinated by Glu329 OE2 and Asn341 ND2, while O2' was coordinated by Glu321 OE2 and Asn323 ND2. SP-D has been reported to exhibit low affinity for Galactose which may reflect the novel binding orientation seen in the rfhSP-D galactose complex (**Shrive et al., 2009**). Inositol phosphate (I-1-P), an integral part of phosphatidylinositol (PI) surfactant lipid has been shown to be recognised by rfhSP-D Ca1 through the hydroxyls groups O4' and O5' (**Fig. 1.12F**), with the phosphate group extending away from the binding pocket (**Crouch et al., 2007**; **Shrive et al., 2009**).

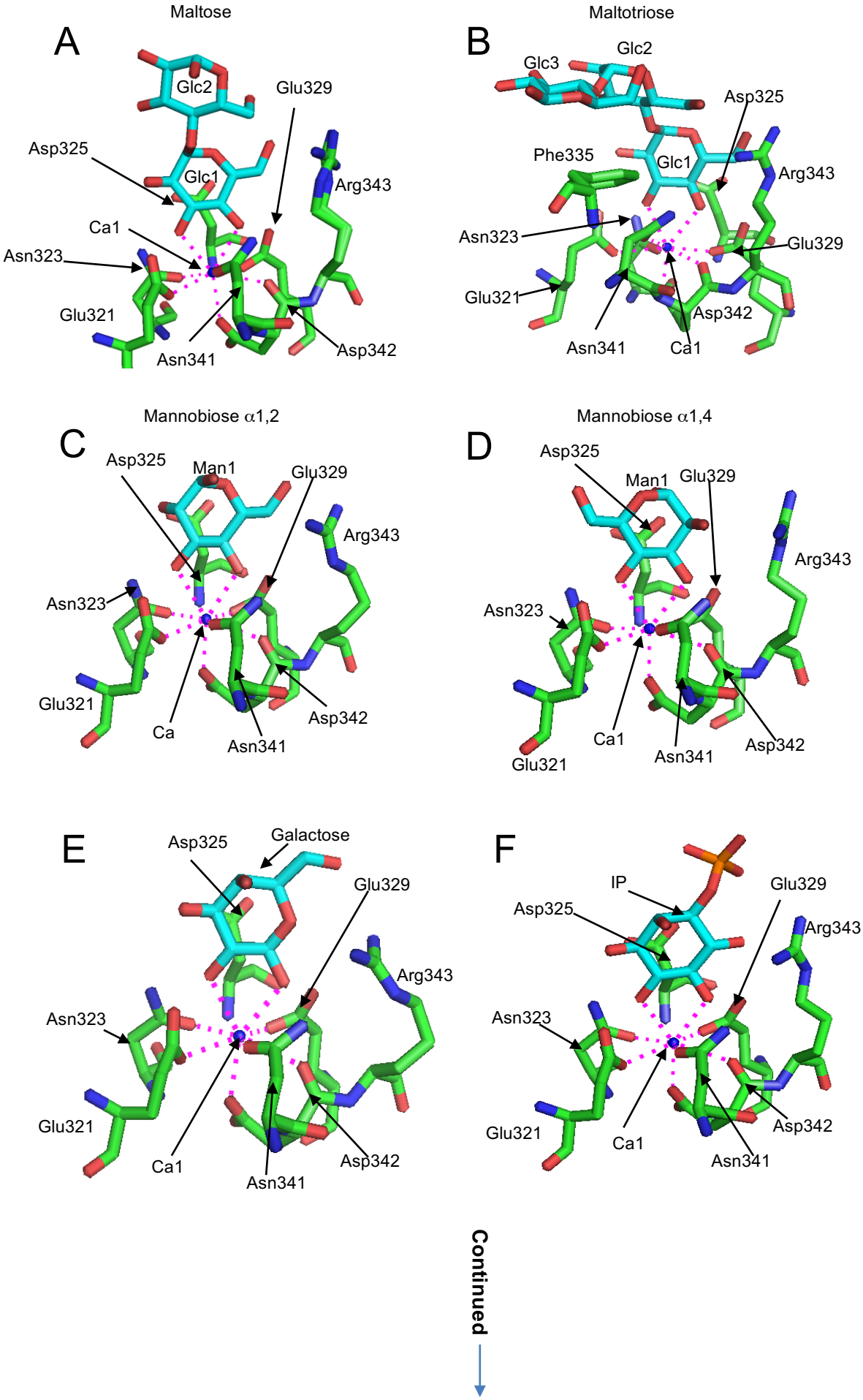
Ligand binding structural studies into the heptose derivatives are of significant biological importance. Heptoses are main constituents of LPS core regions, especially L-glycero-D-manno-heptose. Structural studies of rfhSP-D with a variety of heptoses have been performed, including 2-deoxy-L-glycero-D-manno-heptose, Allyl 7-O-carbamoyl-L-glycero-D-manno-heptopyranoside, L-glycero-D-manno heptopyranosyl-(1-3)-L-glycero-D-manno-heptopyranose, L-glycero-D-manno-heptose, D-glycero-D-manno-heptose (**Wang et al., 2008**) and *Haemophilus influenzae* Eagan 4A PS (**Clark et al., 2016**).

L-glycero-D-manno-heptose bound to rfhSP-D Ca1 via the hydroxyls O6' and O7', with the hydroxyl groups assuming a similar conformation to the equatorial hydroxyl groups O3' and O4' in the maltose structure. This novel orientation had been previously unreported and is unique, as all reported structures show preference for equatorial hydroxyl groups. The authors decided to also investigate a di-heptose α1-3, as this is

representative of Hep1 and Hep2 in inner core LPS. The binding was similar to the heptose, with HepI binding to Ca1 and no further interactions with Hep2 (**Fig. 1.12G**) (**Wang et al., 2008**). D-glycero-D-manno-heptose was shown to bind through the O3' and O4' equatorial hydroxyls (**Fig. 1.12H**), similar to the maltose structure and suggesting that the L-glycero isomer is crucial in achieving the O6' and O7' Ca1 coordination (**Wang et al., 2008**).

The structural study of rfhSP-D and *Haemophilus influenzae* Eagan 4A PS at Keele represents the first natural ligand reported. Structural studies performed previously had used mono, di, and tri-saccharides. The PS of *Haemophilus influenzae* Eagan 4A resembles the natural native state of LPS, with the only difference being the lack of Lipid A (delipidated) and a rearrangement of the Kdo group as a result of the delipidation (**Clark et al., 2016**).

The rfhSP-D in complex with PS from *H. influenzae* Eagan 4A showed the binding to L-glycero-D-manno-heptose, in a similar fashion to that previously reported for mono and di- L-glycero-D- manno- heptose, (**Fig. 1.12I**) (**Wang et al., 2008**). There was no density for the β1-4 linked Glucose attached to the HepI, while HepI-α1-5Kdo showed a clear definition. The electron density for the Kdo suggests a closed 5-membered ring rather than the expected pyranose ring. This supports the findings that mild acid hydrolysis of O4-substituted Kdo causes a conformational change in the KDO. It is interesting to note that rfhSP-D bound to HepI and not the terminal Glc available in the 4A LPS (**Fig. 1.12I**), suggesting a novel interaction preference, where the proximal inner core region (HepI) is preferred over the distal core terminal sugar (Glc) (**Clark et al., 2016**). This however does not reflect on the ability of rfhSP-D to recognise distal core terminal sugars, especially when the inner core is shielded or unavailable.


A
Maltose
Glc2
Glu329
Asp325
Glc1
Arg343
Ca1
Asn323
Glu321
Asn341
Asp342
B
Maltotriose
Glc3
Glc2
Asp325
Phe335
Glc1
Arg343
Asn323
Glu329
Glu321
Asp342
Asn341
Ca1
C
Mannobiose α1,2
Asp325
Man1
Glu329
Arg343
Asn323
Glu321
Ca
Asp342
Asn341
D
Mannobiose α1,4
Asp325
Man1
Glu329
Arg343
Asn323
Ca1
Asn341
Asp342
Glu321
E
Asp325
Galactose
Glu329
Arg343
Asn323
Glu321
Ca1
Asn341
Asp342
F
IP
Glu329
Asp325
Arg343
Asn323
Glu321
Ca1
Asp342
Asn341

Continued

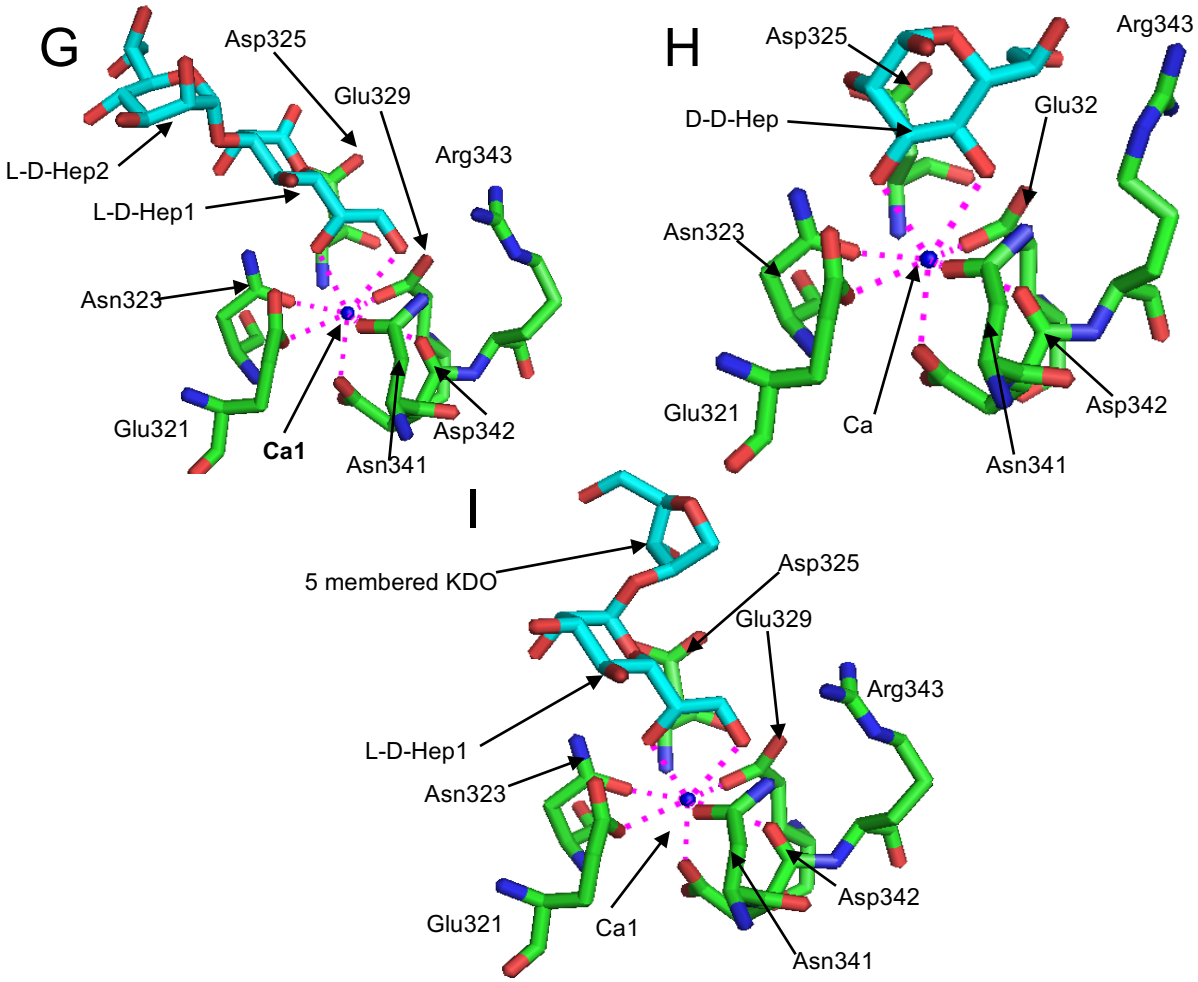


Fig 1.12 Ligand-bound rfhSP-D structures of various ligands tested and reported in the literature (of special interest). All the figures illustrated display similar positioning of the key residues (Arg343 positioned on the right-hand side for direct comparison) and with the same residues illustrated in all the figures; Ca1 coordinating residues - Glu321, Asn323, Glu329, Asn341 and Asp342, with the flanking residues Asp325 and Arg343 also displayed. **Fig. 1.13B** contains an extra residue - Phe335. All rfhSP-D are displayed as sticks with green (carbon), red (oxygen) and blue (nitrogen) colours; Ca1 is displayed as a small blue sphere; ligands are displayed as sticks with light blue (carbon) and red (oxygen) colours; the coordination of Ca1 by the rfhSP-D residues and ligand is displayed as dashed lines (magenta). **A**- Maltose structure (Glc1-α1,4-Glc2) with Glc1 bound by the O3' and O4' equatorial hydroxyls to Ca1 (chain A) (PDB: 1PWB). **B**- Maltotriose structure (Glc-α1,4-Glc-α1,4-Glc = Glc1-Glc2-Glc3) with Glc1 bound by the O3' and O4' equatorial hydroxyls to Ca1 (chain C) (PDB: 2GGU). **C**- Mannobiose α1,2 (Man1-Man2) with Man1 bound by the O3' and O4' equatorial hydroxyls to Ca1 (chain A) (PDB: 3IKQ). **D**- Mannobiose structure α1,4 (Man1-Man2) with Man1 bound by the O3' and O4' equatorial hydroxyls to Ca1 but with the hydroxyls interchanged by 180° (chain C) in comparison to mannobiose α1,2 (PDB: 3IKR). **E-** Galactose structure, with galactose binding through the O1' and O2' equatorial hydroxyls to Ca1 (chain A) (PDB: 3IKN). **F-** Inositol phosphate structure, with the inositol binding through the O4' and O5' hydroxyls (PDB: 3IKP). **G-** di-L-glycero-D-manno-heptose (Hep-α1,3-Hep = Hep1-Hep2) structure, with Hep1 bound by the O6' and O7' hydroxyls to Ca1(chain C) (PDB: 2RIC). **H-** D-glycero-D-manno-heptose, with the D,D-heptose binding through the O3' and O4' equatorial hydroxyls (chain C) (PDB: 2RIA). **I-** *Haemophilus influenzae* Eagan 4A PS binds through the O6' and O7' hydroxyls of the L-glycero-D-manno-heptose (chain C), even though a glucose is available to bind through the O3' and O4' equatorial hydroxyls, suggesting preference of binding for L-D-Hep (PDB: 4E52). Figures generated using *PYMOL*.

Additional interactions between rfhSP-D and bound carbohydrates in excess of the Ca1 coordinating residues (Glu321, Asn323, Glu329 and Asn341) have also been reported. Arg343 and Asp325 have been reported to flank the binding site, providing additional interactions with the ligand. In the maltose structure (**Shrive et al., 2003**) Arg343 NH1 interacts with glc2 O6' and with Man1 O6' of the mannobiose α1-2 structure (**Shrive et al., 2009**; **Crouch et al., 2009**); in the mannobiose α1-4 structure chain C the rotation of the terminal Man1 by 180° (interchanging the hydroxyls interacting with Ca1) causes the C6' previously located near Arg343 to move to the opposite side of the binding site, allowing interactions between Arg343 and O2' and Asp325 OD2 with the O1' of the man1, the latter a previously unreported interaction (**Shrive et al., 2009**); the O5' and O6' of the galactose structure interact with Arg343 NH1 and NH2 respectively (**Shrive et al., 2009**); additional interactions between Hep1 O2' and Asp325 were also reported, suggesting that the flanking residues are important in the recognition of heptose (**Crouch et al., 2006**; **Wang et al., 2008**; **Shrive et al., 2003** and **2009**).

1.5.3 rfhSP-D mutants and the importance of residues 343 and 325

The role of the flanking residues in providing secondary binding potential has led some researchers into exploring the impact that individual amino acids flanking the binding site can have in the recognition of different ligands. These studies have used rfhSP-D mutated by single or double site mutation of site 343 and/or 325 (flanking residues). These sites have been suggested as being important in discriminating glycoconjungates, with the ability to impact on binding affinities.

Single mutant Arg343Val binding properties were investigated in complex with mannobiose α1-2, α1-3 and α1-6 linked and branched tri-mannose cores (**Crouch et al., 2009**). The idea for this mutation arose from the greater ability of a fusion protein containing the SP-D neck and the conglutinin CRD (SP-D/$Cong_{neck+CRD}$) to aggregate, inhibit, and enhance phagocytosis and the oxidative response towards influenza A virus (IAV), than rfhSP-D. Conglutinin possess a Val at position 343 (hSP-D numbering), while

in hSP-D an Arg is present (**Hartshorn et al., 2000**). The authors decided to investigate the impact that Val343 could have on the binding affinity of SP-D towards IAV and created the Arg343Val rfhSP-D mutant. The Arg343Val rfhSP-D structural study showed similar binding to that reported for mannobiose α1-2 through the O3' and O4' hydroxyls, but to the man2, instead of the reported man1 in the mannobiose α1-2 rfhSP-D. Further interactions included hydrogen bonding between O2' and O6' and Arg349 and Glu347 respectively. The study also reported that the mutation Arg343Val resulted in a 2-fold increase in affinity towards mannobiose α1-2 in comparison to rfhSP-D (**Crouch et al., 2009**). The mutant Arg343Val has also been reported in a computational modelling study to be important in binding to N-acetyl-glucosamine and glucose, with increased binding to both carbohydrates when compared to ligand-free SP-D (**Allen et al., 2004**).

Single mutant rfhSP-D Arg343Lys was produced in order to investigate the impact that a single mutation of Arg to Lys at position 343 could have on a variety of carbohydrate, PI and LPS binding affinities. The mutant Arg343Lys showed increased binding to PI, maltose, glucose and rough LPS (*E. coli* J-5 Rc LPS), in comparison with rfhSP-D. It is interesting to note that the Arg343Val mutation showed little impact on the binding to PI and rough LPS (**Wang et al., 2008**; **Crouch et al., 2007**).

Following on from previous work where mutant rfhSP-D (**Crouch et al., 2009**; **Crouch et al., 2007**; **Allen et al., 2004**; **Wang et al., 2008**) containing a single site substitution impacted on binding affinities for different ligands, **Goh et al., 2013** investigated a double site substitution of rfhSP-D, where Asp325 was mutated to Ala325, while Arg343 was mutated to Val343. This new double mutant was investigated in combination with the N-linked glycan of hemagglutinin (HA), present in the envelope of IAV (**Goh et al., 2013**). In this structural study the double mutant rfhSP-D was shown to inhibit IAV in contrast to rfhSP-D which was shown to be ineffective against IAV. The double mutant of rfhSP-D showed similar results to the full length ligand-free SP-D. The double mutant was also shown to be more stable than the rfhSP-D, with stronger binding

reported, as additional hydrogen bonds and hydrophobic interactions with the HA Man-α1,2Man-α1,2Man were seen. This study provided the first structural study of an innate host defence lectin inhibiting a viral glycoprotein target (**Goh et al., 2013**).

In summary, as the evidence presented above suggests, the amino acids flanking the binding pocket at site 325 and 343 can alter the binding affinity for a variety of carbohydrates, making it crucial in the recognition of a variety of PAMPs presented on pathogens. It is interesting to note the potential of using fusion proteins by associating specific structural domains with specific functional properties of different collectins, with the aim of enhancing antimicrobial activities towards specific pathogens.

1.6 Conclusion

The human lung is exposed to continuous physical and environmental challenges, whereby the clearance of pathogenic and commensal bacterium needs to occur by means of an adequate immune response, by minimising inflammation and maintaining normal lung homeostasis. Collectins have the ability to impact on lung homeostasis and act as PRRs, by recognising carbohydrates patterns on the surface of microbes, allergens and necrotic/apoptotic own cells (PAMPs and DAMPs). Collectins share conserved residues in the CRD and a binding pocket that enables pattern recognition of carbohydrates, in a calcium dependent manner, with variations in the sequence of the collectins impacting on the binding affinities for different carbohydrates.

SP-D is the collectin of interest in the experimental work presented here and has been investigated in various studies for its ability in recognising bacteria, viruses, parasites, fungus and allergens and for its enhancement of an immune response by means of agglutination, opsonisation, phagocytosis or direct microbial activity. SP-D has also been investigated due to its dual role in pro and anti-inflammatory responses, dependent on the specific interactions with immune system receptors, in the presence or absence of PAMPs or DAMPs.

Various pathologies have been reported to show abnormal levels of SP-D, suggesting that SP-D levels could be a cause or consequence of disease state. rfhSP-D, a short recombinant fragment of SP-D has been demonstrated to hold similar biological activity as full length SP-D, suggesting that rfhSP-D could be used as a therapy to reduce lung inflammation or for the treatment of bacterial/viral infections. Various structural studies have been reported using rfhSP-D in combination with mono/di/tri-saccharide glycoconjugates normally present as PAMPS on the surface of the various pathogens. These have enhanced the knowledge of the recognition of PAMPS by SP-D, however these have been limited to small glycoconjungates.

LPS is a major component of the outer membrane of gram-negative bacteria that can participate in host-bacterium interactions such as adhesion, colonisation, virulence and symbiosis, enhancing the survival of the bacterium. LPS can also elicit strong immune responses, with the ability to cause septic shock if untreated. The manner in which SP-D interacts with bacterial LPS varies, in part due to the high heterogeneity of the LPS, with smooth (O-antigen) and rough (lacking O-antigen) normally presented together in the bacterial OM, but also due to the ability of bacteria to vary the components of the LPS, enhancing its survival and masking important core carbohydrate ligands targeted by SP-D.

Various studies have investigated the ability of specific hSP-D residues, in particular Arg343 and Asp325 which flank the ligand binding pocket, to impact on the recognition and binding affinities of SP-D towards various carbohydrates, suggesting that mutations in specific residues could result in a higher affinity for specific carbohydrates, and in turn for specific pathogens. Fusion protein studies have also investigated the binding properties of SP-D neck and conglutinin CRD, resulting in increased clearance of IAV. It is interesting to note that a naturally formed heteromeric complex of two separate collectins has recently been reported (**Henriksen et al., 2013**), suggesting that the combination of separate collectins with different binding affinities seems to occur in

nature. Evidence for the therapeutic potential of rfhSP-D is increasing, with the work performed in this thesis focused on adding a further mini-chapter of information.

The studies presented here follow the work of Professor *Trevor Greenhough* and Dr. *Annette Shrive*, and fellow and past members of our research group including, *Ian Burns*, *Carrie Smallcombe*, *Amy Shaw* and *Jamie Littlejohn*.

rfhSP-D has been previously investigated in combination with a variety of simple carbohydrate ligands, resulting in vast knowledge of how SP-D interacts with a variety of ligands. The use of natural ligands represents the next step in the elucidation of how SP-D interacts with a variety of the pathogens. The structure of rfhSP-D bound to *H. influenzae* Eagan 4A LPS has previously been investigated in our research group (**Clark et al., 2016**).

In the work presented here, rfhSP-D has been investigated in complex with both intact and delipidated LPS from *H. influenzae* Eagan CA7 rough and *S. enterica* (serotype minnesota) R5 and R7 rough mutants. Other LPS were also investigated; *E. coli* J-5 rough LPS ligand was data collected to a low resolution, providing little evidence to suggest correct ligand orientation; *S. enterica* minnesota R4 rough mutant LPS and *H. influenzae* Wild-type Eagan LPS produced no crystal growth and therefore these will not be discussed in this thesis, and instead the experimental findings reported will concentrate on the *H. influenzae* Eagan CA7 PS and *S. enterica* minnesota R5 PS and R7 LPS.

The aims of this work were to investigate further the binding mechanisms of rfhSP-D with LPS and natural ligands by means of X-ray diffraction, using co-crystallisation as the main tool of crystal formation, with the aim of increased understanding of the interactions between SP-D and LPS by determination of the detailed binding modes and interactions between hSP-D and bacterial LPS. As more structural data becomes available, showcasing the characterisation of the binding of longer ligands by rfhSP-D, the

more evidence is presented for the medical potential of rfhSP-D, which may one day be exploited for the benefit of human health.

Chapter 2 - Experimental Techniques

An introduction to the experimental techniques used in the experimental work in this thesis is given below, highlighting the expression and purification of rfhSP-D, the preparation of LPS and PS and the methods of crystal formation, data collection, data analysis and structure solution.

2.1 Crystals

A crystal is a solid containing an internal structure, where the atoms are arranged in a lattice constructed of identical unit cells. Distortions caused by lattice imperfection usually adversely affect how well the crystals diffract (X-ray scattering). Therefore, it is important to achieve a near perfect crystal, where virtually identical unit cells offer well-defined, long-range, three-dimensional molecular order.

2.1.1 Unit Cell

The unit cell is the smallest repeating structure (group of atoms) which has the overall symmetry of a crystal, and from which the entire lattice can be built by repetition in three dimensions (3D). By representing each unit cell by points, a crystalline lattice can be created, where each point in the lattice is in the same environment as every other point. Unit cells are described by three crystallographic distance axes a, b and c and three inter-axial angles α, β and γ and are defined by space lattices (3D), which serve to display the symmetry of the lattice in a convenient manner (**Fig. 2.1**).

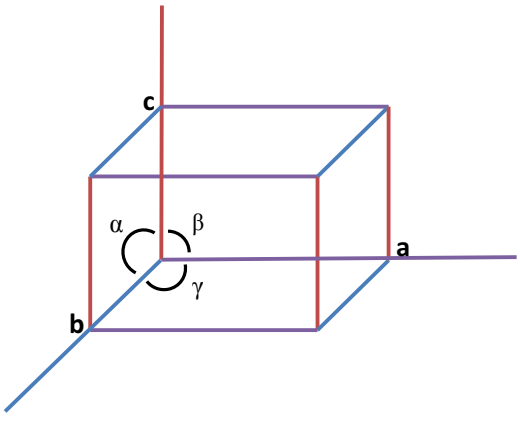


Fig. 2.1 A representation of a typical unit cell. Three crystallographic distances are expressed by a, b and c, with the three inter-axial angles α, β and γ. Together the unit cell parameters and angles can be used to define the unit cell in a space lattice.

The planes of atoms within the unit cell can be thought as sources of diffraction and can be assigned a set of three numbers called lattice indices or Miller indices hkl which are based on the intercepts of the planes with the unit cell axes. For example, a plane that passes through a=a, b=b, c=c is hkl 1,1,1, while a plane which is parallel to the ac face of the unit cell which intercepts b at b/2 is hkl 0,2,0. Within the crystal there are many planes, each part of a family of parallel planes with each plane separated from the adjacent member of the family by a distance d_{hkl} determined by hkl and the unit cell size.

2.1.2 Crystal Symmetry and Bravais Lattices

The molecules in the crystal can be related by different symmetry operations. A symmetry operation can be defined as a set of operations on an object that makes the resulting object indistinguishable from the original except in terms of rotation, translation and reflection. Different types of symmetry are possible, including rotation (with or without translation), reflection and inversion. Reflection and inversion can only occur in crystals where the molecules can exist in both chiral forms of L and R. In biological macromolecules such as proteins, the amino acids are always L-handed, limiting the symmetry operations to rotation and translation only.

The repeated unit cell gives rise to a 3D periodic lattice in one of 14 possible arrangements termed Bravais Lattices. These falls into 7 crystal systems depending on the symmetry and relative dimensions of the unit cell (**Table. 2.1**).

Crystal System	Lattice Type	a,b,c and α, β γ restrictions
Triclinic	P	-
Monoclinic	P, C	α and γ = 90˚
Orthorhombic	P, C, I, F	α, β, γ = 90 ˚
Tetragonal	P, I	α, β, γ = 90 ˚ a=b
Cubic	P, F, I	α, β γ = 90 ˚; a=b=c
Trigonal (Rhombohedral)	R	α, β γ < 120 ˚; a=b=c
Hexagonal	P	γ =120 ˚ α and β = 90 ˚; a=c

Table 2.1 The 14 Bravais lattices (crystal system + lattice type) including a, b, c and α, β γ restrictions for each crystal system and lattice type. P indicates Primitive while C, I and F are unit cells that contain additional lattice points at the centre of one face (C), all faces (F) and at the centre of the unit cell (I). R indicates rhombohedral.

2.1.3 X-rays

X-rays were initially discovered in 1895 by Wilhelm Conrad Rontgen and are a form of electromagnetic radiation with wavelengths of around 0.01-10 nm (0.1-100 Å). X-rays have the ability to penetrate matter and produce relatively strong scattering. In 1912 in Germany Max von Laue discovered that crystals could be used as a diffraction grating for X-rays. Max Von Laue, Paul Knipping and Walter Friedrich directed a beam of X-rays through a copper sulfate crystal and recorded the diffraction on a photographic plate. The photographic plate showed a large number of well-defined spots arranged in a pattern of intersecting circles, around the spot produced by the central beam. Confirmation of this discovery came from England by William Henry Bragg and William Lawrence Bragg, confirming the findings of Max von Laue and colleagues. In 1913 the Braggs proposed an equation (known as Braggs law) which connected the observed scattering with reflections from evenly spaced planes within the crystal. Max von Laue received the Nobel Prize for his discovery of the diffraction of X-rays by crystals in 1914 and in 1915 the Braggs received the Nobel Prize for their services in the analysis of crystal structures by means of X-rays (**Eckert, 2012**).

The wavelength of the X-rays spectrum makes them ideal to measure the inter-atomic bond length, such as C-C bonds with 1.5 Å bond lengths. X-rays interact with the electrons surrounding the atoms in the crystals, making them oscillate and causing the X-rays to diffract. X-ray diffraction is the product of electromagnetic elastic scattering without a change of wavelength and energy, meaning that the outgoing X-rays have the same energy (and wavelength) as the incoming X-rays, with the only difference being the direction of propagation of the electromagnetic wave.

2.1.4 Diffraction by Crystals

Diffraction from crystals creates X-ray “reflections” from the planes of the crystal and uses the concept that X-rays are scattered from the electrons surrounding the atoms

in the crystal in an ordered fashion. Individual molecules scatter X-rays weakly, making them hard to be detected, while the crystals can act as amplifiers of diffraction due to their repeated unit cells (ordered molecules), where each molecule can diffract in similar manner to each other, augmenting the diffraction of X-rays. The amplification of the intensities is proportional to the number of unit cells present in the crystal.

The reflections produced from sets of planes in the crystals can be described by using Bragg's Law. Bragg's Law shows that a set of parallel planes with index hkl and interplanar spacing d_{hkl} produces a diffracted beam when X-rays of wavelength λ, impact upon the planes at an angle θ and are reflected at the same angle if $2d \sin(\theta) = n\lambda$. This means that the rays reflected from successive planes in a family of parallel planes emerge from the crystal in phase, interfering constructively to produce a strong diffracted beam. For other angles where Bragg's Law is not satisfied the waves emerging from successive planes are out phase (they have not travelled distances which differ by a whole number (n) of wavelengths λ), so they interfere destructively and no "reflected" beam emerges at that angle. A representation of Braggs's law is given in **Fig. 2.2**. Large angles of reflection θ correspond to small interplanar distances d with the largest angle observable on the diffraction pattern corresponding to the minimum d that can be resolved. This is called the resolution limit of the crystal and is related to the level of structural detail which can be determined from the crystal.

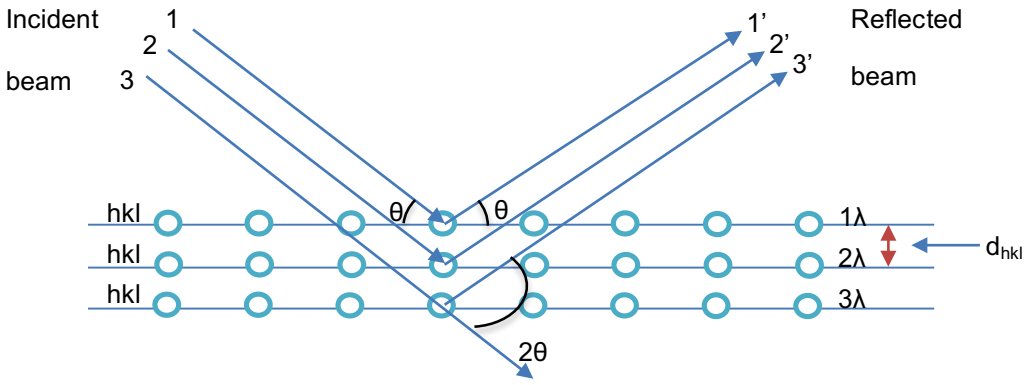


Fig. 2.2 A representation of Bragg's law. Reflected beams are reflected at the same angle θ as the incident beam, from the lattice planes of a crystal, if the condition $2d \sin(\theta) = n\lambda$ is achieved, leading to the successive lattice planes in a crystal producing an in phase reflected beam.

2.1.5 Data Collections and diffraction images

Data collection for the purpose of structural determination is the consequence of two separate processes, the production of high quality crystals suitable for study, and the optimisation of different factors at the X-ray beamline, with the aim of providing detailed structural information about the protein being investigated. Data collection consists of collecting a complete set of reflection intensities (I_{hkl}) with the highest resolution possible (**Fig. 2.3**).

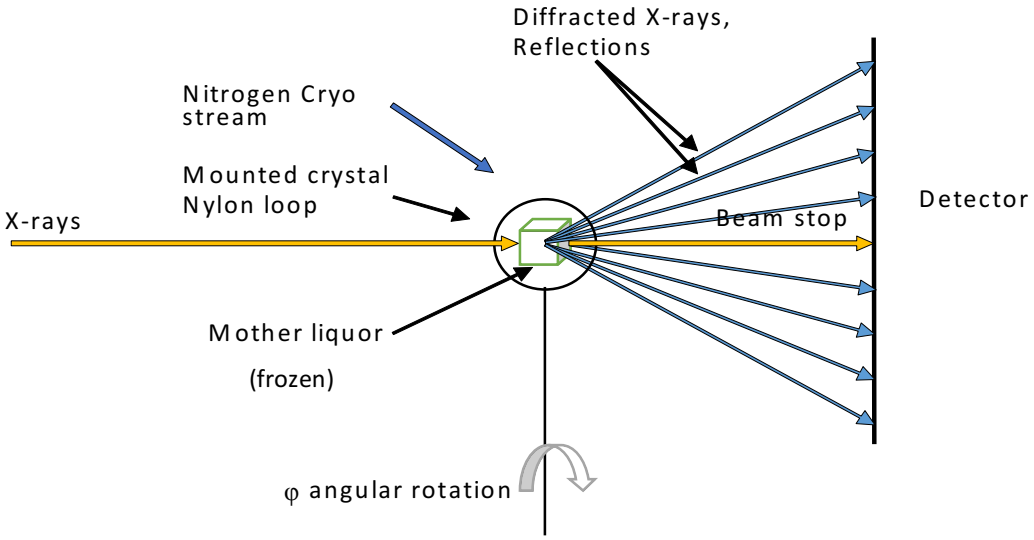


Fig. 2.3 Typical data collection set up performed at the Diamond Light Source (DLS). The crystal is mounted into a goniometer which is rotated by φ ° in relation to the incoming X-rays. The diffracted reflections from the crystal are collected on a detector. The flash frozen crystal is kept at cryogenic temperatures by a stream of nitrogen gas.

Data collection starts when the crystal is mounted on the goniometer and rotated by small angle (0.1-2°), often referred to as φ (phi), with respect to the incoming X-ray beam, with the total angular range varying between 90-180°, depending on the unit cell symmetry of the crystal (**Xuong et al., 1968**; **Leslie, 2006**). The scattered diffraction is collected by the detector, with the aim of collecting accurate measurements of I_{hkl} and standard uncertainties (σI) for all hkl. The detector must have a large active area and good spatial resolving power to be able to measure a large number of closely spaced diffraction spots. The crystal to detector distance can be varied so that the diffraction pattern fills the active area of the detector. The crystal in the goniometer uses a cryo-

stream, which blows a stream of cooled nitrogen gas over the frozen crystal to maintain its temperature.

The detector could be a charge-coupled device (CCD), which provides a rapid readout time and converts the incident photons into electrical charge, forming a 2-D array, with the charge proportional to the intensity of the incident light (**Smyth & Martin, 2000**); or the new generation of X-ray pixel detectors such as PILATUS 6M (specifically dedicated for protein crystallography) a single-photon counting digital detector (**Taguchi et al., 2008**). Each reflection can be used to obtain the direction in which that particular beam was diffracted by the crystal, with the computer constructing the diffraction pattern (diffraction images) using the intensity and direction for each individual diffracted beam that produces a reflection on the detector. Diffraction images (**Fig. 2.4**) consist of images covered in spots (Bragg reflections), with the low resolution spots at the center normally having stronger intensities than those towards the edge of the detector which contain the high resolution information. At the center of the image a blank area is present, where a beam-stop is used, normally made of lead or tungsten, which prevents the beam of undiffracted X-rays burning a hole in the detector, see **Fig. 2.4** for the representation of an X-diffraction image. Each and every spot on the diffracted images contains structural information about each and every atom in the crystal – where it is and how many electrons there are contributing to each diffraction maximum.

Diffraction by crystals can be affected by crystal lattice imperfections (mosaicity), resulting in poor spot shape (overlapping and smearing of spots) and limited resolution. Protein crystals suffer from a mosaic spread, as there is a variety of flexible residues held together by weak forces. Zero mosaicity, where the unit cells are perfectly aligned would create reflections of high intensity, not measurable, suggesting that some mosaicity is essential. In the real world, the crystal is composed of blocks of unit cells slightly misaligned (crystal packing defects). Crystals with high mosaicity usually diffract to low resolution (**Evans et al., 2011**; **Leslie, 2006**). A variety of variables can impact on the

crystal order such as the molecular weight of the protein, solvent content, unit-cell volume, sample-mounting method and sample cooling method.

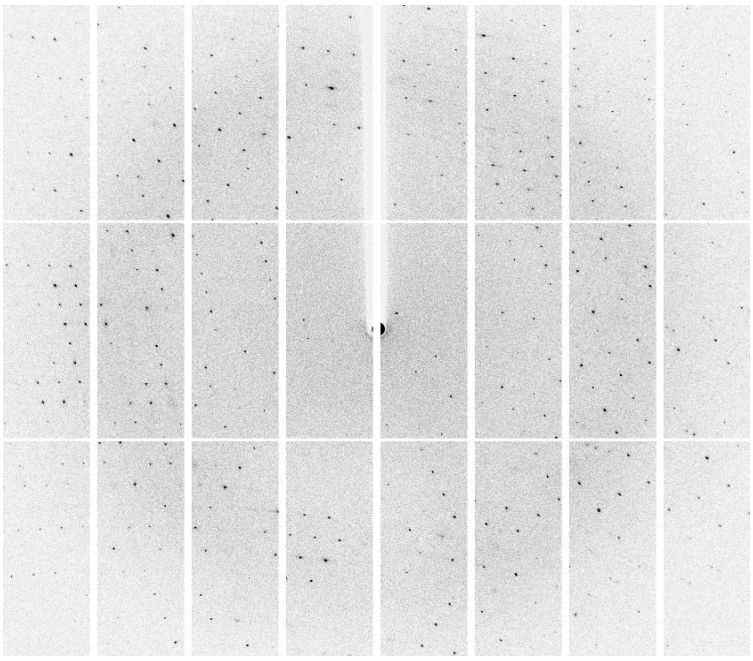


Fig. 2.4 A typical image of an X-ray diffraction snap of a protein crystal on a CCD detector. At the center of the image the beam-stop is visible, with the Bragg reflections (the spots) surrounding the center of the beam. The spots in the center of the image are of higher intensity, while the spots towards the edge represent higher resolution. This image was taken on 30/1/15 at DLS beamline I04-1 from the rfhSP-D crystal RS46 B61. Maximal resolution is 2.6 Å.

2.2 Fourier transform and Structure Factors

A Fourier transform (direct) is the lens simulating operation that a computer performs, where the Fourier transform of the structure is the amplitudes and phases. As a result, the reverse is also true, the reverse Fourier transform of the amplitudes and phases of the diffracted waves gives rise to the generation of the electron density map, providing the link between the diffracted spots that reach the detector and the location and nature of the diffracting atom/electron in the crystal. The direct and reverse Fourier transforms allows for the conversion between a Fourier sum description of the reflections to a Fourier sum of the electron density map (ie structure) and vice-versa.

The reflections that are detected/collected during data collection contain information (in the amplitude and phase) about each and every atom in the crystal (where

it is, i.e. the structure) and these can be described by a Structure Factor ($\mathbf{F}_{hkl}$) equation containing one term for each atom in the unit cell. The electron density in turn can be described by a Fourier sum in which each term is a Structure Factor.

Extrapolating the information that reaches the detector (diffraction pattern), into the position of atoms/electrons (electron density) in the crystal by means of reserve Fourier transform, requires that the angle of diffraction, the amplitude of the diffracted waves and the phases are known. The angle of diffraction is calculated from the detector; the amplitude of the waves is directly related to the intensities (amplitude of $\mathbf{F}_{hkl}$ is proportional to the square root of the reflection intensity I_{hkl}) and is expressed in terms of the structure factor amplitude $|F_{hkl}|$, describing a specific reflection in the diffraction pattern. The other part of the structure factor for each hkl, the phase, will have to be estimated and calculated at a later stage, as this cannot be measured from the diffraction pattern.

2.3 Data processing and Model building

Data processing of the rfhSP-D data sets was performed by using the Collaborative Computational Project number 4 (CCP4) integrated suite of programs that allows the determination of macromolecular structures from X-ray crystallographic data. Various programs of the CCP4 integrated suite were used during the data processing including *iMOSFLM, AIMLESS, TRUNCATE, UNIQUIFY, RE-INDEX* (at times), *REFMAC5* and *COOT* (**Winn et al., 2010**; **Emsley & Cowtan, 2004**).

2.3.1 Data Integration

The data collection stage provides the information required to calculate the reflection intensities I_{hkl} and standard uncertainties (σI) for all hkl, the integration of the different images containing reflections specific to different angles is achieved through three different stages. Stage 1 - the indexing of the diffraction spots determines the unit cell parameters, using two separate diffraction images normally 90˚ apart in φ. Stage 2- the refinement of the initial unit cell parameters and the estimate of the crystal mosaicity.

Stage 3- prediction of the positions of the Bragg reflections on all the diffractions images followed by integration of all the images, generating I_{hkl} and standard uncertainties (σI) for all hkl.

The three stages are completed using the *iMOSFLM* and *AIMLESS* computer programs within the CCP4 suite. During the indexing stage, two diffraction images 90° apart in φ are used in aiding *iMOSFLM* to generate a variety of possible unit cells and parameters with different penalty points, with the unit cell and space group of the lowest penalty normally preferred, unless the space group has been previously determined. The penalty is based on how well spot positions predicted from the unit cell and crystal orientation match those observed on the images. The chosen space group and parameters are then superimposed to fit the observed diffraction spots, miller indices hkl are assigned and the spot intensities are then measured. The merging of the intensities based on symmetry and direct equivalents (multiple measurements) is done by the *AIMLESS* program (**Evans & Murshudov, 2013**; **Leslie, 2006**). At this point we have all the Intensities I_{hkl} and standard uncertainties (σI) for all hkl.

2.3.2 *TRUNCATE, UNIQUIFY* and *RE-INDEX*

The program *TRUNCATE* generates the structure factor amplitudes from the reflection intensity I_{hkl}. *TRUNCATE* is also important in assessing any twinning in the crystals tested, by performing twinning tests. Twinning is the result of two separate crystals, growing in the same crystal lattice, leading to the asymmetric units being added to the growing crystal that do not follow a symmetry operation of the true crystal space group.

The CCP4 *UNIQUIFY* program allows the generation of a unique list of reflections for a given unit cell with a given symmetry and has the ability to set Free-R- flag for each reflection and give indications on the completeness of the data set. The *RE-INDEX* program allows the reindexing of hkl, symmetry and space group.

2.3.3 Phase determination – Molecular replacement

X-rays are electromagnetic waves containing an amplitude and phase. The final component missing in the production of the electron density map is the phase determination. When collecting the diffraction of the scattering X-rays we count the photons (amplitude), without any information of the phases being recorded. The determination of the phase is therefore crucial, and can be achieved by indirect methods or by using phases calculated initially from similar solved structures. The phase of each diffracted ray can be calculated using methods such as molecular replacement, isomorphous replacement and anomalous scattering either by using multiple or single (MAD and SAD) wavelength (**Taylor, 2010**). The experimental work performed in this thesis was based on the structural studies of natural ligands in complex with rfhSP-D. The phases were determined by using ligand-free models of rfhSP-D previously reported (**Shrive et al., 2003**) and by means of molecular replacement, as the rfhSP-D structures are all isomorphous.

2.3.4 Electron density maps

The main objective in X-ray crystallography is to reveal the distribution of electrons within the asymmetric unit such that the features of the molecule in question are revealed. The electron density map is a three-dimensional map of the crystal structure, determined from the scattering of X-rays from the electron clouds of atoms in the crystal lattice. This describes the contents of the unit cell averaged over the whole crystal and not the single unit cell. This is important especially when we consider crystal disorder and how it impacts on the data resolution.

The Fourier (electron density) maps produced by means of molecular replacement (for phase determination) can be divided into two types of map both of which use the best estimate of the phases available at the time. Direct maps are calculated from measured amplitudes Fobs, usually (and throughout this work) combined with calculated structure

factor amplitudes Fcalc, initially using phases derived from molecular replacement. Difference maps use the difference between Fobs and Fcalc which reveals errors in the structure. The difference map uses Fobs - Fcalc (+ phases) and has the ability to show positive density in the map where more electrons are present than have been modelled, i.e. suggesting an additional molecule or ligand present, and negative density suggesting that the model should have fewer electrons, i.e. specific atoms in the model are placed in the wrong site (**Wlodawer et al., 2008**; **Terwilliger et al., 2007**). The Direct maps used here were 2Fobs - Fcalc (+ phases) which combines the Fobs map with a difference map, thus showing both the structure and errors in the structure, see **Fig. 2.5**.

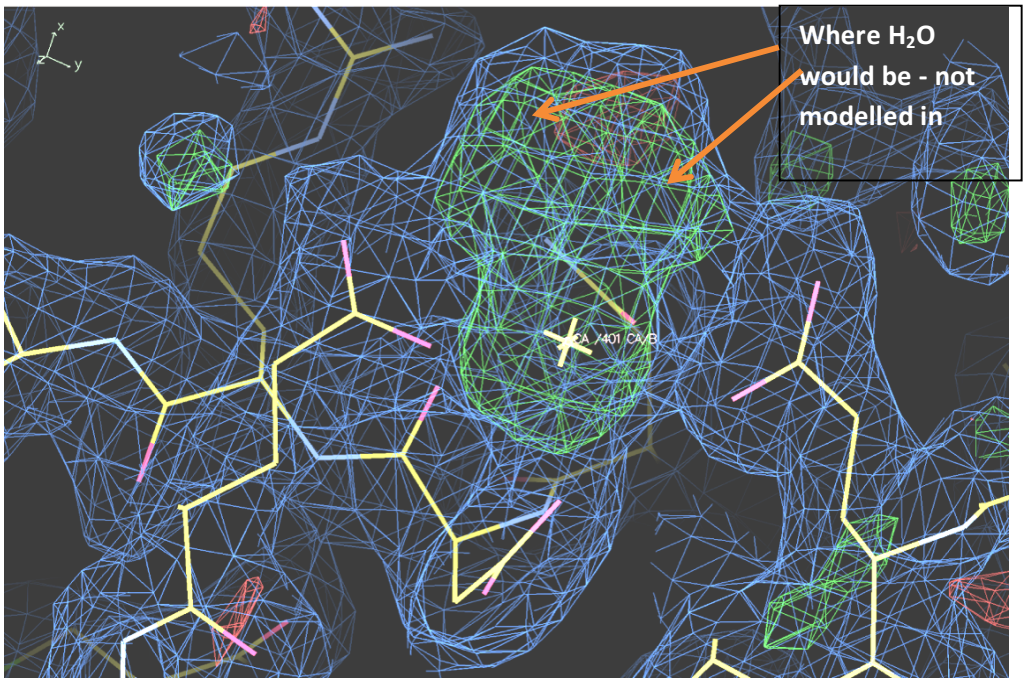


Fig. 2.5 Electron density map of the rfhSP-D ligand-free binding pocket. Two maps are visible in the image. The direct map (2Fobs-Fcalc) in blue and visualised at 1σ surrounds the amino acids and is produced from measured amplitudes Fobs and estimated phases derived from molecular replacement. The (Fobs – Fcalc) difference map can be seen in the image by the green and red contours visualised at 3σ. The green contour suggests that more electrons are present in the data than in the model, in this case two water molecules (not modelled) would sit as indicated. The red contour suggests that the model, should have less electrons, such as an amino acid or amino group is positioned in the wrong site in the map. This map was produced by using the crystal RS46 B63 at 2.02 Å, collected on the 16/1/15 on in DLS beamline I04-1.

The contour of the electron density map produced can be varied, but often the direct map is normally visualised at 1σ (1 standard deviation), while the difference map is normally visualised at 3σ (3 standard deviations), where σ represents the rmsd of all map points from the average value; see **Fig. 2.5** for the representation of a typical electron density map, showing the direct and difference maps. The level of contour is an important feature in the electron density maps and it is important not to use lower contour levels as these might be misleading and emphasise noise and not real features of the protein (**Wlodawer et al., 2008**; **Terwilliger et al., 2007**).

2.3.5 *REFMAC5*

REFMAC5 is a macromolecular refinement program. *REFMAC5* allows models to be refined by means of rigid body refinement, where a similar known model structure is used together with the measured structure factors Fobs. It also provides full crystallographic refinement of the solved structure against the experimental data. The rigid body refinement rotates and translates the known model structure to match the diffraction data, and acts as a starting point in the building and refinement of the unknown structure. The molecular replacement target is essentially to find the best fit of the experimental data to a known model in terms of the match between structure factors calculated from the positioned known model and those measured from the diffraction images. In this experimental work the rfhSP-D data sets were fitted against the known rfhSP-D model, the best fit then being used to calculate an initial electron density map.

The use of rigid body refinement as the first stage of data refinement is normal practice in our research group, as most often we investigate ligand bound structures, where limited conformational changes occur between different ligands and rfhSP-D. The second stage of refinement compares the now solved structure with the experimental data, allowing all structural parameters to be varied in a restrained way so as to maintain good geometry throughout the structure.

2.3.6 Structure refinement and model building

As mentioned above, the work presented in this thesis was based on the X-ray diffraction of crystals containing rfhSP-D in complex with natural ligands from gram-negative bacteria, where the main focus was the interaction between the rfhSP-D binding pocket and the different carbohydrates present on natural ligands. The aim of experiments using X-ray crystallography is the production of a model of interest, in this case rfhSP-D interacting with natural ligands, which, when Fourier transformed, explains the diffraction data as accurately as possible. For this we use refinement (*REFMAC5*) and model building programs (*COOT*). The refinement allows for the optimisation of the different parameters in the model, by refining atomic coordinates and atomic displacements (B factors), while the model building program gives the ability to incorporate bound ligands (LPS and PS) and metal sites (Ca^{2+}) and to make corrections to the maps arising from the refinements. Refinement and model building are used together to add parts of the model and to incorporate structural changes that often the refinement program cannot compute. *COOT* was used as the preferred program of choice in the model building and ligand building steps (**Winn et al., 2010**; **Emsley & Cowtan, 2004**). All the ligands used were similar to those in previous reported structures, as lipopolysaccharides share conserved residues. One important aspect when solving the structure and building the ligand into the final model is over-fitting of the ligand into the model. All the data presented in this thesis is of low resolution (>3Å) and this causes two problems; the risk of over-fitting and the ambiguity of the carbohydrate preference by rfhSP-D. This will be further developed in the results chapters (3 and 4), highlighting proposed reasons as to why the specific crystals did not diffract further.

2.3.7 Crystallographic parameters and indicators of data quality

It is important to understand how different parameters are representative of data quality in X-ray crystallography. Different parameters can be used to assess the quality of the diffraction data and the quality of model (how it was fitted into electron density.

Important parameters include resolution, data completeness, I/σ (signal to noise ratio), multiplicity, R-factors, B-factors, rms deviations (rmsd) of bonds length and angles, and the Ramachandran plot.

2.3.7.1 Data collection parameters and indicators of data quality

Resolution is measured in Angstroms (1 Å = 10^{-10}m) and is the first indicator of data quality, defined as the minimum spacing (d) of the crystal lattice planes that can provide measurable diffraction spots by X-rays. In simple terms, resolution is the level of detail that can be visualised in the electron density maps. The structural detail or resolution of the map can be divided into Low (>3 Å), Medium (3.0 - 2.0 Å), High (2.-0 - 1.2 Å) and Atomic (1.2 - 0.8 Å), where the smallest d spacing is representative of a higher resolution. In ligand bound structures normally a resolution of 2.5 Å or better is required (**Evans et al., 2011**) to identify the binding molecule and remove any ambiguity. Data completeness is the measure of how complete the data is, in terms of I_{hkl} for all unique hkl within the resolution range, or in simpler terms it is the number of X-ray diffraction reflections measured in the data set compared to how many there are. Completeness should be as close to 100% as possible, with an acceptable I/σ. I/σ is a value representative of different variables impacting both the data collection (spatial distortion, radiation damage and detector imperfections, with a higher value being preferred as this represents a higher threshold of measured signal over the averaged noise) and the merged data itself in terms of the number and spread of the different measurements of a reflection contributing to the calculated mean for that reflection. One important aspect in improving I/σ is the use of small φ (angle of rotation) angles, as this improves the I/σ for weaker reflections (**Leslie, 2006**; **Evans et al., 2011**). Multiplicity is calculated as the number of measured reflections divided by the number of unique reflections (measured/unique), with the symmetry in the crystal resulting in the presence of equivalent reflections in the diffraction data (high symmetry crystal yields a high multiplicity). High multiplicity is often the result of multiple measurements of the same

reflection, improving the accuracy of the measured intensities. Residual merge (R_{merge}) and residual symmetry (R_{symm}) are also used as data quality indicators, with the R_{merge} being the measure of the agreement among the multiple measurements of the same reflection and R_{symm} being the measure of agreement between all symmetry related reflections in a data set.

2.3.7.2 Indicators of model quality

Model quality indicators include crystallographic R-factors, consistency with expected geometrical parameters, atomic temperature factors and the Ramachandran plot. The standard crystallographic R-factor or R_{work} is a measure of agreement between the crystallographic model and the experimental X-ray diffraction data in terms of the observed structure factor amplitudes and the calculated structure factor amplitudes, Fobs and Fcalc respectively, with a low R-factor suggesting a better agreement between data and model. The free R-factor R_{free} is similar to the R-factor but it shows the agreement between a subset of reflections not included in refinement. R_{free} is thus a better indicator of model quality than R_{work}, as it provides a less-biased measure of the refinement process. The temperature factors B for each atom is a measure of how much an atom vibrates/oscillates around its position in a specific model. Unusually large B-factors suggest that an atom may be misplaced or perhaps not detectable in the data. Rmsd is a measure of how the final crystallographic models matches expected bond lengths and bond angles, with rmsd values lower than 0.02 Å for bonds lengths and lower than 4 degrees for bond angles expected in a high quality crystallographic model. The Ramachandran plot is based on the main chain conformational angles in the model, highlighting regions or individual amino acids that differ in terms of expected conformation.

2.3.8 Fast data processing (fast_dp)

When collecting data at the DLS, where a data set (set of diffraction images, where φ is varied) is collected, immediately following data collection the fast_dp, a small

python program that uses XDS (integration software using three-dimensional profile fitting) (**Mueller et al., 2012),** CCP4 and CCBX (computational crystallography toolbox), is used to deliver data processing results quickly, by integrating the diffraction images (measuring intensities) and indexing the diffraction spots (assigning unit cell and hkl according to the position on the detector). Fast_dp is often used to offer an indication of data quality, without the need to index all images in-house. If the data looks promising, the indexing and integration of the diffraction images and spots is then carried out more carefully and robustly in-house.

2.4 rfhSP-D protein preparation

2.4.1 rfhSP-D protein

The molecular weight of the rfhSP-D used in this experimental work was calculated using ProtParam sequence analysis tool from the ExPASY proteomics server (http://www.expasy.org), where the molecular weight of the rfhSP-D trimer was calculated to be 56448 Da, while a single monomer was calculated to be 18816 Da.

2.4.1.1 Protein preparation – Expression

The expression of rfhSP-D used in the experimental work was performed by collaborators; Howard W. Clark, Rose-Marie Mackay, Jens Madsen at the University of Southampton, Department of Child Health, Division of Clinical and Experimental Sciences, Faculty of Science, Faculty of Medicine, Sir Henry Wellcome Laboratories, Southampton General Hospital, Southampton, United Kingdom, according to the method reported in **Singh et al., 2003**. The expression of rfhSP-D was performed in *Escherichia coli,* yielding a rfhSP-D that consisted of a short segment of eight N-terminal Gly-X-Y triplets with substitution of Ser for Pro in position 2 (residue 180), followed by an α-helical coiled-coil neck region (203-235) and the globular CRD region (203-355) (**Singh et al., 2003**).

2.4.1.2 Protein preparation – Purification

The protein used for crystallisation should be of high purity, in order to ensure the formation of homogenous crystals. If purification of the protein is not performed correctly, a heterogeneous solution may either prevent crystal growth or may produce crystals which are not homogeneous in terms of their molecular content (**McPherson & Gavira, 2014**).

In this experimental work two different rfhSP-D expressed batches were used, denominated in our research group as Prep-2 and Batch 6 (both rfhSP-D trimers). The initial steps of purification were performed by the collaborators reported in 2.4.1.1, where Prep-2 was affinity purified and size exclusion chromatography treated in TBS and 5 mM EDTA and Batch 6 was affinity purified only. Upon receiving the different batches of the rfhSP-D from the collaborators, additional steps were taken to ensure that the rfhSP-D solution was suitable for protein crystallography, including buffer exchange and concentrating the protein to suitable concentration.

2.4.1.2.1 Dialysis

Dialysis was performed as the method of choice to buffer exchange the rfhSP-D solutions into 50 mM Tris base, 150 mM NaCl, 10 mM $CaCl_2$ pH 7.4, a favourable solution where rfhSP-D is normally stored (termed protein buffer hereafter). The dialysis technique used Visking dialysis tubing, which allows the diffusion of small molecules through the porous tubing, but not larger molecules such as proteins. Prior to using the Visking dialysis tubing, these were cleaned of any heavy metals. An appropriate length of tubing was placed in 500 ml beaker containing 1 mM EDTA and 2% (≈ 0.24 M) sodium bicarbonate solution and heated to ≈ 80° C for 30 minutes on a hotplate with stirring. After 80 minutes the Visking tubing was removed and rinsed thoroughly inside and out with deionised water. The protein was loaded into the Visking tubing and placed into a beaker of protein buffer containing 100x the sample volume inside the Visking tubing. The buffer was replaced after 12 hours and the dialysis was run for 24 hours at 4° C.

2.4.1.2.2 Protein Concentration

The protein concentrations of the different batches of rfhSP-D were established by using the Nano Drop 1000 Spectrophotometer (Thermo Scientific) that allows for the measurements of the absorbance at 280 nm and calculation of concentration using the ExPASy extinction coefficient for rfhSP-D (P35247). The spectrophotometer method is based on the ability of aromatic rings of amino acid residues such as Tyr, Phe and Trp in proteins to absorb UV-light at 280 nm. The UV-light absorbance at 280 nm can be used to calculate the aromatic amino acid content and total concentration in solution, with the Trp residue being responsible for most of the absorbance of UV-light at 280 nm. Once the extinction coefficient has been established for a specific protein with a known amino acid sequence, the protein concentration in solution can be calculated from its absorbance. Protein concentration is an important variable in achieving crystallisation which is dependent on the supersaturation of the protein in solution, i.e. the protein concentration is higher than its solubility (**Vekilov, 2010**).

2.5 Protein Crystallisation

The first publication reporting the crystallisation of proteins occurred in 1840 by Hunefeld (**Ferry, 1923**). Flat plate-like crystals (haemoglobin crystals) appeared when the blood of an earthworm was pressed between two slides of glass and allowed to dry. These observations by Hunefeld suggested for the first time that proteins could be obtained by slow evaporation or slow dehydration. During the 19th century protein crystallisation was developed as a purification tool and was used as validation of chemical purity. Since then, there have been many advances in the knowledge of how crystallisation occurs. Protein crystallography is used as a technique to obtain single crystals of pure protein for X-ray Crystallography. The first characterised protein using X-ray crystallography was myoglobin in 1958 (**Kendrew et al., 1958**).

The crystallisation of specific proteins usually proceeds by , mixing with precipitant buffers in an aqueous environment. Protein solubility properties are governed by their

three-dimensional folded structures, with the hydrophobic regions tucked inside the protein and hydrophilic regions on the surface. Protein crystals can grow in a drop of aqueous mother liquid containing a variety of reagents which promote self-assembly of the irregularly shaped proteins into an ordered and regular crystal lattice structure. The self-assembly into an ordered state is defined by weak non-covalent intermolecular interactions. Due to the irregular shape of the protein molecules (**Weichenberger et al., 2015**), during the self-assembling of protein molecules into a crystal, approximately 30 – 80% of the crystal volume is occupied by solvent content from the precipitant buffers (**Matthews, 1968**; **Matthews, 1976**). Protein solubility is dependent on ionic strength, pH, temperature and the presence of precipitants such as organic solvents, polymorphs and salts (**Vekilov, 2010**; **McPherson & Gavira, 2014**).

Protein crystals are limited in size, very soft, crush easily, disintegrate if dehydrated, exhibit weak optical properties and diffract X-rays relatively poorly. They are also temperature sensitive and are susceptible to radiation damage. The extent of the diffraction pattern from a crystal, and hence the level of detailed information it contains about the structure, is directly correlated with its degree of internal order.

2.5.1 Nucleation and Growth – Crystal formation

Protein crystal formation can be divided into two separate steps; nucleation and growth, with both steps dependent on the supersaturation (protein in solution is supersaturated) of the mother liquor. Nucleation represents the first stage, where the molecules shift from a disordered state to an ordered one and where crystal nuclei are formed, forming a site upon which other molecules are deposited as the crystal grows. The nucleation rate is dependent on the ability to overcome an energy barrier, with low supersaturation suggesting a slower nucleation rate, while a high supersaturation produces a higher nucleation rate (**Vekilov & Vorontsova, 2014**; **Asherie, 2004**). The growth stage of the crystals represents the second step and is dependent on the crystal nuclei formation. The growth stage can occur by different mechanisms of incorporation of

the new molecules, such as transport of atoms through solution, attachment of atoms to the surface, movement of atoms on the surface and the attachment of atoms to edges and kinks (**Malkin et al., 1995**; **McPherson et al., 2000**). Upon the formation of stable nuclei in solution, the crystal will continue to grow until the system reaches equilibrium. As long as non-equilibrium forces and supersaturation are present in solution, this will drive the crystal growth forward.

Proteins will remain in solution until a critical concentration (supersaturation) is reached. This critical concentration normally termed the supersaturated state is crucial in the crystallisation of proteins. Achieving this state can be explained by a transition of four phases **(Fig. 2.6)**, with phase one in the aqueous state and stages two, three and four in the solid phase.

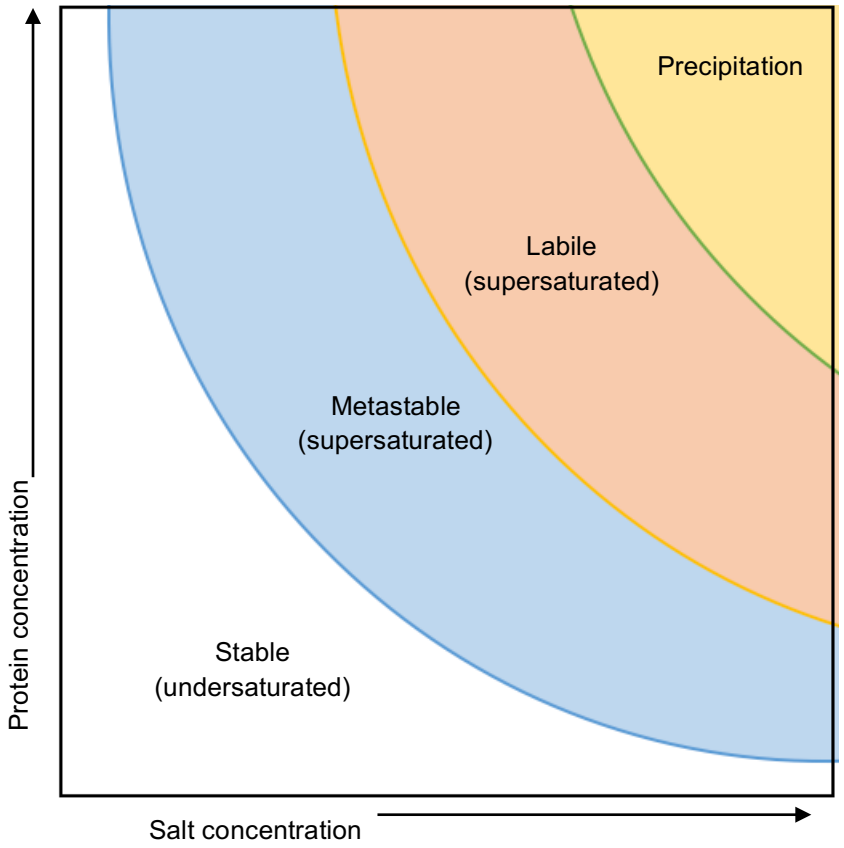


Fig. 2.6 The phase diagram for the crystallisation of macromolecules, displaying the transition of the different regions. In the stable region, the molecules have high solubility; in the metastable region the nucleic develop into crystals; in the labile region both nucleation and growth occurs; the final state is precipitation, which is unfavorable for crystallisation. Image adapted from **McPherson & Gavira, 2014**.

The first phase represents a stable region, where the molecules have high solubility. The second phase, denominated metastable, is where the nuclei develop into crystals. The third phase is known as the labile or supersaturated stage, where both nucleation and growth can occur. The final stage is precipitation, which is unfavourable for crystal formation with precipitates forming faster than the crystals (**Asherie, 2004**). Supersaturation of proteins can be achieved by the modification of an under-saturated solution, where the solubility of the protein is reduced and the attraction between protein molecules in solution is increased. Properties such as pH, salt concentration, buffer concentrations, adding bridging ions and adding ions can be varied with the aim of decreasing protein solubility. Vapour diffusion is a common technique used in achieving supersaturation.

2.5.2 Crystallisation by Vapour Diffusion

Vapour diffusion is a method used in crystallography to produce protein crystals, where variables can be controlled to allow the molecules to change between the aqueous phase and the solid phase, yielding protein crystals. This method has been employed in the production of all rfhSP-D protein crystals used in the experimental work described here. Vapour diffusion relies on the lowered concentration of a solution in the mother liquor (protein + buffer), in comparison to higher concentration in the reservoir (buffer). By osmosis, water molecules will move from the mother liquor, which contains a lower solution concentration, to the reservoir which has a higher solution concentration, “dehydrating” the mother liquor, and raising the concentration of the buffer precipitants and the protein concentration. As water molecules move to the reservoir, protein molecules come closer together in the protein solution reaching a supersaturated state and allowing for nucleation and growth to occur (**Fig. 2.7**).

2.5.2.1 The Sitting-drop Vapour Diffusion Method

The protein crystals used in the structural studies reported here were grown using the sitting drop vapour diffusion method (**Fig. 2.7**). This method consists of using 24 well Linbro plates, where the precipitant buffers are added to the reservoir, with volumes equaling 0.5-1 ml. The precipitant buffers used were a combination of a variety of different components, prior filtered using a SatoriusMinisarts+ cellulose membrane of 0.2 µm. A micro-bridge (Molecular Dimensions) containing a hollow groove is placed within the cell. To the micro-bridge a protein solution aliquot of 1 µl was added with a protein concentration of 7.5-8 mg/ml. From the reservoir one aliquot of 1 µl was added to the protein aliquot producing a 2 µl drop containing 1:1 of protein solution and precipitant buffer. The precipitant buffers used were either structure screens (commercial industrial screens) or were prepared in house. The wells were then sealed with high vacuum grease seal (Dow Corning, USA) and glass cover slips (Menzel-Glaser, Braunschweig, Germany, 22 mm), ensuring that the wells were airtight and preventing evaporation loss from the well. The trays were left to equilibrate at ambient temperature in the cupboards in the laboratory and were observed using a Leica microscope occasionally, noting changes in the solution.

2.5.2.2 Structure/industrial Screens

Structure/industrial screens are based on indications and leads from previously reported buffer precipitants successfully used in protein crystal growth for X-diffraction. These provide hundreds of different mixtures of precipitant buffer with different salts, buffers, precipitant and pHs in an aqueous solution, ready to be used in the initial trials of protein crystallisation. Structure screens are ideal for the identification of suitable conditions that allow specific protein crystal growth. In the case of the experimental work reported in this thesis, for the co-crystallisation of rfhSP-D with a variety of ligands, the suitable conditions had to be identified and the structure screens provided an initial evaluation of potential precipitant buffers. The structure screens used in this experimental

work were purchased from Molecular Dimensions Ltd, Structure Screen 1 – SS1 (MD1-01-CF) and Structure screen 2 –SS2 (MD1-02), offering 100 different crystallization conditions.

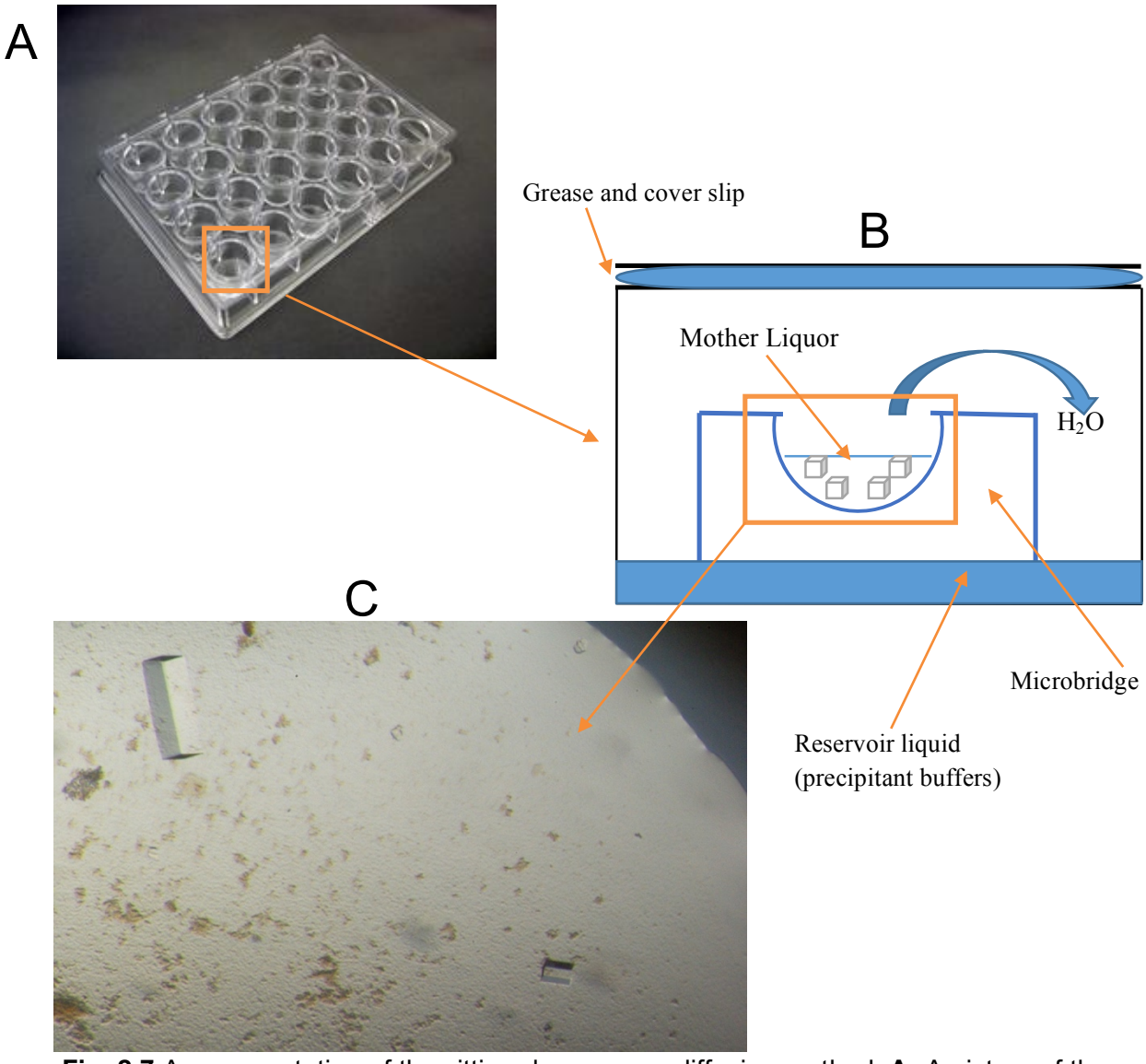


Fig. 2.7 A representation of the sitting-drop vapour diffusion method. **A-** A picture of the 24 well Linbro plate is given. The orange square surrounds one of the wells. **B-** One of the 24 wells. **C-** Representation of successful crystal growth, where crystals are formed in solution.

2.5.2.3 Co-crystallisation vs ligand-soaked crystals

The production of rfhSP-D protein crystals performed in this work differed from previously reported methods. In previous work, rfhSP-D protein crystals were initially produced in various precipitant buffers, with the ligands of interest soaked in to the

crystals at later stages during the cryo-protection stage (**Wang et al., 2008**; **Shrive et al., 2003**; **Shrive et al., 2009**), while in the work presented here a new method attempted to co-crystallise (in the same solution) rfhSP-D in combination with different LPSs. In our research group crystallisation of rfhSP-D ligand-free protein crystals have been a routine practise, where a variety of different molecular weights of polyethylene glycol (Peg) 2000 (2K), 4000 (4K), 6000 (6K), 8000 (8K) and 10000 (10K) and 0.1 M Tris at different pHs ranging between 6 and 8.5, were successful in producing high quality ligand-free rfhSP-D crystals

In the work reported here, an attempt to co-crystallise rfhSP-D in combination with various ligands has been investigated. The co-crystallisation method uses the sitting-drop vapour diffusion method, where the final drop in the microbridge contains a 1:1 ratio of buffer precipitant and rfhSP-D in combination with specific ligands in solution. This means that the nucleation of ligand-bound rfhSP-D molecules will occur surrounded by molecules of the ligand, potentially impacting on the intermolecular contacts and crystal formation. When using the soaking method, if we imagine a ligand-free crystal of rfhSP-D already formed, if we soak ligands that we know can bind to rfhSP-D, this ligand can diffuse through pores in the crystal and hopefully bind to rfhSP-D molecules. Once the ligand is inside the crystal it is constrained to the already formed crystal contacts between the various rfhSP-D molecules. The already formed crystal contacts could impact on the binding of the ligand by rfhSP-D molecules, particularly in the case of longer ligands. The co-crystallisation method could allow for longer ligands to be used, as the crystal contacts are formed with the ligand already bound.

2.5.3 Crystal Preparation and Data Collection

All data collection was performed at the Diamond Light Source, Harwell Science and Innovation Campus, Fermi Ave, Didcot OX11 0QX on the Macromolecular Crystallography beamlines. Different data collections were carried out on specific crystals

of interest, which were treated with cryo-protectant and flash-frozen before diffraction took place.

2.5.3.1 Cryo-protection, Flash Freezing and Radiation Damage

Cryo-protection and flash freezing are two steps employed in the preparation of the crystals for X-ray diffraction and transportation. Both of these steps can greatly reduce radiation damage and reduce the formation of ice crystals, simplifying the handling of the crystals and improving diffraction resolution.

2.5.3.1.1 Cryo-protection

The cryo-protection method employed in this experimental work was based on two separate strategies of cryo-protection, strategy 1 and strategy 2. Cryo-protection is the first step in preparing the crystals for X-ray diffraction. Crystals of interest and suitable for X-ray diffraction were cryo-protected using Glycerol, Polyethylene glycol (Peg 400) and 2-methyl-2,4-pentanediol (MPD) all hydrophilic cryo-protectants.

Specific cryo-protection details for specific crystals will be given in chapters 3 and 4, but in a general manner, strategy 1 was based on sequential additions to the mother liquor of increasing concentrations of the cryo-protectant in a buffer corresponding to the mother liquor and any ligand to be soaked in to the crystal. The cryo-protectant was added at increasing concentrations of 5, 10, 15 and 20%. 2 µl of each concentration (20% twice) was added with 2 minutes' intervals between. The final volume of cryo-protectant added equalled 10µl, which was then exchanged, where 10µl of the mother liquor is removed and replaced with the same volume of the highest concentration (10 µl of 20%). For specific rfhSP-D crystals the gram negative bacterial lipopolysaccharide (LPS) and polysaccharide (PS) were added in combination with the cryo-protectants. Strategy 2 was based on the addition of 0.5 µl (twice) of 50% cryo-protectant. This strategy was initially used due to limited availability of specific ligands as it reduces the volume necessary to

achieve cryo-protection. Both strategies will be discussed in more details in the respective chapters.

2.5.3.1.2 Flash freezing

Cryo-protection allows for the flash freezing of crystals without major impacts on the crystal lattice, freezing the crystal in an amorphous vitreous state. Protein crystals in the mother liquor contain and are surrounded by water molecules. When flash freezing the crystals, if no cryo-protectant is used, ice crystals are formed disrupting the crystal lattice, as water expands when it freezes impacting on the X-ray diffraction and data quality. The crystals of interest were mounted using a free-standing method for mounting crystals, where the crystal is held by surface tension in a nylon loop in a pin, containing the crystallisation precipitants and hydrophilic cryo-protectants (**Teng, 1990**), and submerged in liquid nitrogen (100K) for a rapid freezing mechanism. Flash freezing has the ability to greatly reduce the impact of X-ray radiation upon the crystal lattice, by absorbing up to 70x more radiation than a crystal tested at room temperature, before significant degradation in the crystal order occurs (**Nave & Garman, 2005**).

2.5.3.1.3 Radiation damage

Radiation damage is the consequence of X-ray radiation on the ordered lattice of the crystal, whereby the biological molecules absorb the X-ray radiation. The impact of X-rays on the crystals can cause heating, breakage of disulphide bonds and production of free radicals, which can cascade through the crystal causing damage (**Garman, 2010**).

Before the introduction of cryo-protection and flash freezing, biological crystals were exposed to x-rays at room temperature. The high radiation dose in the crystals meant that multiple crystals needed to be exposed in order to collect sufficient data. The introduction of cryo-protection and flash freezing in 1990 reduced the need for multiple crystals to be exposed and simplified the transportation requirements. The cryogenic temperatures used slow the radical species produced, allowing time for the data to be

collected, before radiation can induce disorder in the crystal lattice (**Garman, 2010**). Another direct benefit is that only one crystal often needs to be used to collect an entire dataset.

In summary, the use of cryo-protectants and flash freezing has contributed to improving the data quality and reducing radiation damage, however some considerations must be employed when using both methods. Cryo-protectant must be optimised as it has the potential to impact on and damage crystal formation and in the worst case scenario to dissolve the crystals; flash-freezing has been reported to have the potential to cause crystal contraction, causing significant change in the molecular packing and perturbation in the crystal lattice (**Juers & Matthews, 2001**).

2.6 Lipopolysaccharides (LPS)

Various bacterial Lipopolysaccharides (LPS) were investigated as natural ligands for ligand bound structural studies with rfhSP-D: *Escherichia coli* J-5 LPS (Sigma-Aldrich-L5014), *Salmonella enterica* (serotype minnesota) R5 (Enzo Lifesciences- ALX-581-017-L002) and R7 (Enzo Lifesciences ALX-581-018-L002) and *Haemophilus influenza*e Eagan CA7 LPS was provided by collaborators; Mary E. Deadman and Derek W. Hood at Mammalian Genetics Unit, MRC Harwell, Harwell Science & Innovation Campus, Oxfordshire, United Kingdom.

2.6.1 LPS in Crystallography

In crystallography the evaporation method (for crystal production) requires a liquid phase, composed of protein solution, buffer precipitant and water. The liquid phase creates a problem when dealing with LPS as its amphipathic nature makes it relatively insoluble in solution. As described in 1.2, the LPS is composed of a hydrophobic lipid A linked to a hydrophilic core and O-antigen (when present). The glycosidic bond linking the lipid A and the rest of the LPS is labile. The method used in this thesis for the delipidation of the LPS into PS was mild acid hydrolysis.

2.6.2 Hydrolysis of LPS

The LPS were hydrolysed and the Lipid A was removed, leaving the PS of interest. The chemical hydrolysis was performed for the *Escherichia coli* J-5 LPS, *Salmonella minnesota* R5 LPS and *Haemophilus influenzae* Eagan CA7 LPS. The hydrolysis method is a chemical reaction that occurs by the action of a mild acid. The hydrolysis of LPS breaks the glycosidic bond holding the lipid A and the remaining of the polysaccharide together, yielding water molecules and two separate fractions that can be separated by centrifugal action (**Morrison & Leive, 1975**). The hydrolysis method was carried out various times for the different LPS, with ≈ 80% hydrolysis yield (expected weight PS/ measured weight PS x100) typically achieved.

2.6.2.1 Hydrolysis Method

Preparation of the PS from LPS is achieved by the hydrolysis method reported in **Masoud et al., 1994** and **Masoud et al., 1997**, where the method of using 2% aqueous acetic acid and refluxing it at 100° C for 2.5 hours is performed. The LPS is a very light and fluffy white powder and care was taken when weighting 10-25 mg. The weighed LPS was then added into a 25 ml round bottom flask, containing a magnetic stirrer. To this 5 ml of 2% acetic acid was added and gently stirred using a magnetic stirrer device. The round bottom flask was then attached to the reflux condenser and placed in a sand bath at 100° C; a thermometer was used to regularly check the temperature. The hydrolysis timing started when the temperature of the sand bath reached 100° C and was left for 2.5 hours. After that time the heating was turned off and the round bottom flask was raised out of the sand bath and allowed to cool. Once the round bottom flask had cooled sufficiently, the water reflux condenser was switched off. The contents of the round flask were then transferred into a 15 ml plastic centrifuge tube and left to cool further in the fridge for 30-60 minutes at 4° C. The centrifuge tubes were then placed in a centrifuge (HERMLEZ400) and centrifuged at the top speed of 2424 g for 30 minutes at 4° C. The top supernatant was carefully removed using an automatic pipette and transferred into another storage

tube, without disturbing the pellet. The pellet was then re-dissolved and re-suspended in 1 ml of deionised water and further centrifuged a further two times as mentioned above. Upon the completion of the second and third spin, 1 ml of each supernatant was removed from the centrifuge tubes and added to the storage tube containing the PS. The storage tube containing the PS was placed in a savant speed vacuum and lyophilised into a dry powder to be weighted out.

2.6.2.2 Acetic Mild Acid Hydrolysis - Impact on LPS structure

The use of mild acid hydrolysis has been suggested to cause a structural and chemical rearrangement of the normal 3-Deoxy-α -D-manno-oct-2-ulosonic acid (Kdo), especially in LPS species with a single Kdo, such as *H. influenzae,* with no reported information regarding LPS with two or more Kdos, such as is the case for *E .coli* and *S. minnesota*. Hydrolysis causes the normal 6-membered Kdo ring to rearrange into a 5-membered ring anhydro Kdo. This novel 5-membered anhydro Kdo has been visualised in X-ray diffraction data (**Clark et al., 2016**) where the best electron density model fit for the Kdo was a 5 membered cyclic 4,7 anhydro Kdo derivative. Mild acid hydrolysis impacts on the phosphate groups present on the heptoses and Kdo group of the inner core of the LPS (**Danan et al., 1982**; **Auzanneau et al., 1991**), with a β-elimination of the phosphate group attached to the Kdo at position 4, leading to the pyranose ring opening and subsequent 4,7 ring closure (**Masoud et al., 1997**; **Clark et al., 2016**). The position 4 of the Kdo in *E. coli* and *S. minnesota* LPS is normally linked to the KDO-II. The ability of acetic acid mild hydrolysis to impact on the structure of LPS during the delipidation stage does create some uncertainty in terms of the physiological validity of the rfhSP-D interactions with the LPS. At present, only a few methods have been reported, such as sonication and hydrogen fluoride, to solubilise LPS without using hydrolysis. Due to the necessity of solubilising LPS in order to visualise the natural interactions between rfhSP-D and LPS, some time was spent investigating and testing potential methods to solubilise the full LPS.

2.6.3 Triton X-100 and Absolute Ethanol

Triton X-100 (Sigma-Aldrich) is a non-ionic detergent composed of an aromatic hydrocarbon (hydrophobic group) and polyethylene oxide chain (hydrophilic group) with a reported non-denaturant nature (**le Marie et al., 2000**), important when co-crystallising the LPS with rfhSP-D. Absolute ethanol is also an amphipathic molecule consisting of an ethyl group linked to a hydroxyl group. Ethanol has been reported to have aided in the solubilisation of smooth LPS (**Sharma et al., 2013**). The potential of both amphipathic molecules was investigated in three separate experiments.

2.6.3.1 The solubilisation mechanism in polar solutions – like dissolves like

The introduction of a non-polar group into a solution containing water molecules disrupts the hydrogen bonding network of the H_2O molecules and increases the entropy (ordered systems have low entropy) in solution (**Ahmed et al., 2016**; **Privalov & Gill, 1989**), resulting in the aggregation of non-polar solutes and limited solubilisation potential. At low concentrations, LPS could in theory be solubilised in solution, forcing the water molecules into creating a “cage” or hydration shell surrounding the hydrophobic lipid A. Fatty acids, which are normally poorly hydrated are limited to a low concentration that can be incorporated in solution. The concentration of hydrophobic solutes and the number of hydration shells and water molecules available for hydration have also been reported to be inversely correlated (**Di Tommaso et al., 2014**; **Hsieh & Wu, 1996**; **Raschke & Levitt, 2005**; **Xu & Dill, 2005**; **Schmid, 2001**).

The low concentration of LPS that can be incorporated into solutions is the main reason why the hydrolysis of the LPS is performed, in order to increase the concentration of the soluble PS in solution. The concentration of ligand in solution often needs to be in excess of the concentration of the protein in order to achieve significant binding. Increasing the concentration of the LPS without using hydrolysis leads to the precipitation

and non-solubilisation of the LPS, with eye visible “chunks” of LPS in suspension, irrelevant for the actual concentration in solution, as its solubility or hydration shell limit has been exceeded. The idea was to use both amphipathic molecules to increase the solubility (and concentration) of LPS in solution with rfhSP-D, in the hope that a protein co-crystal would form with the LPS bound to the rfhSP-D CRD.

2.6.3.2 Triton X-100

Triton X-100 is a non-ionic surfactant/detergent consisting of a hydrophilic polyethylene oxide chain (repeated 9-10 times) and a hydrophobic aromatic hydrocarbon group. Triton X-100 surfactant is a non-denaturant, with no reported impact on protein structure (**Fig. 2.8**) (**le Marie et al., 2000**). Triton X-100 is widely used in the solubilisation of lipid bilayers and in the isolation of integral membrane proteins (**le Marie et al., 2000**). An important feature when using surfactant Triton X-100 is to understand its critical micellar concentration (CMC).

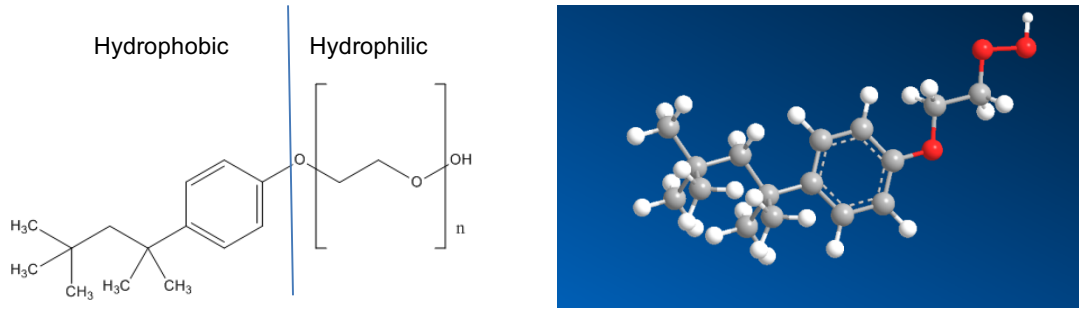


Fig. 2.8 A representation of the Triton X-100 molecule. The picture on the left shows a 2-D representation of the molecule and the image was produced using *CHEMDRAW*. The picture on right the shows a 3-D representation of Triton X-100 and the image was produced using *CHEMBIO3D ULTRA.* The hydrophilic polyethylene oxide chain can be repeated 9-10 times, while the hydrophobic component is made up of the aromatic hydrocarbon group.

A micelle is an aggregate of surfactant molecules dispersed a polar liquid. The micelle forms colloidal-sized clusters in solutions, where the hydrophilic component is in contact with the polar solvent, while the hydrophobic group is sequestrated in the centre of the micelle. Each surfactant exhibits a specific CMC below which the surfactants are

present in solution as monomers and individual chains; if the concentration of surfactants in solution is greater than the CMC, they come together forming a micelle (**Fig. 2.9**). The formation of micelles is dependent on the specific surfactant used, as these vary in hydrocarbon molecules chemistry and length, with different packing effects and hydration shells, producing varying CMCs (**Heerklotz et al., 1997**; **Lichtenberg et al., 2013**). The CMC for triton X-100 has been reported to be ≈ 0.25 mM (**le Marie et al., 2000**).

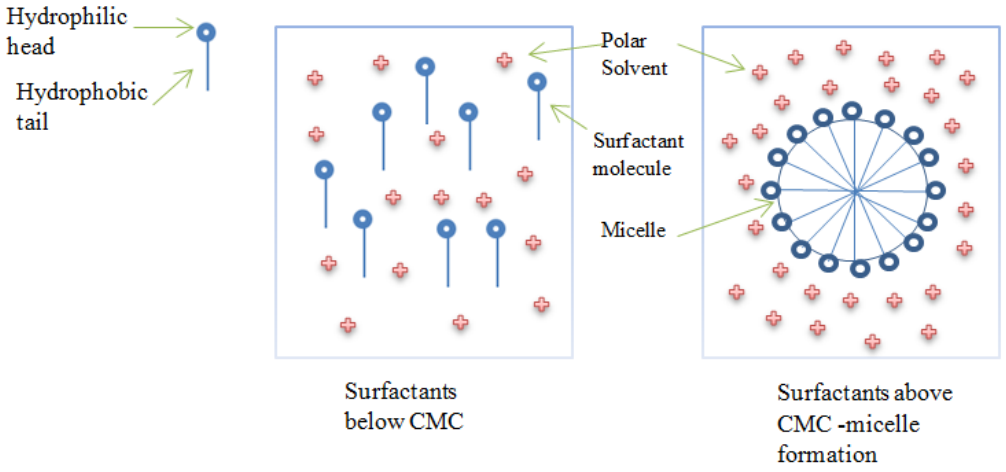


Fig. 2.9 A representation of the surfactant Triton X-100 in a polar solution. Below the CMC the surfactants will be found in solution as monomers. When the concentration of surfactant is above the CMC specific to the surfactant, the monomers in solution will come together forming a micelle. Image produced using **Lichtenberg et al., 2013** as a source.

2.6.3.3 Absolute Ethanol

Absolute ethanol is a pure form of ethanol, containing less than 1% of H_2O molecules and is composed of an ethyl group linked to a hydroxyl group (**Fig. 2.10**). Ethanol ranks high in the polarity index (a measure of how it interacts with various polar solvents) (**Torres et al., 2011**) with the ability to impact on the solubility of lipids in the cell membrane (**Sun & Sun, 1985**) and in increasing dispersion properties of lipid molecules of poorly water soluble drugs (**Sek et al., 2006**), making it a good candidate for the LPS solubilisation trials.

Ethanol has been reported to have a detrimental effect on protein structures, by denaturing the protein and altering hydrophobic interactions. Ethanol's denaturing

properties have been investigated in hen egg white lysozyme, where up to 63% ethanol (v/v) did not induce conformational changes in the protein structure, with lysozyme protein molecules reported to have increased repulsive electrostatic forces. Above 63% of ethanol (v/v) concentration led to conformational changes in lysozyme, including the transformation from an α-helix-rich structure to a β-sheet rich structure, or rod-like structures with a high content of α-helices (**Tanaka et al., 2001**; **Pace et al., 2004**). The above suggests that the concentration of ethanol used will have to be rather low so not to impact on the structural organisation of rfhSP-D.

Fig. 2.10 A representation of an ethanol molecule. The picture on the left shows a 2-D representation of the molecule produced using *CHEMDRAW*. The picture on the right shows a 3-D representation of Triton X-100 and the image was produced using *CHEMBIO3D ULTRA*.

2.6.3.4 Triton X-100 and absolute ethanol in combination with LPS

Triton X-100 and absolute ethanol have been investigated and reported in the literature to interact with LPS. Triton X-100 has been reported to have the ability to associate into micelles containing Triton X-100 and LPS molecules (**Weiser et al., 1967**). Triton X-100 has also been shown to have the ability to impact on the ratio of micelles to monomers of LPS, with a high concentration of detergent used yielding more monomers than micelles, disrupting the LPS micelle structure (**Jang et al., 2009**). This disruption of the LPS micelles, when Triton X-100 is used above the CMC, occurs by means of non-polar interactions between the alkyl chains of LPS and Triton X-100 (**Li et al., 2004**). Ethanol has been reported to have been used in crystallisation attempts of peptidoglycan recognition protein in combination with LPS. In the reported crystallisation the crystals of peptidoglycan protein were soaked for over 48 hours in a solution containing 30%

precipitant buffer and 30% ethanol, into which the LPS was dissolved at approximately 20mg/ml (**Sharma et al., 2013**).

2.6.3.5 Impact of Triton X-100 and Ethanol on LPS solubility

The determination of the success of solubilising LPS in Triton-X-100 and/or absolute ethanol was assessed by three key conditions: (1) did a specific concentration of triton X-100 and/or a % (v/v) of absolute ethanol improve the solubility of the full LPS, evaluated by means of direct observations using light microscopy? (2) where the first condition was met, did the specific concentration of triton X-100 and/or % (v/v) of absolute ethanol in combination with specific LPS and rfhSP-D yield protein crystals? (3) where the first and second conditions were met; was there any bound ligand present and what was the preference of crystal packing (space group)? The introduction of LPS rich in fatty acids (4-6 hydrocarbon chains composed of 12-14 carbon atoms each) (**Kabanov & Prokhorenko, 2011**) into a polar solution creates a hydrophobic effect. LPS can be presented in a membrane bound form or as free LPS, where the free LPS often contains a low CMC, dependent on the length of LPS (smooth vs rough) (**Parikh & Chorover, 2007**).

In these experiments free LPS was used. The introduction of free LPS into a polar solution will create a hydrophobic effect (**Fig 2.11**), where the hydrophilic component (core and O-antigen if present) will be able to hydrogen bond with water molecules, while the hydrophobic lipid A will have limited hydration by water molecules. At a high concentration (above CMC) the LPS will assemble into micelles, where the lipid A is internalised and excluded from the polar solvent (**Parikh & Chorover, 2007**). When considering the LPS it is important to highlight that the Lipid A is highly conserved, with its chemical structure commonly composed of two phosphorylated glucosamine units typically acylated with four to six fatty acids (**Wang et al., 2015**; **Frecer et al., 2000**). The remainder of the LPS, which normally includes Kdo, heptose, glucose and galactose, is hydroxyl rich and able to form hydrogen bonds with polar solutions and so able to solubilise in solution (**Parikh & Chorover, 2007**). Smooth LPS, containing the hydrophilic O-antigen, has a higher

hydrophilic to hydrophobic ratio, higher CMC and higher solubility compared to rough LPS which lacks the O-antigen (**Park et al., 2009**; **Kabanov & Prokhorenko, 2011**).

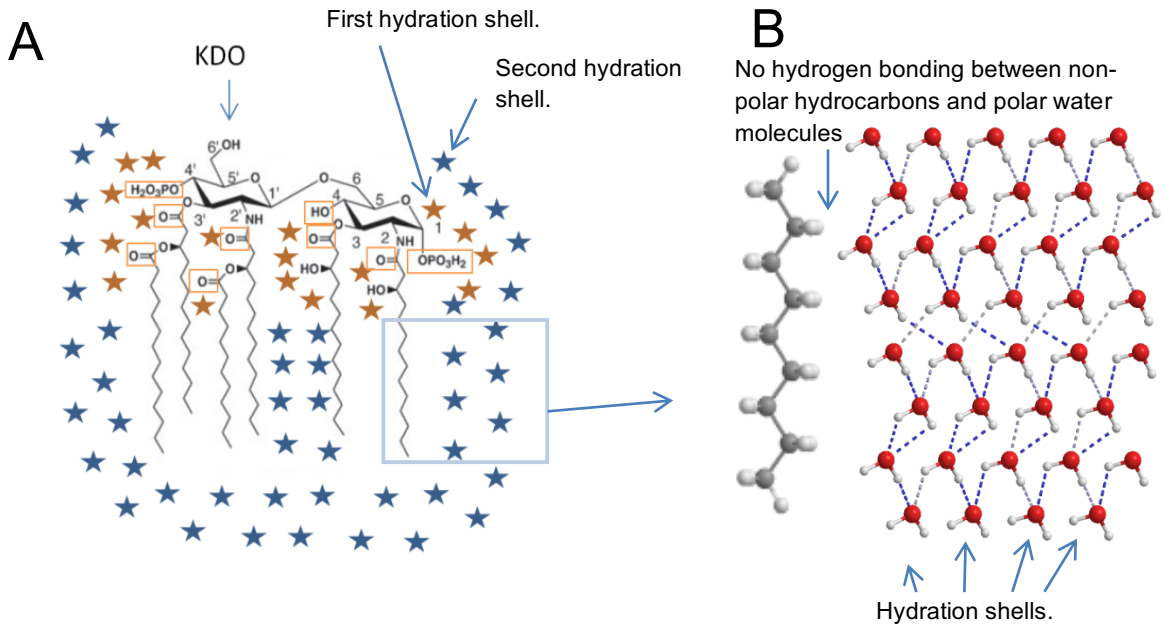


Fig 2.11 A- Image adapted from **Que et al., 2000**. This is a representation of the hydration of lipid A by water molecules. Blue stars represent water molecules solvating the hydrophobic fatty acids and interacting with other water molecules without any hydrogen bonding occurring between water molecules and the hydrocarbons. Orange stars represent water molecules which interact with other water molecules and are placed to hydrogen bond with the hydrophilic components of the Lipid A. Inside the small orange boxes are the components of the lipid A able to form hydrogen bonds with the water. **B-** Hydration shells surround the LPS hydrocarbon without any interaction between the two. The first hydration shell water molecules are ordered by the second shell, which in turn are ordered by the third shell and so on. The number of hydration shells is dependent on the number of hydrophobic solutes in solution and the number of water molecules available. Image produced using *CHEMBIO3D ULTRA.*

The introduction of LPS into a polar solution is limited to a small hydration potential and a small concentration of LPS (**Fig 2.11**). The lipid A exhibits little hydrogen bonding potential, with the glucosamine units reported to consist of 27 hydrogen bond acceptor atoms, including 8 of which are also hydrogen donors, with on average approximately 18 hydrogen bonds being made with the water molecules. Over 50% of all the hydrogen bonds occur by means of the phosphate groups in the di-glucosamine, while the 6 carbonyl groups provide approximately 16%. The di-glucosamine hydroxyl groups offer limited hydrogen bond potential (**Murzyn & Pasenkiewicz-Gierula, 2015**), see **Fig 2.11**.

The LPS core region and O-antigen (if present) offer a much higher hydrogen bonding potential. The phosphate groups have been reported to produce a pattern of negative molecular electrostatic potential, allowing engagement in electrostatic interactions or hydrogen bonding with polar solvents or cations (**Frecer et al., 2000a**). The hydration of the LPS in a polar solvent will be dependent on its concentration, with a limited number of LPS molecules that can be accommodated in the "cages" surrounded by hydration shells of water molecules. These LPS molecules will be organised through repulsion of the polar water molecules (first hydration shell) by the lipid A fatty acids and by means of van der Waals interactions between them (**Chandler, 2005**). In the first hydration shell water molecules will reorganise themselves and optimise hydrogen bonding with neighbouring water molecules, with no interaction between hydrophobic solutes and the first hydration shell water molecules (**Fig 2.11B**) (**Chaplin, 2007; Ahmed et al., 2016**; **Tomlinson-Phillips et al., 2011**).

The solubility of hydrophobic solutes such as LPS is influenced by the nature and extent of how they group together, especially for a high concentration of LPS where all the molecules are trying to be accommodated in "cages" surrounded by a fixed and limited number of polar molecules. This is expected to create a negative order (high entropy) and favour precipitation/insolubility of the LPS. However, the grouping of the hydrophobic solutes together increases the solvation of the hydrophobic solutes thus decreasing the entropy and increasing the solubility. The coming together of hydrophobic residues increases the order in the system, suggesting that the solute-water interactions do not contribute to the solubility potential, but instead the solute-solute interactions are the main reason for the limitation in aqueous solubility of hydrophobic solutes in a polar solution (**Privalov & Gill, 1989**; **Zhang & Gobas, 1995**; **Ruelle et al., 1994**).

The introduction of the Triton X-100 and/or absolute ethanol in combination with the LPS into an aqueous solution will initially contribute to the disruption of the hydrogen bonding of ordered molecules as both molecules have a hydrophobic component. This

forces the hydrophobic regions of both lipid A and Triton X-100 and/or absolute ethanol to group together, minimising the surface area in contact with the polar solvent. The hydrophilic components of Triton X-100 and absolute ethanol form the maximum possible number of hydrogen bonds with water molecules in the first hydration shell, (**Lazaridis & Paulaitis, 1992**; **Skvarla, 2001**), see **Fig 2.12**. The concentrations of Triton X-100 and ethanol should not exceed their reported CMC. At the CMC, the number of monomers available to interact with the lipid A is maximised.

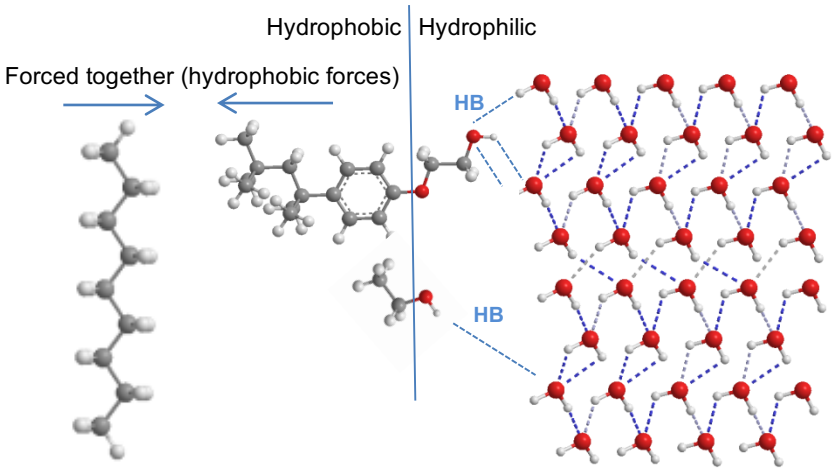


Fig 2.12 Not to scale. A representation of the proposed mechanism on how Triton X-100 and absolute ethanol can aid in the solubility of LPS in polar solutions. In **Fig 2.10B** there was no hydrogen bonding between the hydrocarbons and the water molecules. By being forced together, the hydrophobic fatty acids and the hydrophobic components of both molecules could allow the hydrophilic components of Triton X-100 and absolute ethanol to hydrogen bond with water molecules, increasing the order in solution and the LPS solubility. Image produced using *CHEMBIO3D ULTRA*.

Chapter 3 - Ligand binding structural studies of co-crystallised rfhSP-D and *H. influenzae* Eagan CA7 PS

3.1 Introduction

The crystallographic studies performed in this chapter were aimed at producing co-crystallised protein crystals of a recombinant fragment of human SP-D (rfhSP-D) in complex with *Haemophilus influenza* Eagan CA7 polysaccharide (delipidated - minus lipid A) to investigate the recognition mechanism and structural interactions between protein and ligand. A similar and shorter LPS than CA7 LPS has been previously investigated in our research group, *H. influenzae* Eagan 4A PS (delipidated), containing Kdo-HepI-GlcI and it was demonstrated that the preferential binding occurs through the HepI (**Clark et al., 2016**). The binding of the HepI was similar to previous reported binding of heptose sugars (**Wang et al., 2008**), where the rfhSP-D recognised the Eagan 4A LPS through the O6' and O7' hydroxyls groups of HepI. Binding affinity studies using rfhSP-D with Eagan 4A (1 Hep), CA7 (2 Heps) and Wild-Type (3 Heps) *H. influenzae* LPS, have demonstrated that the binding affinity of rfhSP-D is higher for the Eagan 4A (established by sugar competition assays with the results calculated through ELISA analysis), with the increase of branching and heptoses in CA7 and wild-type yielding reduced binding affinity in comparison with the Eagan 4A (**Clark et al., 2016**). Strains containing more heptoses induced stronger inflammatory responses in mice models, suggesting that *H. influenzae* has developed the ability to shield the core from immune surveillance and that the binding and recognition of core saccharides by SP-D seems more favourable when compared to the non-core saccharides. The study of rfhSP-D with CA7 LPS offered the opportunity to investigate the preference of rfhSP-D binding for terminal outer core saccharides (Glc) or the non-terminal inner core saccharides (HepI and HepII) and the opportunity to characterise the binding and orientation of the ligand. Both terminal and non-terminal

oligosaccharides have been suggested as the primary site of binding and recognition of LPS by SP-D.

3.1.1 *H. influenzae* Eagan wild-type type b LPS

*H. influenza*e Eagan type b is a typeable strain with the ability to cause various different disease (**King, 2012**), especially in infants between 6-12 months of age, with 80% of all *H. influenzae* induced disease occurring in the first five years of life (**Hambleton & Davies, 1975**). The LPS of *H. influenzae* has been shown to be able to cause symptoms without the presence of the whole organism, suggesting that the LPS holds the toxicity in *H. influenzae* (**Syrogiannopoulos et al., 1988**). The LPS from *H. influenzae* Eagan has been suggested to have the ability to mimic human glycolipids and the ability to reversible switch its terminal epitopes (phase variation) in the oligosaccharide portion (**Li et al., 2001**). The mimicking of the hosts epitope Galα1-4Gal disaccharide by means of presentation in the LPS allows *H. influenzae* bacteria to become resistant to killing, masking it from the immune system (**Weiser et al., 1998**). Phase variation allows for the position of the carbohydrates to be varied, including the pattern of non-carbohydrate components such as phosphate substituents, O-acetyl groups and ester linked glycine moieties (**Schweda et al., 2007**). The phase variation is proposed to be an evolutionary mechanism that allows for the survival of the bacteria by means of rapid adaptation to the changing microenvironments in the host (**Hood et al., 1999**). Phase variation occurs by the loss or gain of one or more of the four base repeats during DNA replication by means of slipped-strand misparing, producing frame shifts and resulting in high frequency phase variation of the oligosaccharide core (**Fox et al., 2005**; **Hood et al., 1996**). This high frequency phase variation is the main reason why the *H. influenzae* LPS biosynthesis is considered diverse. This results in the carbohydrate structure and non-carbohydrate substituents varying and affecting the recognition of the LPS structures by the host immune system, with a positive effect on the biological properties of the organism (**Raetz & Whitfield, 2002**). The capsule of typeable *H. influenzae* surrounds the OM of

the bacteria and in *H. influenzae* Eagan type b it is composed of five-carbon sugars (ribitol). Capsular expression is limited to one capsular serotype per clone of *H. influenzae*, with no signs of antigenic variation; only the quantity or thickness of the capsule seems to vary between individual bacteria (**Crisel et al., 1975**; **Corn et al., 1993**).

The *H. influenzae* Eagan Wild-Type LPS structure was first reported in 1997 to be composed of a lipid A glycosidically linked to a single Kdo, which in turn is linked to HepI (**Fig. 3.1**). The inner core structure contains a trisaccharide heptose with each heptose substituted by a hexose residue. Kdo, HepII and HepIII were shown to be phosphorylated by either phosphoethanolamine (PEA) or by a phosphate group (P). The Kdo can have a PEA or P group in position 4, giving rise to two subpopulations. *H. influenzae* antigenic diversity is seen in the various epitopes reported for HepII, with Hex4 (Glc-Glc), Hex5 (Gal-Glc-Glc) and Hex6 (Gal-Gal-Glc-Glc) all reported to be present. The Hex 6 terminal structure is Gal-Gal similar to P^k blood group antigen (**Massoud et al., 1997**).

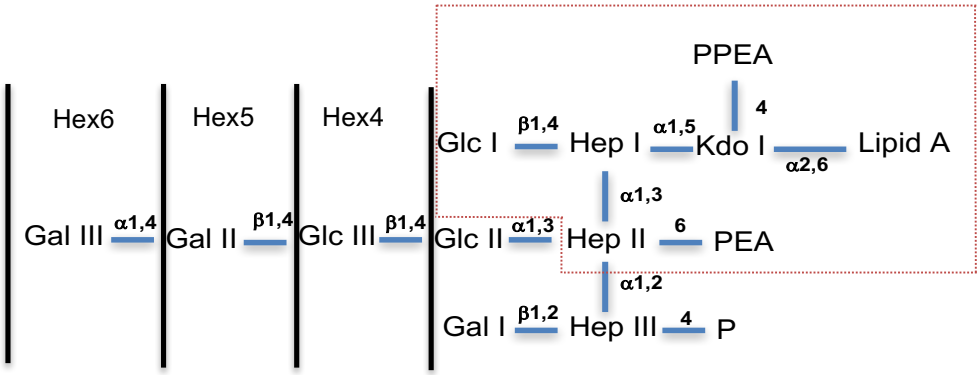


Fig. 3.1 *H. influenzae* Eagan Wild-Type LPS. In this LPS 3 hexoses are possible as extensions from the HepII. In dotted lines (orange) the CA7 LPS is shown. Glc - Glucose; Hep - L-D-Heptose, Kdo - 3-deoxy-D-manno-octulosonic acid; Gal - Galactose; P - Phosphate; PEA - Phosphoethanolamine. Image adapted from **Masoud et al., 1997.**

3.1.2 *H. influenzae* Eagan 4A and CA7 mutant LPS

H. influenzae Eagan 4A PS has been used in structural studies in combination with rfhSP-D (PDB: accession code 4E52). The Eagan 4A PS used in the structural studies contained a trisaccharide of Kdo-HepI-Glc, where the Kdo structure suffered

conformational changes upon the delipidation of Lipid A and loss of the P or PEA group in position 4, see 2.6.2.2 for more information. Mutant CA7 PS was used in the co-crystallisation attempts in combination with rfhSP-D. The CA7 mutant LPS used in this experimental work in comparison with the Eagan 4A LPS contains an extra Hep (HepII) phosphorylated by a PEA, with no hexose extensions off HepII (**Fig. 3.2**). CA7 Eagan LPS is produced by a *orfH* deletion, impacting on the addition of Glc to HepII (**Fig. 3.2**) (**Hood et al., 1996a**).

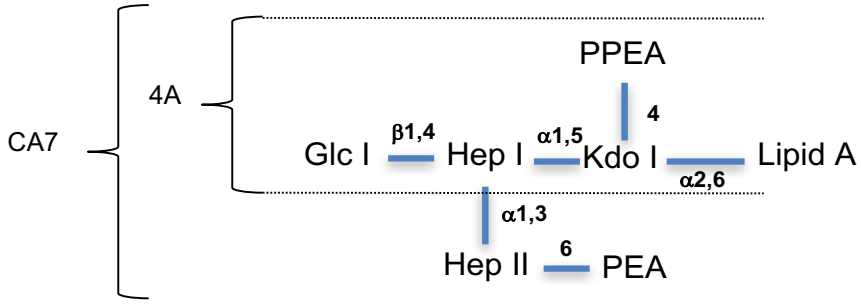


Fig. 3.2 The Eagan 4A LPS containing a trisaccharide Glc-Hep-Kdo linked to the Lipid A and with Kdo phosphorylated by either Phosphate and/or Phosphoethanolamine. The LPS from Eagan CA7 containing an extra phosphorylated HepII in comparison to the 4A LPS is displayed. Glc - Glucose; Hep - L-D-Heptose, Kdo - 3-deoxy-D-manno-octulosonic acid; P - Phosphate; PEA - Phosphoethanolamine.

3.2 Results

The aim of the experiments reported in this chapter was to co-crystallise rfhSP-D and CA7 PS bound together in order to investigate by means of X-ray diffraction the binding interactions between CA7 PS and rfhSP-D in detail. All the crystallisation experiments detailed in this chapter were carried out at Huxley Building, Keele University, Staffordshire. All data collection took place at Diamond Light Source (DLS), Diamond House, Harwell Science and Innovation Campus, Fermi Ave, Didcot, Oxfordshire. Protein, LPS and buffer constituents.

3.2.1 Protein, LPS and buffer constituents

rfhSP-D supply, purification and preparation were performed as previously described in 2.4.1. The rfhSP-D batch used in combinations with CA7 PS was prep-2 and the structure/industrial screens were from Molecular Dimensions.

3.2.2 Hydrolysis of CA7 LPS

The hydrolysis of Eagan CA7 LPS was performed according to the method previously described in 2.6.2 and was performed by Ian Burns in our research group. The supplied Eagan CA7 LPS molecular weight was calculated by the addition of individual glycoconjungates and lipid A, yielding a molecular weight of approximately 2856 Da, with Lipid A (1800 Da) accounting for 63% of the Mw (**Que et al., 2000**). 10 mg of CA7 LPS were hydrolysed and yielded 8 mg of PS. This value may be an overestimation with the potential for some of the Lipid A to be present in solution with the PS. The 8 mg were dissolved in 100 µl of deionised water and produced a concentration of 40 mM for storage and to be used in the co-crystallisation attempts with rfhSP-D.

3.2.3 Co-crystallisation of rfhSP-D and CA7 PS

The co-crystallisation of rfhSP-D in combination with CA7 PS was performed according to the methods described in 2.5. The final crystallisation drop was composed of a 1:1 ratio of specific precipitant buffers and rfhSP-D at 7.5 mg/ml (0.42 mM), CA7 PS at 10 mM and Ca^{2+} concentration at 5 mM in the drop. Structure screens 1 and 2 (SS1 and SS2) were initially used to establish growth conditions. If positive results were achieved (crystal growth), follow ups were carried out where the initial conditions were varied slightly in order to improve crystal morphology and in order to identify the best precipitant conditions. Follow ups consisted of the refinement of the initial conditions in the SS1 and SS2, in order to find the optimum environment to produce suitable high resolution crystals for X-ray diffraction. A total of 2 µl was placed in the micro-bridge, with the CA7 PS in excess of the number of rfhSP-D monomer binding sites by 23.8 x (molar excess), with the aim of ensuring that the ligand bound to rfhSP-D with high occupancy.

3.2.3.1 Initial structure screens

Five 24-well Linbro plates were initially set down on 17/12/2013 containing SS1 and SS2 with the following tray codes, RS20, RS21, RS22, RS23 and RS24, and

containing the 100 structure screen buffer precipitants and rfhSP-D at 7.5 mg/ml in combination with CA7 PS at 10 mM (23.8x ligand molecular excess to rfhSP-D monomer). Positive crystal growth was achieved for structure screen 1 condition 5 (tray RS20, well A5), structure screen 1 condition 28 (tray RS21, well A4), structure screen 1 condition 40 (tray RS21, well C4), structure screen 1 condition 44 (tray RS21, well D2) as shown in **Table 3.1**. SS1 conditions 5, 28 and 44 gave the most promising looking crystals. Condition 5 produced crystals resembling thin squares, condition 28 and 44 produced crystals with variety of shapes, including rectangles and small squares.

Well ID	Comments and conditions
RS20 A5	This well was set down on 17/12/2013. The condition used in this well comes from SS1 condition 5. This well contained 0.1 M Na acetate pH 4.6 and 2.0 M Ammonium Sulphate.
RS21 A4	This well was set down on 17/12/2013. The condition used in this well comes from SS1 condition 28. This well contained 0.1 Na Hepes pH 7.5 and 0.8 M NaH_2PO_4/ H_2KO_4P
RS21 D2	This well was set down on 17/12/2013. The condition used in this well comes from SS1 condition 44. This well contained only 2.0 M Ammonium sulphate.
RS21 C4	This well was set down on 17/12/2013. The condition used in this well comes from SS1 condition 40. This well contained 0.2 M Ammonium Phosphate Monobasic, 0.1 M Tris pH 8.5 and 50% MPD.

Table 3.1 The initial SS1 co-crystallisation trial wells that showed positive crystal growth

3.2.3.1.1 SS1 condition 5: RS20 A5

SS1 condition 5 produced various crystals suitable for X-ray diffraction and 3 crystals were collected and frozen for testing at Diamond Light Source. This condition produced mostly multiple layered squares and some singles ones shown in **Fig. 3.3**

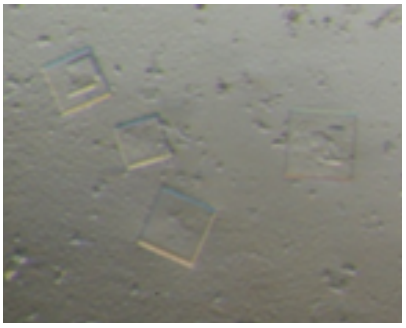


Fig 3.3 The typical morphology produced in the RS20 A5 well is displayed, with various crystals produced resembling thin and often multiple square plates.

3.2.3.1.2 SS1 condition 28: RS21 A4

SS1 condition 28 produced multiple crystals with a total of 8 crystals cryo-protected and stored in liquid nitrogen to be tested at the DLS. Squares and triangles were the typical morphology seen and shown in **Fig. 3.4**. RS21 A4 crystals took over 8 months to grow and when the crystals were seen, limited quantities of CA7 PS were available and so this condition was not followed up any further.

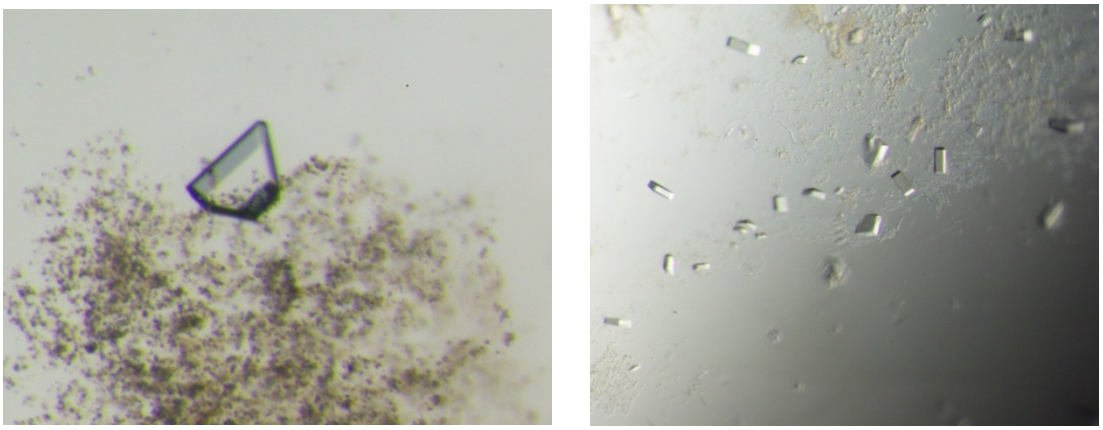


Fig. 3.4 The typical morphology of the crystals grown in the RS21 A4 well is shown. This well was particular successful in growing crystals. Even though many crystals grew over time, the majority of crystals diffracted to a very low resolution.

3.2.3.1.3 SS1 condition 44: RS21 D2

Similar to the RS21 C4 well, RS21 D2 initially produced promising results with nucleation and crystal growth observed and a follow up was produced, RS34 (discussed below). The initial nucleation and crystal growth was somehow halted, with no further crystal growth or new nucleation occurring as time passed. This was rather surprising, especially considering that the follow up RS34 yielded various crystals in a short period of time.

3.2.3.1.4 SS1 condition 40: RS21 C4

RS21 C4 well (and follow up RS33) wells initially produced promising results where nucleation and crystal growth was observed. However, as time passed it became evident that no further nucleation and growth was occurring, with only a few small

nucleated crystals produced and not suitable for X-ray diffraction. Both trays were regularly examined for over three years, with no further crystal growth observed.

3.2.3.2 Follow up screens

Positive crystals growth in SS1 and SS2 conditions were followed up in order to find the optimum environment to produce suitable high resolution crystals for X-ray diffraction. The concentration of rfhSP-D and CA7 PS remained the same in the drop as in the SS1 and SS2 co-crystallisation attempts, producing a 23.8 x ligand molecular excess to rfhSP-D monomer. Three out of the four positive hits from the initial structure screens were followed up as shown in **Table 3.2**, where each precipitant buffer was varied slightly in order to optimise the crystal growth.

SS tray ID and well	Follow up tray	Initial conditions and variations
RS20 A5 (SS1 5)	RS32	This well was set down on 14/02/14. SS1 condition 5 contained 0.1M Na acetate pH 4.6 and 2.0M Ammonium Sulphate. In the follow up a 3 by 3 (9 wells) were produced, where the pH of the Na acetate and concentration of Ammonium Sulphate were varied, pH 4.3, 4.6 and 4.9; and 1.8M, 2.0M and 2.2M respectively.
RS21 C4 (SS1 40)	RS33	This well was set down on 14/02/2014. SS1 C.40 contained K/Na tartrate 0.4M. In the follow up 3 wells were produced with varying concentrations of K/Na tartrate, 0.35M, 0.4M and 0.45M.
RS21 D2 (SS1 44)	RS34	This well was set down on 13/3/2014. SS1 C.44 contained 2.0M Ammonium Sulphate. The follow up 3 wells were produced with varying concentrations of Ammonium Sulphate, 1.8M, 2.0M and 2.2M. On the 19/3/14 another 2 wells were made at 2.0M ammonium sulphate (B2 and C2).

Table 3.2 The three follow up trays produced from the initial positive hits in SS1.

The final positive indication from the structure screen, RS21 A4, resulted in positive crystal growth, but unfortunately it took over 8 months for the crystals to develop and at that time limited quantities of CA7 LPS were available and so it was decided not to follow up the condition. In total 11 24-well Linbro plates were produced and ≈ 150 precipitant buffers and/or variations investigated.

3.2.3.2.1 Follow up to SS1 RS20 A5: RS32

RS32 was produced as a follow up from RS20 A5. In this tray a 3 by 3 screen was produced, shown in **Table 3.3**. This follow up proved to be very successful as all 9 wells produced protein crystals. The crystals produced were often multiple or layered, with a few single crystals. In total over 30 crystals were collected from the different wells and exposed at DLS. The typical morphology of the crystals resembled thin squares plates, with triangles and irregular shaped crystals also seen; see **Fig. 3.5** for the typical morphology of crystals produced. All the crystals exposed from this tray were shown to be protein. RS32 A1 well (in combination with RS34 B2 – see below) produced some of the crystals with the best diffraction and resolution achieved.

	1	2	3
A	0.1M Sodium Acetate pH 4.3 Ammonium Sulphate 1.8M	0.1M Sodium Acetate pH 4.6 Ammonium Sulphate 1.8M	0.1M Sodium Acetate pH 4.9 Ammonium Sulphate 1.8M
B	0.1M Sodium Acetate pH 4.3 Ammonium Sulphate 2.0M	0.1M Sodium Acetate pH 4.6 Ammonium Sulphate 2.0M	0.1M Sodium Acetate pH 4.9 Ammonium Sulphate 2.0M
C	0.1M Sodium Acetate pH 4.3 Ammonium Sulphate 2.2M	0.1M Sodium Acetate pH 4.6 Ammonium Sulphate 2.2M	0.1M Sodium Acetate pH 4.9 Ammonium Sulphate 2.2M

Table 3.3 Tray RS32, the follow up to RS20 A5 (SS1 5). All wells produced protein crystals with RS32 A1 well produced some of the best diffraction (+ and resolution) of crystals - shaded in grey.

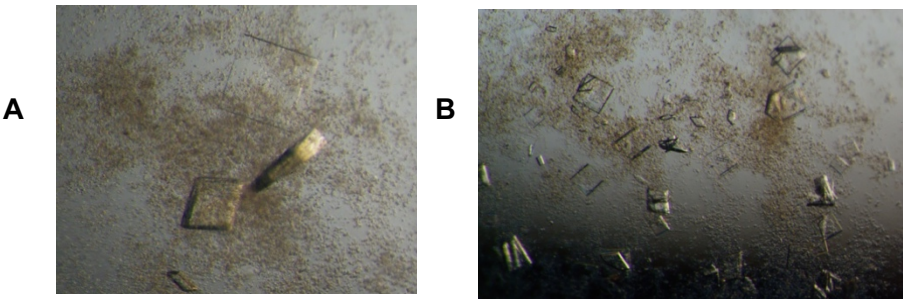


Fig. 3.5 The typical morphology of the crystals produced in RS32 wells. Crystals were most often multiple with an undefined morphology. The brown precipitate in the solution is the CA7 PS. **A** RS32 well A3. **B** RS32 well B2.

3.2.3.2.2 Follow up to SS1 RS21 D2: RS34

RS34 tray was produced as a follow up from RS21 D2 well. In this tray 3 wells were initially set down, with two other wells added at a later date, see **Table 3.4**. Similar to

the RS32 wells, RS34 wells were also very successful in producing many protein crystals, with all five wells producing various crystals. The crystals were shown to be most often multiple or layered, with only a few single crystals adequate for X-ray diffraction. The typical morphology of the crystals produced was also similar to that seen for the RS32 A1 well, with mainly squares and triangles produced, see **Fig. 3.6**. A total of 12 crystals were collected and tested at DLS. RS34 B2 well (in combination with RS32 A1) also produced some of best resoltution achieved.

	1	2	3
A	Ammonium Sulphate 1.8M	Ammonium Sulphate 2.0M	Ammonium Sulphate 2.2M
B		Ammonium Sulphate 2.0M	
C		Ammonium Sulphate 2.0M	

Table 3.4 Tray RS34, the follow up to RS21 D2 (SS1 44). All wells produced protein crystals. B2 and C1 were set down at a later stage

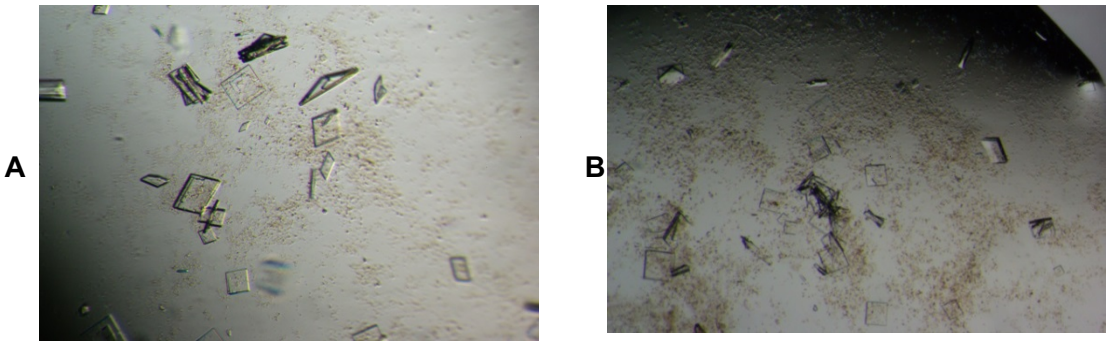


Fig. 3.6 The typical morphology of the crystals produced in RS34 wells. Crystals were most often multiple with an undefined morphology. The brown precipitate in the solution is the CA7 PS in solution. **A-** RS34 well A2. **B-** RS34 well B2.

3.2.3.2.3 Follow up to SS RS21 C4: RS33

No suitable crystals were observed in either RS21 C4 or the follow up tray RS33, and these will not be discussed further.

3.2.3.3 Cryo-protection of rfhSP-D and CA7 PS crystals

3.2.3.3.1 Cryo-protection strategies

Cryo-protection of crystals of interest was performed according to the method described in 2.5.4.1.1, where both strategy 1 and strategy 2 were employed. Over 50 co-

crystals of rfhSP-D and CA7 PS were cryo-protected and flash-frozen with X-ray data being collected in each case. 2-methyl-2,4-pentanediol (MPD) was chosen as sole cryo-protectant to be used, as previous trials in our research group demonstrated that MPD is normally a good cryo-protectant for rfhSP-D crystals. Strategy 1 was initially used where increasing concentrations of MPD were added in 2 µl volume aliquots at 2 minute intervals to the micro-bridge. In this strategy each of the four different concentrations of MPD (5, 10, 15 and 20%) contained 10 mM of CA7 PS in a solution which matched the particular well conditions. This strategy produced a final concentration of MPD of ≈ 20%, CA7 PS ≈ 10 mM and Ca^{2+} ≈ 1.2 mM in the micro-bridge and this was used in the cryo-protection of crystals from RS20 A5 and from follow up RS32 C2. The cryo-protected crystals were subsequently looped out of the drop and stored in liquid nitrogen.

Strategy 2 was employed in order to ration the limited quantity of CA7 PS while maintaining the CA7 PS present in micro-bridge at a reasonable concentration. Strategy 2 was based on the addition of 0.5 µl (twice) of 50% MPD, with 2-minute interval between additions, with no CA7 PS in the cryo solutions. Strategy 2 produced a final concentration of MPD of ≈ 16.6%, CA7 PS ≈ 7 mM and Ca^{2+} ≈ 3.33 mM in the micro-bridge. Strategy 2 was used in the cryo-protection of crystals from RS32 and RS34 follow ups. The cryo-protected crystals were subsequently looped and stored in liquid nitrogen.

3.2.3.3.2 RS20 A5 (SS1 - C5): Cryo-protection information

The cryo-protection used for the SS1 condition 5 crystals (RS20 A5) was based on the strategy 1, where CA7 PS was soaked in together with the different concentrations of MPD (see **Table 3.5**). The cryo-protection was successful in cryo-protecting the crystals and no ice rings were detected during data collection.

Well Conditions	MPD %	Volume (µl)	Time in solution (minutes)	Comments
0.1 M Na Acetate pH 4.6, 2.0M ammonium sulphate and CA7 PS at 10mM.	5	+2	2	The first cryo-protection using a co-crystal of both rfhSP-D and PS CA7.
	10	+2	2	
	15	+2	2	
	20	+2	2	
	20	+2	2	
	20	-10 (solution reservoir) + 10 of 20% MPD – Exchanged	7	

Table 3.5. The cryo-protection performed in the RS20 A5 well. Strategy 1 was used as the cryo-protection method and included R5 PS at 10mM in each of the solutions of the 5, 10, 15 and 20% MPD

3.2.3.3.3 RS32 (follow up to RS20 A5, SS1 - C5): Cryo-protection information

In the RS32 tray both strategies for cryo-protection were applied. Strategy 1 containing the PS CA7 as a soak Strategy 1 containing the PS CA7 as a soak (10mM) was initially used for the RS32 C2 well. Due to the limited stock of CA7 PS, strategy 2 was then chosen as a preferential choice for the remaining wells - RS32 A1,2,3; B1,2,3 and C1 and 3. Cryo-protection by using both strategy 1 and 2 was rather successful, with most crystals tested showing no signs of ice rings. Only the crystals from the RS32 A1 well showed signs of some ice rings, suggesting poor cryo-protection. Due to the high number of crystals tested it is unfeasible to detail every single cryo-protection here; instead the information regarding the important well RS32 A1 will be given below. RS32 A1 well is of significant interest, as this produced a crystal RS32 A12 (crystal 2) used in the structure solution and molecular building of the CA7 PS into rfhSP-D model.

The RS32 A1 well was cryo-protected with the MPD strategy 2. During the cryo-protection the crystal seemed rather happy with the two additions of 50% MPD. However, in the diffracted data ice rings were visible, suggesting poor cryo-protection of the crystals in this well. The RS32 A12 crystal used in the structure solution at later stages also showed ice rings in the diffracted data. The method used to cryo-protect the RS32 A1 well is shown in **Table 3.6**. In strategy 2 no CA7 PS ligand was present in the solution.

Well Conditions +	MPD %	Volume (µl)	Time in solution (minutes)	Comments
0.1M Na acetate pH 4.3 and 1.8M Ammonium Sulphate	50	+0.5	≈2	RS32 A12 crystal seemed to sustain the additions of 50% MPD
	50	+0.5	≈2	

Table 3.6 The cryo-protection performed in the RS32 A1 well. Strategy 2 was used as the cryo-protection method with no ligand present in the 50% MPD cryo solution.

3.2.3.3.4 RS34 (follow up to RS21 D2, SS1 - C44) – Cryo-protection information

In RS34 tray only strategy 2 was used, where no CA7 PS was added during the cryo-protection stages. The crystals from this well were able to sustain the additions of 50% (twice) MPD, with no signs of crystals degradation. The X-ray diffracted data also showed that the crystals were cryo-protected efficiently with no signs of ice in the diffraction images. Similar to the RS32 well, the high number of crystals cryo-protected and tested makes it unfeasible to detail every single cryo-protection. Instead, the cryo-protection of the RS34 B2 well will be detailed below. This well produced a crystal, RS32 B21 (crystal 1) used in structure solution and molecular building of the CA7 PS and the rfhSP-D model.

The RS34 B2 well was cryo-protected with the MPD strategy 2. The crystals collected maintained their morphology during the cryo-protection stages, with the diffraction data suggesting successful cryo-protection. The method used for the cryo-protection of RS34 B2 is shown in **Table 3.7**. The RS34 B21 crystal was present in the cryo for 6 minutes.

Well Conditions +	MPD %	Volume (µl)	Time in solution (minutes)	Comments
2.0M Ammonium sulphate	50	+0.5	≈2	The RS34 B21 crystal seemed to sustain the additions of 50% MPD
	50	+0.5	≈2	

Table 3.7 The cryo-protection performed in the RS34 B2 well. Strategy 2 was used as the cryo-protection method, with no ligand present in the 50% MPD cryo solution.

3.2.4 X-ray diffraction and Data collection

Due to the high number of co-crystals tested only the crystals that yielded significant results will be discussed in detail. Successful data collection from crystals with resolution up to 3.5Å or above were further investigated by means of data analysis using the CCP4 software (see 2.5.5), with the aim of investigating the presence of CA7 in the co-crystallised rfhSP-D crystals. From the structure screen SS1, 11 crystals were collected and exposed at Diamond Light Source. RS20 A5 and RS21 A4 produced various proteins crystals. From the follow ups, an impressive 42 crystals were prepared and exposed at Diamond where most of the crystals tested were shown to be protein. Five crystals produced the best resolution; RS34 A22 (Tray RS34, Well A2, Crystal 2), RS32 A12, RS32 A31, RS34 B21 and RS32 B27 (**Table 3.8**) and these were data analysed further in order to determine the quality of the electron density maps.

Crystal Code	Date exposed	Data Collection resolution	Data Analysis	Beamline	Structure Solution	Notes
RS34 A22	27/4/14	DC ≈3.73 Å	Yes	I03	No	The best resolution achieved for the co-crystals CA7 PS
RS32 A12	5/7/14	DC ≈3.05 Å	Yes	I04-1	Yes	
RS32 A31	5/7/14	DC≈ 3.37 Å	Yes	I04-1	No	
RS34 B21	5/7/14	DC ≈3.24 Å	Yes	I04-1	Yes	
RS32 B27	25/9/15	DC≈4.18 Å	Yes	I04-1	No	

Table 3.8 Summary of data collection for the crystals with the best resolution (fast_dp), RS32 A12 and RS34 B21 data sets offered the best resolution data (shaded in grey).

A total of 53 crystals were tested at DLS, with 90% of the crystals being protein crystals and approximately 10% small molecule (salts). Additional details of the data collection for the two crystals (RS32 A12 and RS34 B21) used in the production of the electron density maps of rfhSP-D complexed with Eagan Ca7 PS are given below.

3.2.4.1 RS32 A12 Crystal – Data collection

This co-crystal was exposed on 5/7/14 on the I04-1 beamline. The parameters used in the snap and data collection are shown in **Table 3.9**. The final concentration of CA7 PS was ≈ 7 mM (16.67 x molecular excess of ligand to rfhSP-D monomer) following cryo-protection. The morphology of the crystal resembled a small and fairly thin and very pretty square plate, shown in **Fig. 3.7A**. The initial exposures of this crystal (snap) showed some signs of ice rings, see **Fig. 3.7B**. Following the snap of the crystal, EDNA, the diffraction characterisation and data collection strategy software at DLS, suggested a maximum resolution of 3.5 Å.

SNAP or DC	Start Angle (˚)	φ (˚)	Total φ (˚)	Number of Images	Exposure per Image (s)	Total Exposure (s)	Maximum Resolution (Å)	Detector Distance (mm)	Transmission (%)
SNAP	0	0	90	3	0.5	1.5	1.8	292	47
DC	115	0.2	134	670	0.034	22.78	3	393.1	47

Table 3.9 The parameters used for the initial SNAP and for the data collection of the RS32 A12 crystal. SNAP- first diffraction; DC- data collection; φ (phi) – rotation angle.

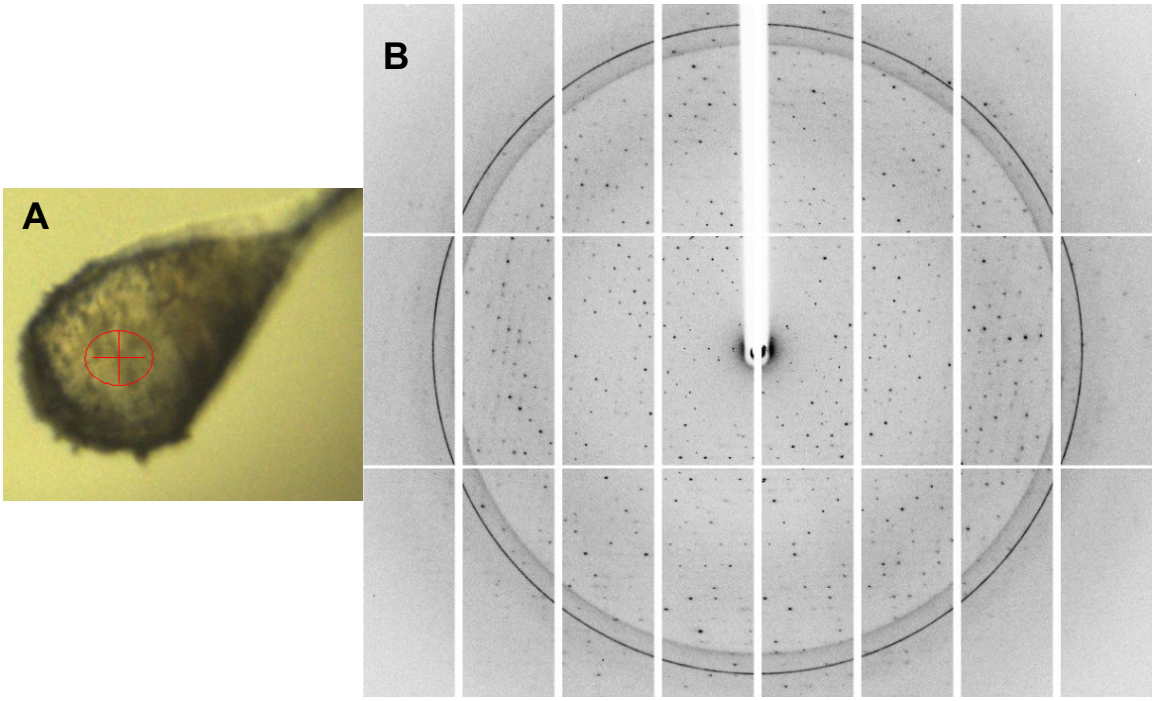


Fig. 3.7 Images collected at DLS for the RS32 A12 crystal. **A** The crystal before a SNAP was taken to evaluate the quality of the crystal. **B** the diffraction pattern produced by one of the SNAPS. Some ice rings can be seen in the image.

3.2.4.2 RS34 B21 Crystal - Data collection

This co-crystal was also exposed on 5/7/14 on the I04-1 beamline. The final concentration of CA7 PS was ≈ 7 mM (16.67 x molecular excess of ligand to rfhSP-D monomer) following cryo-protection. This crystal was initially snapped, with no signs of ice rings. EDNA suggested a maximum resolution of 3.55 Å and data was collected using the parameters shown in **Table 3.10**. This particular crystal was a nice square plate (**Fig. 3.8A)**, with the one of the corners of the crystal broken.

SNAP or DC	Start Angle (˚)	φ (˚)	Total (˚)	Number of Images	Exposure per Image (s)	Total Exposure (s)	Maximum Resolution (Å)	Detector Distance (mm)	Transmission (%)
SNAP	0	0	90	3	0.5	1.5	3.5	463	75
DC	45	0.25	124	496	0.034	16.864	3	393.1	100

Table 3.10 The parameters used for the initial SNAP and for the data collection of the RS34 B21 crystal. SNAP- first diffraction; DC- data collection; φ (phi) – rotation angle.

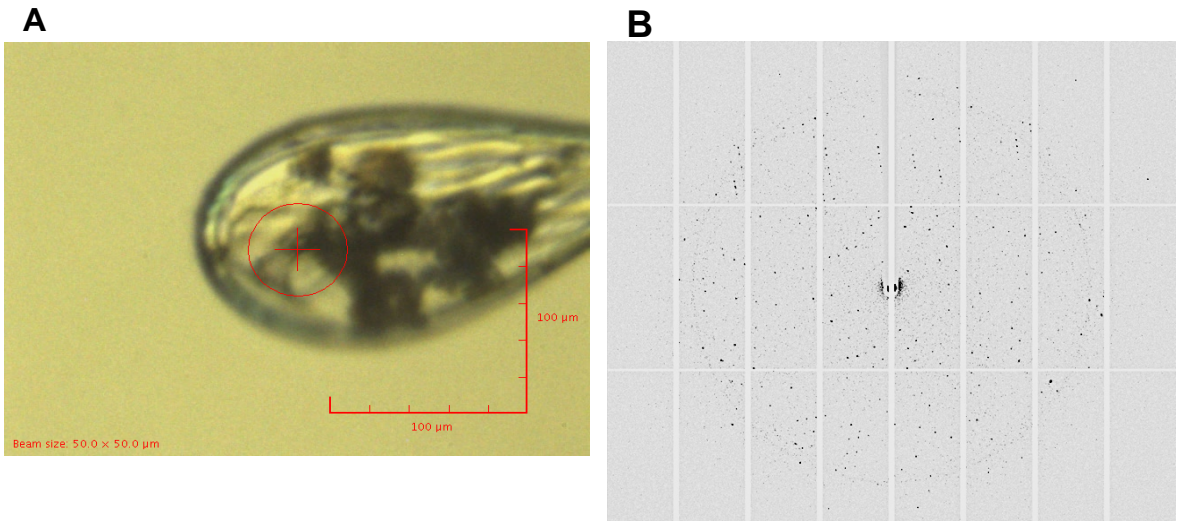


Fig. 3.8 Images collected at DLS for the RS34 B21 crystal **A-** On the left, the picture of the crystal before a snap was taken to evaluate the quality of the crystal. The loop was rather dirty making it hard to visualise the crystal. **B-** On the right, the diffraction pattern produced following one of the snaps.

3.2.5 Data Processing

3.2.5.1 fast_dp data processing

The RS34 A22, RS32 A12, RS32 A31, RS34 B21 and RS32 B27 data were initially analysed at Diamond via fast_dp (see **Table 3.8**), with RS34 A22 diffracting to 3.73 Å,

RS32 A12 to 3.24 Å, RS32 A31 to 3.37 Å, RS34 B21 to 3.05 Å and RS32 B27 to 4.18 Å. All the five data sets analysed indicated a novel space group, the tetragonal P4 2 2, previously unreported for rfhSP-D. The novel space group encouraged the possibility that CA7 PS was bound to the rfhSP-D protein, impacting on the crystal contacts and yielding a new molecular arrangement in the crystal (space group). The decision was made to investigate the data sets that looked more promising, by indexing and integrating the spots and images in house, RS32 A12 and RS34 B21 were decided as the best out the five. The remaining three data sets, contained one or more poor statistical parameters, making them less useful in the structure solution and molecular building of rfhSP-D and CA7 PS. The poor statistical parameters were: the RS34 A22 - data set had a low completeness (data not shown); the RS32 A31 - data set gave a higher R_{merge}, (data not shown) and the RS32 B27 - data set had the lowest resolution out of the five data sets and so it was not investigated further. RS32 A12 and RS32 B21 offered better statistical parameters and were chosen as the best data sets to be used in data solution.

3.2.5.2 *iMOSFLM* / *CCP4* data processing

Following on from the data analysis using the fast_dp of RS32 A12 and RS34 B21 data sets, the images and spots were integrated and indexed in house using *iMOSFLM*. During this process the space group was confirmed as $P42_12$. Using *AIMLESS*, various different runs were investigated, where batches of images containing a high R_{merge} were unselected and the resolution of the data limited (decreased). After various runs of *AIMLESS*, the data resolution was limited to 3.10 Å for RS32 A12 and RS34 B21 to 2.98 Å, with a great improvement in the R_{merge}. *TRUNCATE* was used to investigate the twinning of the both data sets, with no twinning reported. *UNIQUEIFY* was used to add the free R-flag, while the *RE-INDEX* program was used to re-index the space group into $P42_12$. See **Table 3.11** for the data processing crystallographic parameters following in house indexing.

Data collection Parameters	RS32 A12			RS34 B21		
Overall, InnerShell and OuterShell	Overall	InnerShell	OuterShell	Overall	InnerShell	OuterShell
Wavelength (Å)	0.92000			0.92000		
Low Resolution limit (Å)	77.01	77.01	3.31	77.06	77.06	3.16
High Resolution limit (Å)	3.10	8.77	3.10	2.98	8.94	2.98
Space Group	P4 2_1 2			P4 2_1 2		
Unit Cell lengths (Å)	a=112.45, b=112.45, c=105.67			a=112.52, b=112.52, c=105.75		
α, β, γ (°)	90, 90, 90			90, 90, 90		
N. of reflections	59195	2679	10517	79268	3004	13072
N. of Unique reflections	12450	603	2233	12781	510	2058
Average I/σ	8.7	23.6	2.7	9.5	22.3	3.4
CC (1/2)	0.993	0.999	0.715	0.994	0.996	0.843
Completeness %	97.9	94.2	98.5	90.2	86.6	91.1
$R_{merge\ \%}$	13.0	3.4	52.2	13.3	5.2	50.1
Multiplicity	4.8	4.4	4.7	6.2	5.9	6.4
Fast_dp (DLS) or processed in house	Processed in house			Processed in house		

Table 3.11 Data processing parameters, following the in house indexing and integration of images and spots for the RS32 A12 and RS34 B21 data sets.

3.2.6 Rigid Body Refinement

The molecular replacement was performed by Dr. *Annette Shrive* using the ligand-free rfhSP-D structure (PDB: accession code 1PW9) and using *AMoRe* program (CCP4). The molecular model produced was denominated rig_rs34. The R-factors for the *REFMAC5* rigid body refinement for the RS32 A12 data were 36.91% for the R_{free} and 36.14% for the R-factor; for RS34 B21 38.33% for R_{free} and 36.72% for the R-factor. This model contained the coordinates for the three chains of rfhSP-D minus the calcium and waters. The rigid body refined model allowed for the inspection of the binding site in terms of the presence of the Ca^{2+} and the investigation of the overall structural organisation of key residues in each of three subunits.

3.2.7 Initial Electron Density maps

The direct electron density map (2Fobs - Fcalc) and the difference map (Fobs - Fcalc) were initially created by using the rigid body refined ligand-free rig_rs34 model for the phases. The electron density maps for both RS32 A12 and RS34 B21 produced following rigid body refinement showed the usual overall structure of rfhSP-D, previously reported in the literature (Shrive et al., 2003; Shrive et al., 2009; Wang et al., 2008). The electron density maps and the superimposed model rig_rs34 corresponded well. The electron density maps showed the three chains A, B and C and the overall structure seen in the trimeric aggregate with the three C-terminal globular domains (236-355) and the three chains linked through the α-helical coiled-coil neck region (203-235). Both the RS34 B21 and RS32 A12 electron density maps showed suitable density for ligand modelling (**Fig. 3.9 A and Fig 3.9 B** respectively) and the findings of both maps will be detailed together in the subsequent pages.

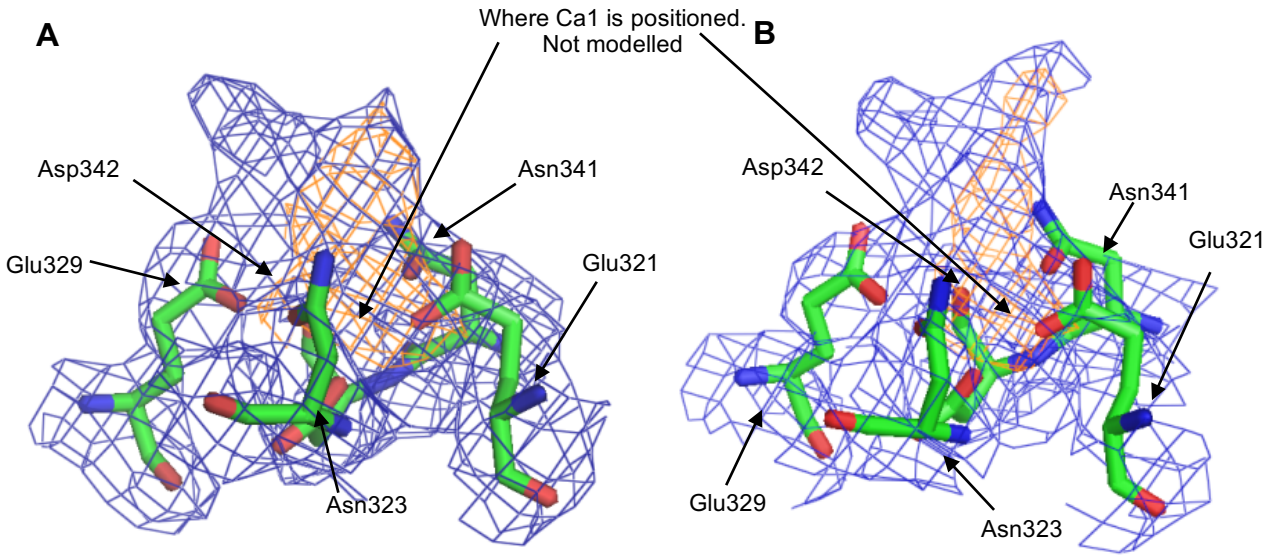


Fig. 3.9 RS34 B21 and RS32 A12 electron density maps for the binding pocket in chain B produced following the rigid body. In both images, the direct map (2Fobs-Fcalc) is in blue and contoured at 1σ and the difference map (Fobs-Fcalc) positive density is in orange and contoured at 2σ. The electron density displays a similar shape, suggesting that the same ligand (Hep or Glc) had bound in both separate crystals. In both images, important Ca1 coordinating residues are displayed, including Glu321, Asn323, Glu329, Asn341 and Asp342; no Ca1 was modelled at this stage. **A**. RS34 B21 electron density map. **B**. RS32 A12 electron density map. Figures A and B generated using PYMOL.

It is important to mention that these two maps showed remarkable similarities, with both electron density maps producing similar density (in terms of shape), suggesting a similar ligand orientation and overall similar structural organisation of rfhSP-D. **Fig 3.9 A** displays the RS34 B21 map, while **Fig. 3.9 B** displays the RS32 A12 map of chain B.

Even though the electron density maps produced following the rigid body refinement for RS32 A12 and RS34 B21 showed the overall structure of rfhSP-D previously reported in literature for the $P2_1$ space group (**Shrive et al., 2003**; **Shrive et al., 2009**; **Wang et al., 2008**), significant differences were noted in all three chains, particularly in terms of Ca1, Ca2 and Ca3 occupancy.

3.2.7.1 Chain A

In both data sets, chain A showed no density for Ca1, Ca2 and Ca3, with the binding pocket located in the vicinity of a symmetry related molecule. Chain A showed an unusual binding pocket with no suggestion of Ca1 and with some shifts seen for the Ca1 coordinating residues (Glu321, Asn323, Glu329, Asn341 and Asp342). The density surrounding the pocket in chain A was very weak, but an attempt was made to characterise the shifts seen in some residues upon the loss of the Ca1. The loss of Ca1 in the binding pocket resulted in an increased electronegativity in the pocket with Asn341 and Glu321 suggested to rotate away from their normal position (**Fig. 3.10 A).** The symmetry related molecule chain A sits right above the binding pocket of chain A, leaving no space for ligand. **Fig. 3.10 B** shows the binding pocket of chain A and the nearby symmetry related molecule.

3.2.7.2 Chain B

In both data sets chain B showed density for ligand and positive density for the Ca1, however no density for Ca2 and Ca3 could be visualised. The density surrounding the binding pocket was poorly defined in terms of the orientation of the ligand, creating some problems and ambiguity when characterising the right ligand and its orientation. A

symmetry related rfhSP-D molecule is again located in the vicinity of the binding pocket potentially impacting on the orientation of the bound ligand. Chain B will be discussed in more detail below, as it was the only chain of rfhSP-D with electron density which suggested bound ligand.

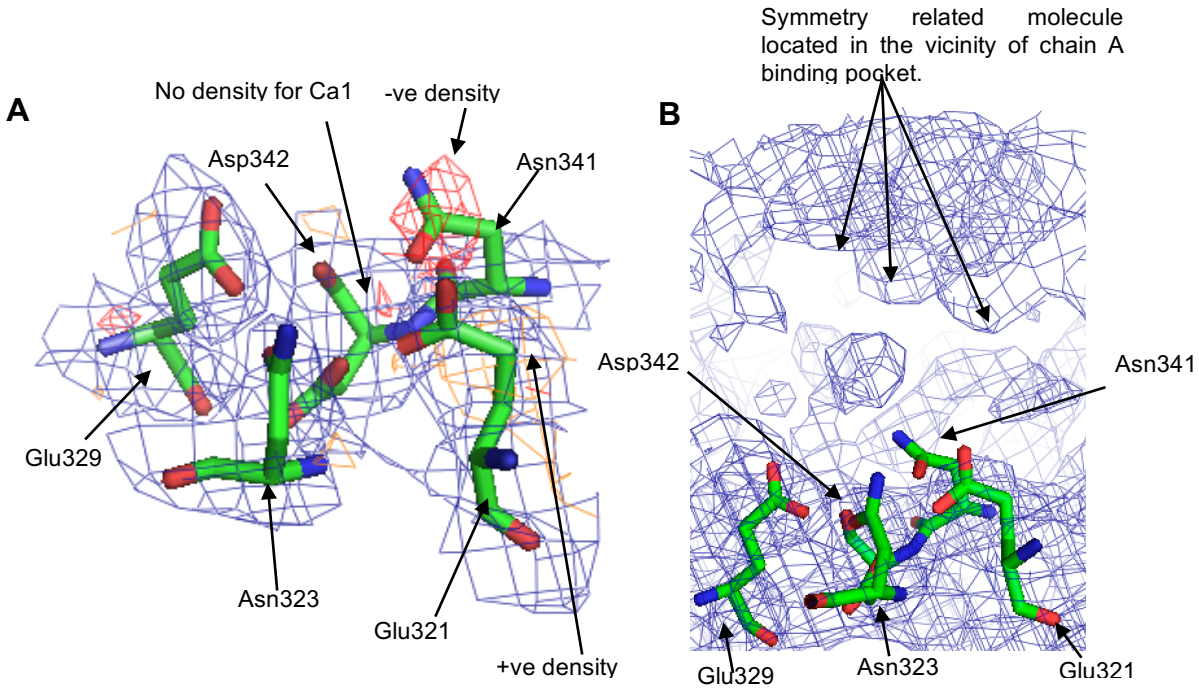


Fig. 3.10 Electron density map of RS34 B21 data set showing chain A binding pocket, the lack of Ca1 and the nearby symmetry related molecule. In both images Glu321, Asn323, Glu329, Asn341 and Asp342 are labelled; the direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. **A-** in addition to the direct map the difference map (Fobs-Fcalc) is also displayed in orange for positive density and red for negative density; contoured at 2σ. Positive (+ve) density surrounding Glu321 suggests that the residue is rotated away from the binding pocket as no Ca1 is present for coordination. The head of Asn341 is surrounded by negative (-ve) density, suggesting that this particular residue also suffers a rotation away from the normal position when Ca1 is not present. **B-** the location of the symmetry related molecule suggests that even if Ca1 was present no ligand could fit in the tight space between the chain A binding pocket and symmetry related molecule. In this image the density has been expanded (radius of contour) to display the symmetry related molecule. Figures A and B generated using *PYMOL*.

3.2.7.3 Chain C

In both data sets chain C resembled the typical unliganded rfhSP-D binding pocket, with density for Ca1 and coordinating waters. Similar to chain A and B, no density

for Ca2 and Ca3 could be visualised. Similar to chain A, a symmetry related molecule sits very near the chain C binding pocket, potentially impacting on the binding of the ligand.

3.2.7.4 Ca4

No density could be visualised for the presence of Ca4 in either data set. Ca4 has been suggested to be linked to the levels of calcium during the cryo-protection stage, with a concentration of Ca^{2+} of 2 mM or above needed for binding in the neck funnel (**Shrive et al., 2003**). For both RS32 A12 and RS34 B21 crystals, strategy 2 was used as the cryo-protection of choice. When using cryo-protection strategy 2, the final concentration of Ca^{2+} achieved (3.33 mM) should have been enough for the Ca4 to be present, suggesting that the presence/absence of Ca4 is not exclusively dependent on the final concentration Ca^{2+}.

3.2.7.5 *BLEND* software – Merging RS32 A12 and RS34 B21 into one data set

BLEND software is part of the suite of programs present in the CCP4 suite, allowing for the management and merging of multiple data sets into a single merged data set. *BLEND* software is used following the integration and indexing of the spots and images by *iMOSFLM*. In this case RS32 A12 and RS34 B12 were integrated and indexed as reported in 3.2.5.2. *BLEND* software can be used to offset challenging problems when merging various data sets into a single one, with the *BLEND* software facilitating the task of building a complete data set from partial or separate data from multiple crystals, by merging these together. *BLEND* works by optimally assisting in the selection of optimal groups of data sets from multiple crystals prior to scaling, by using cluster analysis. Cluster analysis is a technique often used in statistical data analysis, whereby groups, or in this case data sets, are grouped together in a cluster according to how similar they are with respect to other data sets in other clusters. Both the data sets RS32 A12 and RS34 B21 were very similar with the same molecular arrangement, making the process of grouping these together very easy (**Foadi et al., 2013**).

BLEND software works in two separate stages. The first stage is the analysis mode where both data sets are analysed individually and tested for overall radiation damage. At the first stage the user can choose the space group and in this case $P42_12$ space group number 90 was chosen. The analysis mode produces a variety of output files including Blend_Summary.txt, containing tabulated information for all datasets being processed with each data set given a serial number, and a Clusters.txt file containing information about the Linear Cell Variation (LCV). The LCV is an important analysis of the merged data and a good indication of isomorphism (similarity of crystalline form between two separate data sets) with a value below 1% indicating a good degree of isomorphism. The analysis mode of the *BLEND* software for both RS32 A12 and RS34 B21 produced a LCV of 0.07% showing a high degree of isomorphism (**"CCP4 Program Suite: *BLEND*"**).

The second stage is the combination mode where the information from the stage analysis mode is used and the data is clustered together. This produces a variety of Output files including Merging_Statistics.info, Unscaled_001.mtz, Scaled_001.mtz and an Aimless_001.log. The combination mode offers the possibility of running *AIMLESS* in combination with the *BLEND* program producing a scaled mtz, or the possibility of using the unscaled mtz and manually running it through *AIMLESS*. In this experimental work the unscaled mtz was used in various runs of *AIMLESS* with the data limited to 3.10 Å. *TRUNCATE* and *UNIQUEIFY* programs were used in order to prepare the merged data to be used in *REFMAC5* and for the production of the electron density map, see **Table 3.12** for the data analysis crystallographic parameters of the *BLEND* merged data.

3.2.7.5.1 Electron density map of *BLEND* merged data (RS32 A12 + RS34 B21)

The electron density map for the *BLEND* merged data was produced as previously reported in 3.2.7 where the direct ($2F_{obs}$ - F_{cal}) and the difference map (F_{obs} - F_{cal}) were initially created by using the rigid body refined ligand-free rig_rs34 model for the phases. The R-factors for the rigid body movement for the *BLEND* merged data were 36.37% for

the R_{free} and 38.46% for the R-factor (for the two individual data sets: RS32 A12 R_{free} – 36.91% R-factor – 36.14%; RS34 B21 R_{free} – 38.33% R-factor – 36.72%).

The electron density of the *BLEND* merged data failed to show any new density not present in the electron density maps of the individual data sets, RS32 A12 and RS34 B21. The problem in this case when merging both data sets together is that the RS34 B21 electron density map (in terms of ligand occupancy) is a significantly better data set than RS32 A12. When merging both data sets some of the limitations seen in the RS32 A12 data set will be merged with the RS34 B21 better data, yielding in this case a higher R_{merge} for the merged data at 3.1Å. As *BLEND* merged data failed to provide extra new density and actually produced a less informative map than RS34 B21 in terms of bound ligand, no further work was performed using the *BLEND* merged data, and instead the RS34 B21 was chosen as the only data set to be used to display the experimental findings and for model building.

Data collection Parameters	Merged data - Blend		
Overall, InnerShell and OuterShell	Overall	InnerShell	OuterShell
Wavelength (Å)	0.92000		
Low Resolution limit (Å)	77.03	77.03	3.31
High Resolution limit (Å)	3.10	8.77	3.10
Space Group	P4 2_1 2		
Unit Cell lengths (Å)	a=112.48, b=112.48, c=105.70		
α, β, γ (°)	90, 90, 90		
Number of reflections	219585	9896	39120
Unique reflections	12853	657	2287
Average I/σ	15.3	32.8	6.0
CC (1/2)	0.994	0.993	0.945
Completeness %	99.9	99.3	100
$R_{merge\ \%}$	17.5	6.5	56.6
Multiplicity	17.1	15.1	17.1
Fast_dp (DLS) or processed in house	Processed in house		

Table 3.12 Data analysis parameters, following in-house merging of the RS32 A12 and RS34 B21 data sets through the *BLEND* and *AIMLESS* software.

3.2.8 Restrained refinement

Restrained refinement was carried out for RS32 A12 and RS34 B21. RS34 B21 offered the best electron density map, resolution and data to illustrate the best estimate of the rfhSP-D bound to *H. influenzae* CA7 PS, in terms of ligand orientation and protein structural organisation. It is important to mention that all the findings reported below for the RS34 B21 electron density map were checked against the RS32 A12 data, with the two data sets showing similar electron density shape, suggesting similar ligand orientation, preference and protein structural organisation. As both maps share common features only the RS34 B21 data set will be discussed below.

The restrained refinement is the second stage of refinement, where rigid body refined models are refined, while maintaining good geometry. The RS34 B21 best model was completed using *REFMAC5* and *COOT.* R-factors improved slightly from the initial rigid body refinement for the final model of RS34 B21 and were 20.69% for the R-factor (or R_{work}) and 25.21% for the R_{free}. The Ramachandran analysis showed that 95.96% (*MOLPROBITY*) of the residues modelled were in the preferred regions. No waters were modelled as this resolution makes it very difficult to determine the waters location. Some positive density in the map was not modelled, such as for weakly defined solvent molecules.

3.2.9 Molecular Model Building - RS34 B21

The initial rigid body density surrounding the binding pocket provided little evidence for the correct orientation of the ligand. In this case for CA7 PS, only three ligand carbohydrates are available to bind to CA1 and protein in the chain B binding pocket; HepI, HepII and Glc, represented in **Fig. 3.2**. Kdo has been reported to be a low affinity ligand for rfhSP-D (**Wang et al., 2008**).

Initially, both heptose and glucose carbohydrates were positioned according to previous reported binding orientations and hydroxyl locations in relation to Ca1, where the

heptose bound to rfhSP-D Ca1 via the hydroxyls O6' and O7' (**Wang et al., 2008**; **Clark et al., 2016**), while glucose bound by the equatorial hydroxyls O3' and O4' (**Shrive et al., 2003**).

Positioning both possible ligand carbohydrates in similar positions to those previously reported in the literature was carried out to ensure that the coordinating hydroxyls were located in the correct place with respect to Ca1. To date all the ligand bound rfhSP-D structures show fairly similar hydroxyl positions and distances between the Ca1 and the hydroxyls of the coordinating ligand. **Fig. 3.11** displays both heptose and glucose orientations as reported in the literature and how they fit into the density of RS34 B21, before any refinement attempt was performed.

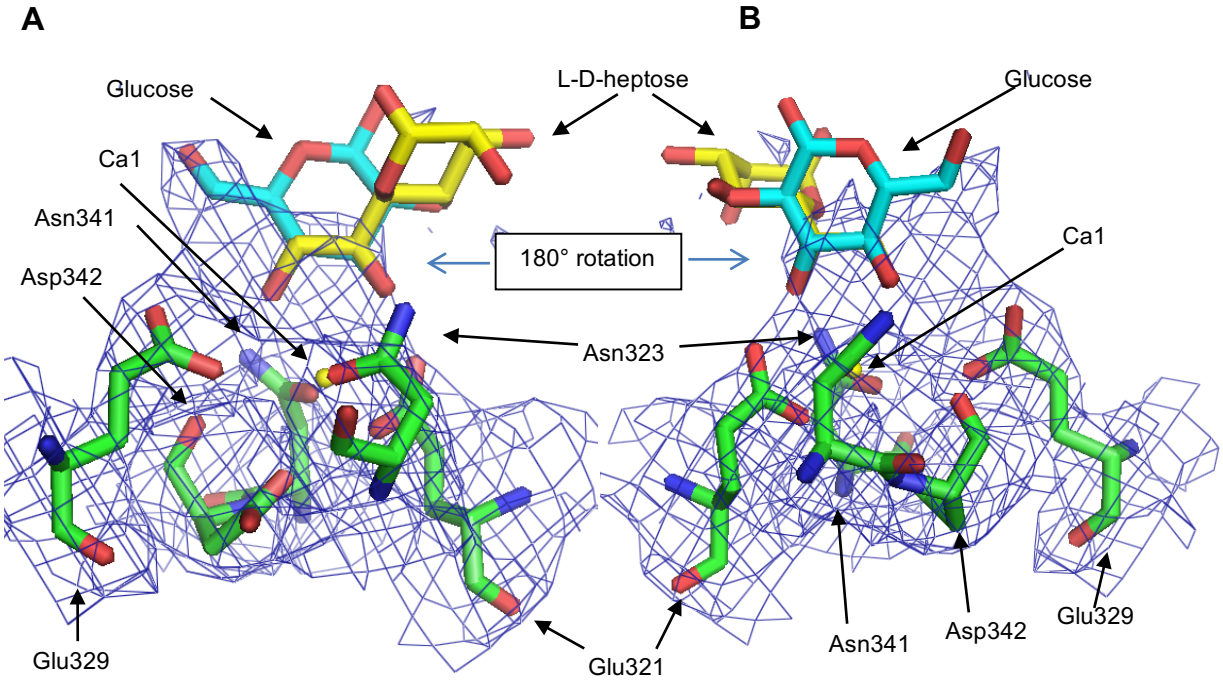


Fig. 3.11 The fitting of both a L-D-heptose and glucose as reported in literature (**Clark et al., 2016**; **Shrive et al., 2003**), PDB: 4E52 and 1PWB respectively and before refinement. In both figures Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342); glucose is displayed with the carbons in blue and with the oxygens in red; L-D-heptose is displayed with the carbons in yellow and the oxygens in red. The ligands in Figure A and B are rotated by 180° to each other. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. Figures A and B generated using *PYMOL* .

When considering which carbohydrate ring was coordinated to Ca1, glucose in the same orientation as that reported in the literature (**Shrive et al., 2003**; **Crouch et al. , 2006**) was used to attempt to explain the density surrounding the binding pocket (**Fig. 3.11**). The equatorial O4' hydroxyl group was placed in a similar location to that in literature and interacting with OE2 of Glu329 and ND2 of Asn341; the O3' hydroxyl group was placed near OE2 of Glu321 and ND2 of Asn323. The refinement of the glucose model against the electron density yielded a R_{free} of 25.26% (**Fig. 3.12 A**). A second ligand model was also tested whereby the glucose was rotated by 180° such that the positions of the hydroxyls O3' and O4' were interchanged, yielding a significantly worse map (data not shown).

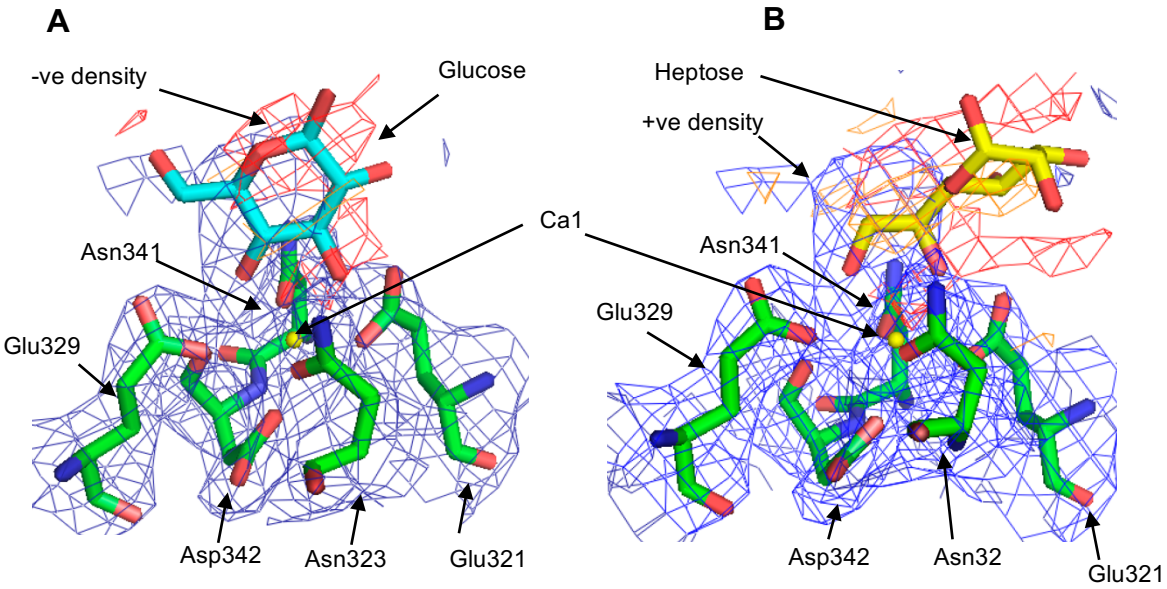


Fig. 3.12 The fitting of heptose or glucose in Chain B as reported in (**Clark et al., 2016**; **Shrive et al., 2003**), PDB: 4E52 and 1PWB respectively and subsequent refinement results. In both figures Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342); glucose is displayed with the carbons in blue and the oxygens in red; heptose is displayed with the carbons in yellow and the oxygens in red. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. The difference map (Fobs-Fcalc) is displayed in orange for positive (+ve) density and red for negative (-ve) density and contoured at 2σ. **A-** The glucose model produced an R_{free} of 25.26% with the suggestion that the glucose did not fit the well, with negative density features seen near C1' and O1'. **B-** The heptose model produced an R_{free} of 25.50%, with positive density suggesting that the O6' and O7' hydroxyls could be interchanged by rotation of the ligand in relation to the reported orientation. Figures A and B generated using *PYMOL.*

The heptose model was also positioned according to the orientation reported in the literature (**Clark et al., 2016**; **Wang et al., 2008**), with the O6' and O7' hydroxyls coordinating to Ca1 (**Fig. 3.11**). The refinement of the heptose model yielded a significant higher R_{free} than the glucose model, of 25.50%. The refined map of the heptose model displayed contrasts of positive density on the opposite side where the pyranose ring structure was present and negative density where the heptose (as reported) was placed, suggesting that the ligand could be rotated to interchange the O6' and O7' hydroxyls, changing the manner in which the ring is positioned (**Fig. 3.12 B**).

The refinement of glucose (**Fig. 3.12 A**) and heptose (**Fig. 3.12 B**) initially placed as reported in the literature does not seem to offer a good density fit between model and data. The heptose refinement displayed positive density opposite to where the pyranose ring was located, suggesting an interchange of the hydroxyls coordinating Ca1. A variety of models (20+) where the hydroxyls were interchanged and the sugar rotated by 180° in relation to the reported heptose orientation were placed in the density seen following rigid body refinement (**Fig. 3.13**), in order to investigate the possibility that the heptose hydroxyls were interchanged and this was refined against the data (**Fig. 3.14**). Of the various models attempted only two models are displayed in **Fig. 3.13**, model 1 and model 2 with the only difference between the **two m**odels being a 180° rotation about the C5-C6 bond.

The refinement of the heptose models (model 1 and model 2) as shown in **Fig. 3.13** produced improvements in the R_{free} with model 1 producing an R_{free} of 25.21% and model 2 25.22%. The density surrounding the hydroxyls and the adjacent C5 was good, however, the density surrounding the remainder of the ring was rather poor, making it very difficult to define a correct orientation for the ring and to discriminate between model 1 and model 2. The model 1 refined structure is displayed in **Fig.3.14**. The orientation of the O6' and O7' hydroxyls and the residues which they interact with are similar to both structures reported in the literature (**Wang et al., 2008**; **Clark et al., 2016**), suggesting that this

structure could offer evidence of a novel mode of orientation and binding for the heptose sugar. The minimal difference in R_{free} between model 1 and model 2 suggests that more than one orientation of the heptose ring could have bound to the rfhSP-D, causing the averaged density to produce a weaker density surrounding the pocket and creating problems when defining the right orientation.

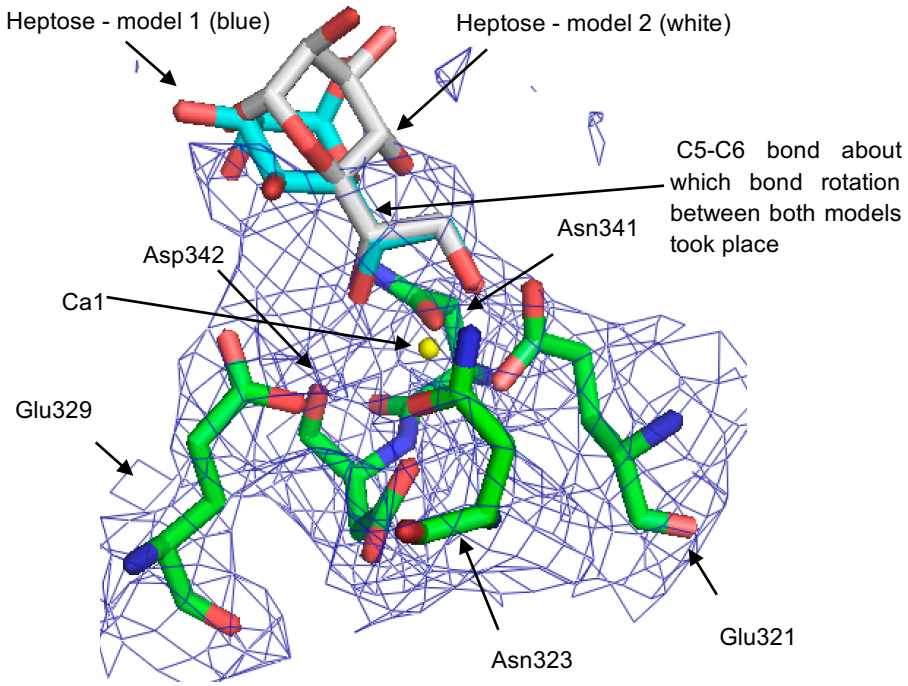


Fig. 3.13 The fitting of both models of the heptose where the hydroxyls are interchanged and the sugar rotated 180° in relation to the orientation reported in the literature in the rigid body refinement density (**Clark et al., 2016; Wang et al., 2008**), before any refinement was performed. In the figure Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342); heptose model 1 is displayed with the carbons in blue and the oxygens in red; heptose model 2 is displayed with the carbons in white and the oxygens in red, showing ≈ 180° ring rotation with respect to the C5-C6 bond. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. Figure generated using *PYMOL.*

Of all the models tested and refined and using the knowledge of the R_{free} and direct visualisation of the electron density maps for discrimination between models, the best model suggests that a heptose and not a glucose is bound to the rfhSP-D chain B binding pocket. The fact that the glucose did not fit the density as well as the heptose and gave a higher R_{free} is not surprising as recent work by Clark and co-workers showed that

the *H. influenzae Eagan* 4A mutant LPS binding to rfhSP-D occurred through the non-terminal heptose, even with a terminal glucose available for binding (**Clark et al., 2016**). A second point to consider and reported in 1.4.5.1 is that SP-D and rfhSP-D have been shown to have a higher affinity for heptose (L-D-Hep) than glucose, so it is not surprising to see the heptose bound instead of the glucose, when both sugars are available.

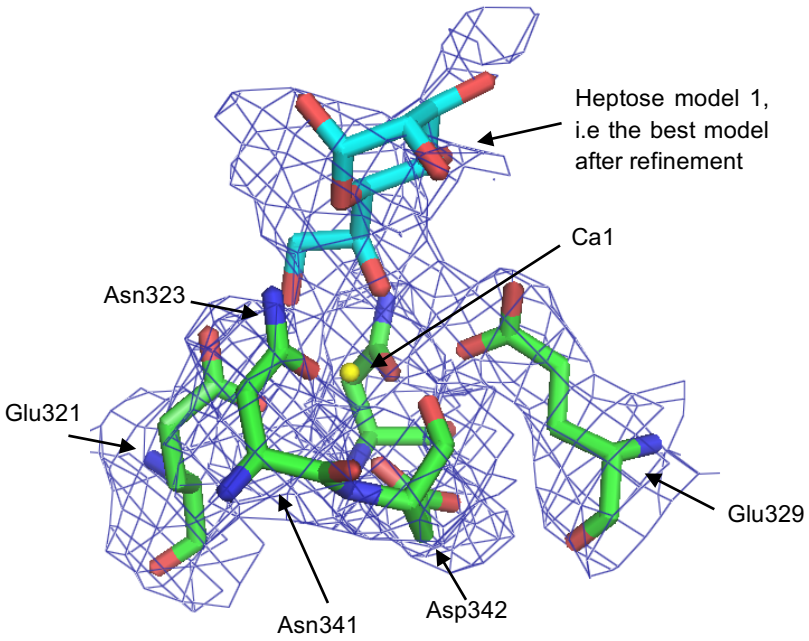


Fig. 3.14 The final model (model 1) displayed following refinement, with the heptose sitting 180° rotated in relation to previously reported structures (**Clark et al., 2016**; **Wang et al., 2016**), with the positions of the coordinating hydroxyls then interchanged. In the figure Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342). The heptose model is displayed with the carbons in blue and the oxygens in red. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. Figure generated using *PYMOL.*

3.2.10 Final Model

The electron density map of the ligand binding site in chain B (**Fig. 3.14 and Fig. 3.15**) suggests that the orientation and binding of heptose occurs by rotating the sugar by 180°about C5-C6 so that the positions of the hydroxyl groups O6' and O7' are interchanged in comparison to reported structures (**Wang et al., 2008**; **Clark et al., 2016**) as shown in **Fig. 3.14** and **Fig. 3.15**. The hydroxyl O6' in the novel orientation is interacting with Glu329 OE1 (2.78 Å) and Asn341 OD1 (2.93 Å) which are also

coordinating Ca1. The hydroxyl O7' in the novel orientation is interacting with Glu321 OE2 (2.30 Å). Further interactions seen include the heptose ring O5' coordinating with Asn341 ND2 (2.82 Å) and the O4' linkage which connects to the Glc possibly interacting with Asp325 (2.67 Å). The strong electron density surrounding C5 clearly suggests that the heptose is indeed rotated, with C5 coming off at a different angle from the reported structures. The reasons why the Eagan CA7 heptose shows a different orientation from previous heptose structures will be discussed in more detail in 3.3.6. The refinement statistics for the best model are given in **Table. 3.13**.

Refinement statistics	
Data Range (Å)	77.01 - 3.10
No. of protein atoms	3457
No. of residues chain A and B	205-355
No. of residues chain C	206-355
No. of water molecules	N/A
No. of calcium ions	2
No. of ligand atoms	14
R_{work} (%)	20.69
R_{free} (%)	25.21
RMSD bond length (Å)	0.015
RSMD bond angle (°)	1.9
Average B-values (Å^2)[a]	
Protein main chain	56.6
Water	N/A
Ligand	58.466
Ramachandran plot values (%)	
Favoured	95.96
Outliers	0.45

Table 3.13 The refinement statistics of the best model. [a] defined according to *Baverage* defined according to *MOLPROBITY*.

The best model suggests that the heptose C4 is directed towards the flanking residue Asp325 within hydrogen bonding distance and that heptose C2 is inclined towards the Arg343 flanking residue placing the C2-OH within hydrogen bonding distance of Arg343 NE, NH1 or NH2. The Arg343 orientation in the density suggest a novel orientation, where the Arg343 moves away from the C2 of the heptose in order to accommodate the sugar in its novel orientation, leading to the head of Arg343 directed towards the nearby Glu347. Glu347 in turn moves towards the Arg343 head group such that both residues are within distance to form a salt bridge between the Glu347 OE1/OE2

and Arg343 NH2. The novel orientations for both Arg343 and Glu347 provide additional strong evidence that the final model is correct.

Arg343 and Asp325 have been suggested in the literature to be important in discriminating glycoconjugates and to have the ability to impact on binding affinities. Arg343 presented in this model assumes a novel conformation previously unreported for the unliganded and liganded structures of rfhSP-D. In the reported structures Arg343 flanks the binding site with only slight differences in conformation reported, but always with the head group of Arg343 facing up, flanking the binding pocket (**Shrive et al., 2003**; **Crouch et al., 2006**; **Wang et al., 2008; Clark et al., 2016**). In the model presented here the density for the Arg343 shows density to suggest a novel conformation of Arg343, where the head group of Arg343 tilts away from its normal orientation, in order to accommodate the novel orientation of the HepI (**Fig. 3.16** and **3.17**). This leaves the C2 hydroxyl of Hep within hydrogen bonding distance of Arg343 NH2, NH1 and NE. When the Arg343 head group tilts away from the C2-OH of the Hep, it comes in close contact with Glu347, causing the Glu347 OE1 and OE2 to assume a preferred orientation in order to hydrogen bond with Arg343 NH2, OE1-NH2 and OE2-NH2 distances are 2.49 Å and 2.94 Å respectively and this is displayed in **Fig. 3 17**. The interaction reported between Arg343 and Glu347 is a novel interaction, with no previous structural work displaying such an interaction.

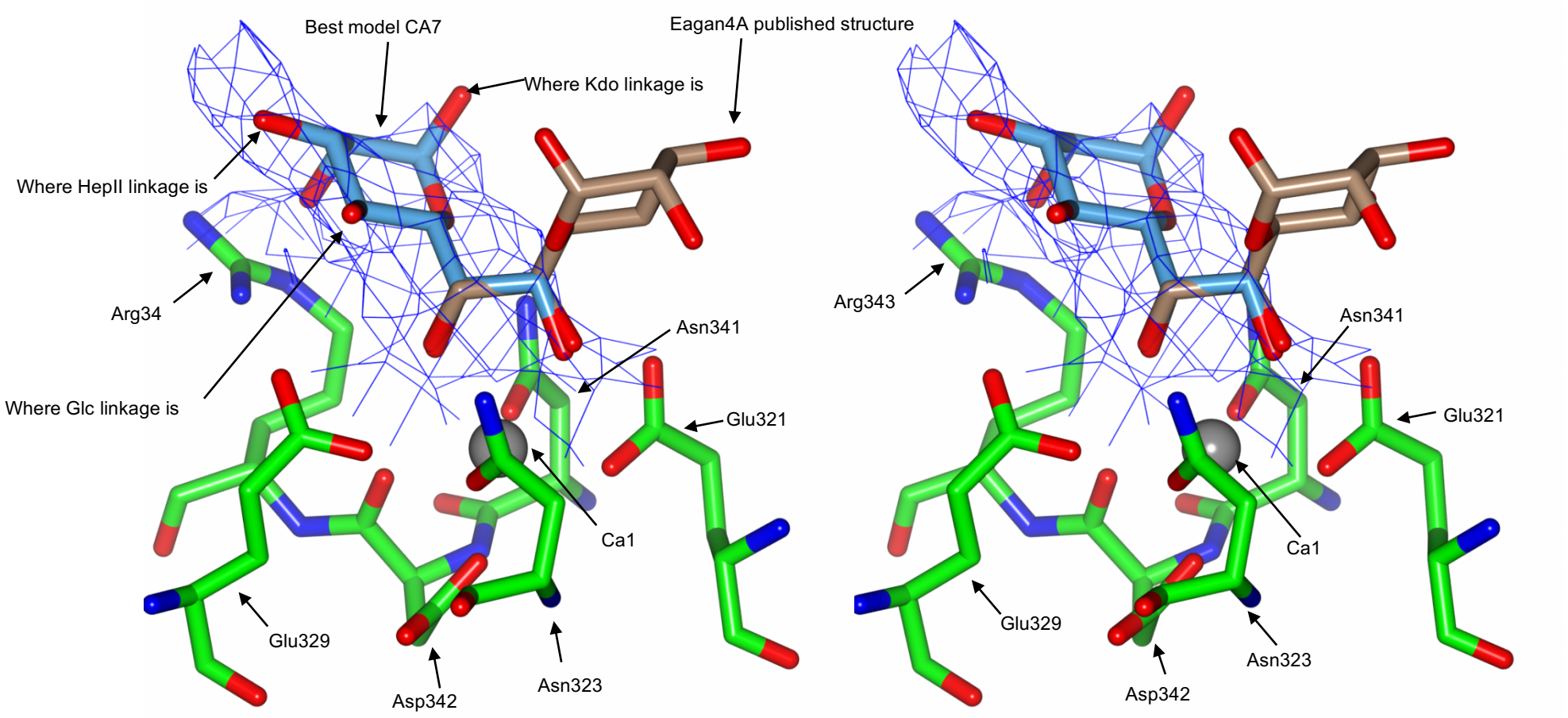


Fig. 3.15 The novel orientation of HepI with the Hep rotated by 180° resulting in the positions of the O6' and O7' hydroxyls being interchanged. The Arg343 head group moves away from C2-OH in order to accommodate the novel orientation of HepI. The positions where HepI would link to Kdo, Glc and HepII are shown. In the figure Ca1 is displayed as a grey sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342) and Arg343; heptose model is displayed with the carbons in blue and with the oxygens in red; *H. influenzae* 4A structure (PDB: 4E52) is displayed for direct comparison with the carbons in brown and the oxygens in red. Stereo image of the best model, 4A structure and direct map (2Fobs-Fccalc) in blue surrounding the HepI and contoured at 1σ, figure generated using Molecular Graphics CCP4 (CCP4mg).

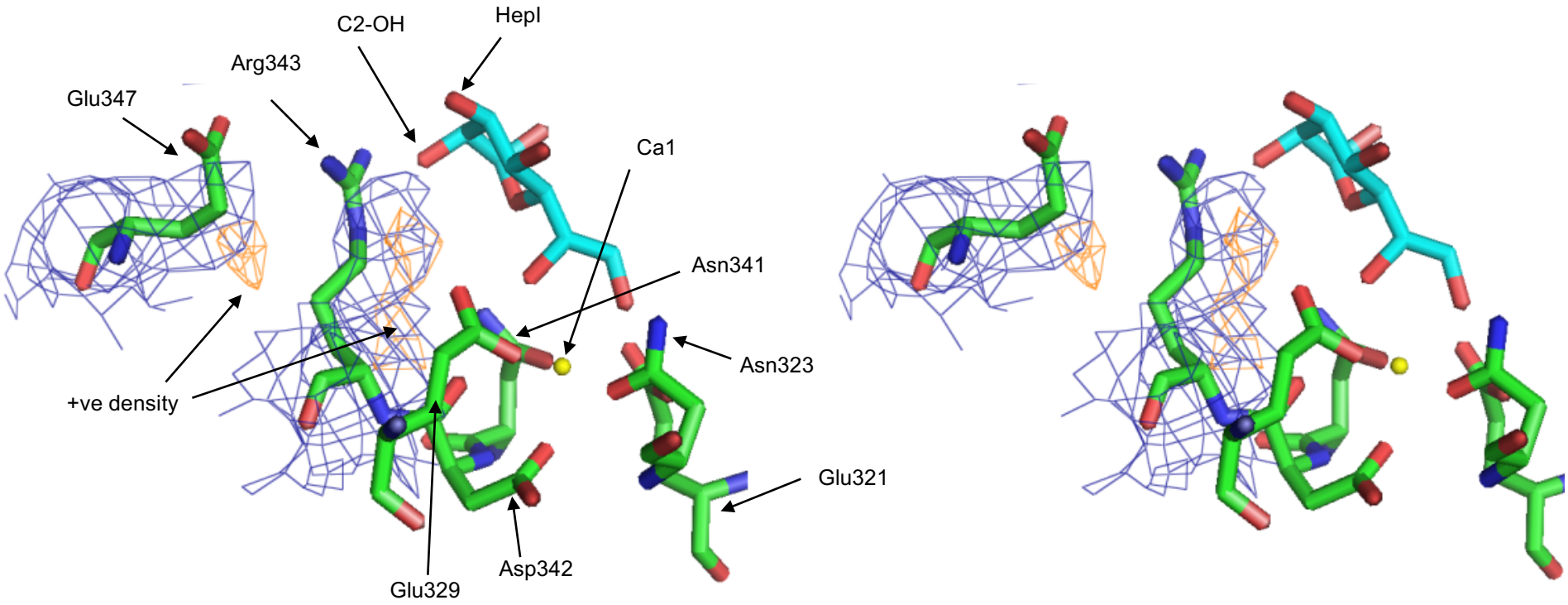


Fig. 3.16 An illustration of the shifts seen in Arg343 and Glu347, in order to accommodate the novel orientation of the Hep. In the figure Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342); heptose model is displayed with the carbons in blue and with the oxygens in red. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ, while the difference map (Fobs-Fcalc) positive (+ve) density is displayed in orange and contoured at 2.5σ. The Arg343 and Glu347 residues are located near positive density suggesting a different conformation for both residues. Stereo image generated using *PYMOL*.

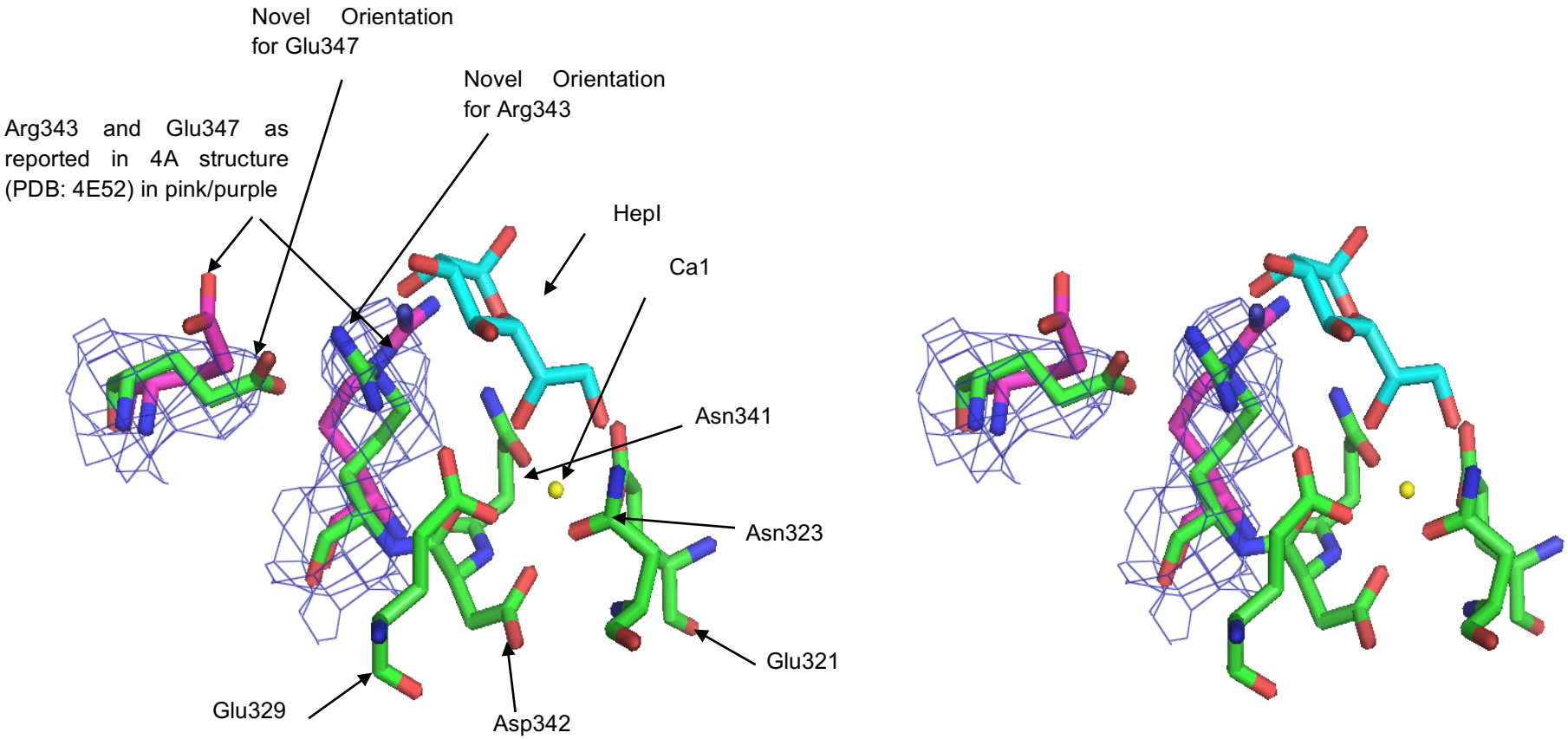


Fig. 3.17 An illustration of Arg343 and Glu347 after refinement, with both residues in the novel orientation. NH2 of Arg343 is within distance to hydrogen bond to Glu347 OE1 (2.49 Å) and OE2 (2.94 Å); a salt bridge between Arg343 and Glu347 is suggested. In pink/purple Arg343 and Glu347 are displayed as reported in the 4A *H. influenzae* structure (PDB: 4E52) (**Clark et al., 2016**) with residues in pink (carbon), blue (nitrogen), red (oxygen) showing the Arg343 and Glu347 in the reported structure. In the figure Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342) with the carbons (green), oxygen (red) and nitrogen (blue); heptose (Hep1) model is displayed with the carbons in blue and with the oxygens in red. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. The interaction suggested between Arg343 and Glu347 represents a previously unseen and unreported interaction, a consequence of a novel orientation of binding by HepI. Stereo image generated using *PYMOL.*

The best model and density suggests that HepI and not HepII is bound, with density coming off the C3 of the best model and where HepII would be, with the electron density map showing space for the Kdo at position 1 and Glc at position 4, but with no density seen for either sugar suggesting conformational freedom about the glycosidic linkage. As illustrated in **Fig. 3.2** HepI is linked to Kdo at position 1, HepII at position 3 and Glc at position 4, while HepII is linked to HepI at position 1 and a PEA group at position 6. The best model and density suggest that HepI is binding with some density coming off C3, suggesting that only HepI could bind this way. The rotation of the ring of the Hep in order to check if HepII was binding and subsequent best fit into the density around C1 failed to produce adequate density for the ring. A second point to consider is that HepII contains a PEA group at position C6, which to the best of our knowledge could be present or absent, as the full impact of hydrolysis on the phosphate derivatives remains incomplete. If the PEA was present, HepII could not bind through the O6' and O7' hydroxyls. It is important to mention that even though the best model suggests an orientation of binding, this does not exclude that other orientations of the main ring heptose could have bound, but with the density shown it was impossible to provide evidence for other orientations of the main ring heptose.

A novel orientation of heptose (HepI) binding to the rfhSP-D is therefore proposed and the model illustrated in **Fig. 3.18** and **Table 3.14** and summarised as follows; the electron density suggests that a heptose sugar rotated by 180° with respect to the examples in the literature, with the positions of hydroxyls O6' and O7' interchanged offers the best fit; the electron density present at the C3 linkage suggests that only HepI could sit in this density, as HepII contains no C3 linkage; rotating the HepII ring in order to fit C1 of HepII near C3 of HepI failed to produce adequate density, suggesting that HepI and not HepII is bound; finally the changes seen in the Arg343 and Glu347 residues, where Arg343 moves away from C2 of the heptose, coming into proximity of Glu347 which in

turn moves towards Arg343, leaving both residues within adequate distance for a salt bridge, gives strong evidence to support the model presented here.

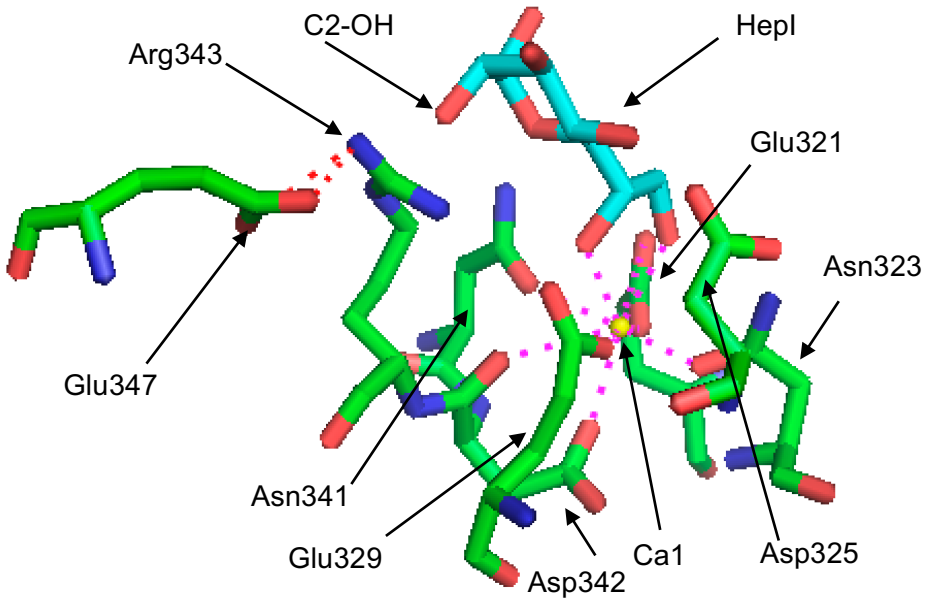


Fig. 3.18 The proposed novel orientation of HepI bound in the rfhSP-D binding pocket chain B. In the model the C2 of the HepI occupies the position where normally Arg343 would be located, causing the Arg343 to move away into proximity of Glu347. Glu347 in turn moves towards Arg343 and is within hydrogen bonding distance. C2-OH of the HepI is within hydrogen bonding distance of NE, NH1 or NH2 of Arg343, with Glu347 OE1 and OE2 within distance for a salt bridge with Arg343 and displayed with red dashed lines. In the model Ca1 is seen in yellow sphere with the coordinating residues (Glu321, Asn323, Glu329, Asn341 and Asp342) and Asp325 also shown. Pink dashed lines demonstrate the interaction of Ca1 and coordinating residues. Image generated using *PYMOL.*

Atom 1	Ca1 coordinating residues/ligand	Chain B (Å)-Best model	Chain C (Å)-Best model	4A Chain B (Å)	4A Chain C (Å)
Ca1	Glu321 OE1	2.34	2.71	2.57	2.55
Ca1	Asn323 OD1	2.33	2.38	2.42	2.37
Ca1	Glu329 OE1	2.84	2.43	2.44	2.31
Ca1	Asn341 OD1	2.58	2.33	2.40	2.42
Ca1	Asp342 OD1	2.35	2.35	2.27	2.25
Ca1	Asp342 O	2.94	2.64	2.56	2.59
Ca1	Hep Hydroxyl - O6’	2.35	N/A	2.42	2.44
Ca1	Hep Hydroxyl - O7’	2.37	N/A	2.58	2.53

Table. 3.14 Ca1 and coordinating residues interatomic distances for RS34 B21 Chain B and C. For RS34 B21 chain B the distances to the Hep hydroxyls O6’ and O7’ are the distances calculated for the best model and shown in **Fig. 3.14** and **Fig. 3.15**. For comparison, the *H. influenzae* Eagan 4A bound structure (PDB: 4E52) chains B and C are shown (**Clark et al., 2016**). No Ca1 was visualised for Chain A in the RS34 B21 data and so no interatomic distances could be calculated. Chain C contained Ca1 but no ligand bound with the N/A implying no ligand presence.

3.3 Discussion

3.3.1 Co-crystallisation of rfhSP-D in combination with CA7 PS

The co-crystallisation of rfhSP-D in combination with CA7 PS produced many crystals in the structure screens SS1 and SS2, with RS20 A5 (SS1 C.5), RS21 A4 (SS1 C.28), RS21 C4 (SS1 C.40) and RS21 D2 (SS1 C.44) showing positive crystal growth. These conditions produced a variety of morphologies including thin squares, triangles and non-square plates. The squares produced in SS1 and SS2 resembled the ligand-free rfhSP-D crystals normally produced in our research group, with the only difference being that these were rather thin, problematic when using X-ray radiation.

3.3.2 Follow up crystals

Three follow ups were produced from SS1 and SS2 conditions; RS20 A5 – RS32, RS21 C4– RS33 and RS21 D2 – RS34, where the initial conditions were varied slightly, in order to optimise the crystal growth. From the three follow ups produced, only RS32 and RS34 tray produced positive protein crystals. Both of these conditions were fairly similar, with both conditions containing ammonium sulphate, suggesting that a high concentration of ammonium sulphate >1.8 M, favours the formation of rfhSP-D co-crystals bound to CA7 PS. RS32 and RS34 trays produced similar morphologies to the initial SS1 conditions, where most of the crystals formed resembled thin squares however, most often layered. Single squares were often the crystal morphology of choice looped for X-ray diffraction. Both trays were successful in growing a myriad of crystals, with easily over 200 crystals produced in both wells. Unfortunately, these were often multiple (twinned) and odd-shaped, with limited number of single crystals formed.

All 42 crystals tested from the follow ups were shown to be protein, offering several proteins co-crystals to be used in X-ray diffraction in the hope of achieving a high resolution (< 2 Å) and accurately characterise the binding of CA7 PS to rfhSP-D. Unfortunately, the resolution seen led to some ambiguity in the characterisation of rfhSP-

D and CA7 PS. The resolution of 80% of the crystals tested was very low >6 Å, with the resolution of the remaining 20% of the crystals tested and data collected, often around the 3.05 - 4.00 Å, which is considered rather low in accurately characterising ligand binding studies (**Evans et al., 2011**).

3.3.3 Data Analysis of promising crystals

Out of the 53 crystals of rfhSP-D co-crystallised in complex with *H. influenzae* PS CA7 and tested at DLS, only 2 crystals achieved an acceptable resolution to be used in the characterisation of structural ligand studies. Of the 53 crystals tested 90% were protein, even though these were often poor diffracting crystals, suggesting that conditions used during crystallisation favoured crystal growth. However, there is evidence to suggest a high internal disorder upon formation of the crystals, especially when we consider that ligand-free rfhSP-D crystals have been grown in our research group and diffracted to 1.4Å.

3.3.4 Data processing and analysis

RS32 A12 and RS34 B21 electron density maps were both used for model building, as these offered the best overall resolution limited to 3.10 Å and 2.98 Å respectively. Both crystals were indexed and integrated in house using *iMOSFLM*. *AIMLESS* was used to scale and merge the multiple observations of reflections together into an average intensity. The use of the *AIMLESS* software was particularly important in modifying the statistical parameters, as it allows batches of images to be selected on the basis of acceptable statistical parameters and R_{merge} for integration in *iMOSFLM*

Initially for both RS32 A12 and RS34 B21 the resolution used was 3.3Å and 3.1Å respectively. Batch selection was carried out to achieve a R_{merge} overall ≈ 12% and outer shell of ≤40% for both data sets (data not displayed) and a final resolution of 3.3 Å and 3.10Å respectively. There are conflicting views in the literature on role of R_{merge} as the criterion for choosing the cut-off resolution (**Evans, 2011a**; **Wlodawer et al., 2008**). The

R_{merge} has the disadvantage that it increases with the data multiplicity, although the more observations that are averaged the better the merged data (**Evans, 2011a**). In the highest resolution shell the target R_{merge} has been reported as 30-40% for low symmetry crystals and up to 60% for high-symmetry crystals (**Wlodawer et al., 2008**). This was the main idea when aiming for a R_{merge} ≤40% in the outer shell, however the $P42_12$ tetragonal space group is a high symmetry space group with eight symmetry operations possible and yielding high multiplicity, suggesting that the R_{merge} could be allowed to reach 60% in the outer shell. Not over-limiting the data resolution allowed a higher resolution model with a lower R_{free} and R_{free}-R_{work} gap, all important when defining the ligand and its correct orientation.

3.3.5 Electron Density Maps

The RS32 A12 and RS34 B12 electron density maps were used in model building in the binding pocket of chain B. Both electron density maps were created using a rfhSP-D model identical to that in the unliganded $P2_1$ space group (**Shrive et al., 2003**). The electron density maps shown in **Fig. 3.9**, were very similar at 1σ and both datasets were merged in to one to produce an electron density map of the merged data. *BLEND* was used to merge RS34 B21 and RS32 A12 after *iMOSFLM*, into one merged dataset. The electron density map of the merged data was unfortunately very poor and did not add any novel information not seen in the RS32 A12 and RS34 B21 individual electron density maps. Of the two, the RS34 B21 map appeared somewhat better than the RS32 A12 map and the RS34 B21 electron density map was used to model the CA7 PS bound to rfhSP-D. The best model was also checked against the RS32 A12 data, with similar findings to those reported for RS34 B21 data.

3.3.6 The final model

Chain A, B and C showed the overall structure of rfhSP-D similar to that reported in literature for the $P2_1$ space group (**Shrive et al., 2003**; **Shrive et al., 2009**; **Wang et al.,**

2008), with the trimeric aggregate with the three C-terminal globular domains (236-355) and the three chains linked through the α-helical coiled-coil neck region (203-235). The chain A and C binding pockets are located in the vicinity of a symmetry related molecule with little or no space to accommodate the long ligands tested, which explains why no binding was seen. Chain A is particularly tight with the symmetry related molecule facing the binding pocket in chain A. For the chain C binding pocket there is more room than in chain A, however it seems to hinder carbohydrates larger than monosaccharides, with only space to accommodate one carbohydrate. Ligand binding was only seen in chain B, but again the symmetry related molecule is present in vicinity of the binding pocket, potentially impacting on the orientation of longer bound ligands.

The lack of Ca1, Ca2 and Ca3 in chain A, Ca2 and Ca3 in chain B and Ca2 and Ca3 in chain C and no Ca4, is somewhat surprising (note that this is also seen for the ligands reported in chapter 4 for the P4 2_1 2 co-crystallised rfhSP-D). At first sight it seems probable that the loss or absence of these calcium ions was in part due to a low concentration of Ca^{2+} during the cryo-protection stages. However, the finding that $P2_1$ crystals (reported in chapter 4), produced using the exact same method of co-crystallisation showed Ca1, Ca2 and Ca3 present in all chains suggests that other variables are also important. As the loss of Ca1, Ca2, Ca3 and Ca4 is also reported in chapter 4, an overall discussion and hypothesis for the reason why we saw the absence of the Ca^{2+} will be given in chapter 5.

Modelling the most likely ligand and the correct orientation that had bound, into the binding pocket density in chain B, required the investigation of many different possibilities. *H. influenzae* CA7 PS provides four possible epitopes for recognition by rfhSP-D: 1- through the Kdo linked to the lipid A by C7-OH (O7') and C8-OH (O8'); 2- through the inner core HepI by C6-OH (O6') and C7-OH (O7'); 3- through the terminal Glc by C3-OH (O3') and C4-OH (O4'); 4- through the HepII bound to HepI by means of C3-OH (O3') and C4-OH (O4'). If the PEA group was lost during the hydrolysis process HepII O6' and O7'

would be available to bind rfhSP-D however, at present it remains unknown if following hydrolysis the PEA group would be present or absent. Kdo has been reported to be a very low affinity ligand for rfhSP-D and so it was discarded as the potential sugar bound to Ca1 in rfhSP-D chain B, leaving HepI, HepII and Glc as potential ligands (**Wang et al., 2008**) **Fig. 3.19** displays the possible binding sites of the *H. influenzae* CA7 PS, excluding the Kdo. High resolution structural studies of rfhSP-D complexed with glucose (**Shrive et al., 2003**; **Crouch et al., 2006**) and heptose (**Wang et al., 2008**; **Clark et al., 2016**) have been reported in literature, allowing for a direct comparison between the possible ligand bound models and those reported in the literature.

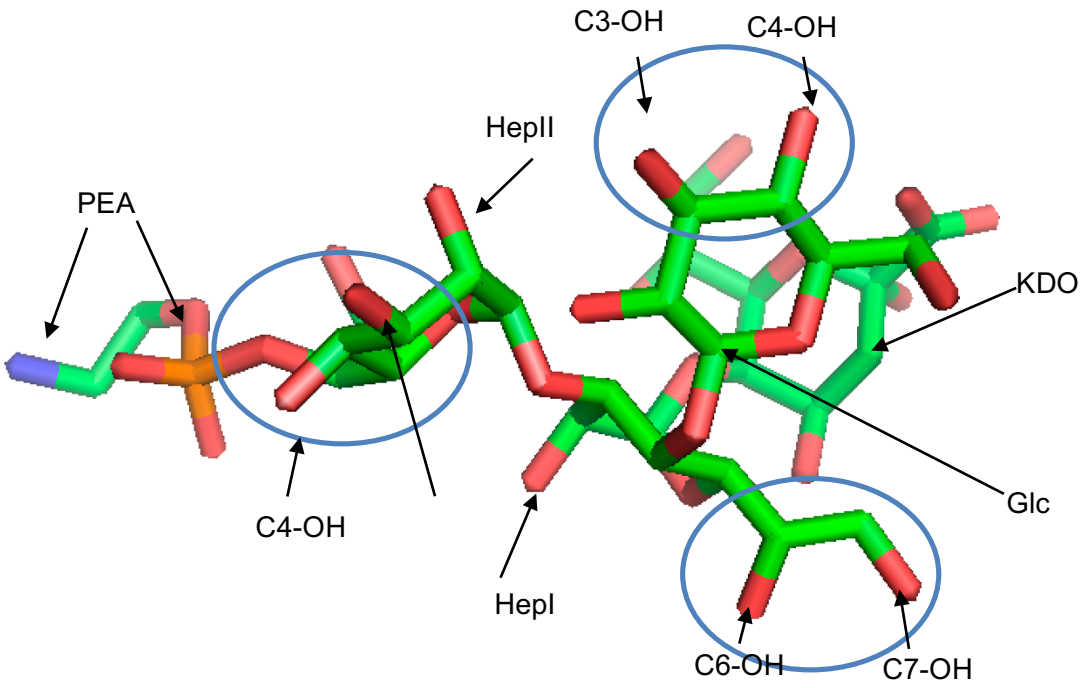


Fig. 3.19 An illustration of the *H. influenzae* CA7 PS molecule and possible sites of binding to rfhSP-D using current knowledge and reported data. rfhSP-D has been reported in the literature to bind glucose (Glc) through the equatorial C3-OH (O3') and C4-OH (O4') hydroxyls (**Shrive et al., 2003**); rfhSP-D has been reported to bind to heptose through the C6-OH (O6') and C7-OH (O7') in the L-D-heptose, while the D-D-heptose isomer has been shown to not bind through the O6' and O7' hydroxyls, but instead through the equatorial O3' and O4' hydroxyls, in similar manner to glucose (**Clark et al., 2016**; **Wang et al., 2008**). Both L-D and D-D heptoses contain a mannose ring structure, near identical to glucose (only C2-OH different - C2 epimers). The Kdo has been reported to be a very low affinity ligand and so it was discarded as a potential ligand. Figure generated using *PYMOL.*

Previous structural work and binding affinity studies have shown that SP-D and rfhSP-D binding to LPS is Ca^{2+} dependent, with the recognition dependent on the binding of terminal residues (such as glucose) or inner core residues (such as heptose). Kuan and co-workers reported that SP-D displayed poor binding of longer Ra mutants and smooth LPS strains, with the suggestion that long LPS can "mask" the binding sites associated with preferred inner core binding. The shorter LPS mutants Rc and Rd of *S. enterica* minnesota demonstrated a higher binding by SP-D than the Ra mutant rough LPS, suggesting that inner core residues such as heptose and terminal core glucose are the preferred binding epitopes (**Kuan et al., 1992**).

SP-D and rfhSP-D binding affinities have been investigated using different mutants of purified *H. influenzae* Eagan LPS and whole bacteria, *H. influenzae* Eagan 4A LPS (1 heptose), *H. influenzae* Eagan CA7 LPS (2 heptoses) and *H. influenzae* Eagan Wild-type LPS (3 heptoses). Both SP-D and rfhSP-D showed a high binding affinity for 4A LPS and whole bacteria, with subsequent increase in the number of heptoses and branching yielding poorer binding, 4A> CA7> Wild-type. (**Clark et al., 2016**).

Crystallographic studies using delipidated *H. influenzae* Eagan 4A LPS soaked into rfhSP-D ligand-free crystals, reported that the PS bound to the rfhSP-D via the heptose and through the hydroxyls O6' and O7', even though terminal glucose was also available for binding (**Clark et al., 2016**). The binding of the heptose in the *H. influenzae* 4A PS was reported to be similar to the heptose previously reported by Wang and co-workers, with the only difference being that *H. influenzae* Eagan 4A PS HepI showed a ring rotation away from flanking Asp325 (**Wang et a., 2008**; **Clark et al., 2016**). The rfhSP-D bound to *H. influenzae* 4A study offers a direct comparison to the experimental work reported here, extending the knowledge of how different mutant *H. influenzae* LPS binds to rfhSP-D, with the only difference being that the *H. influenzae* CA7 LPS mutant contains an extra heptose (HepII) phosphorylated at position 6.

In order to correctly characterise the ligand binding in pocket chain B a variety of models with different ligands of Hep or Glc were used, with different orientations (hydroxyls rotated) tested. The glucose model as reported in the literature (**Shrive et al.,** 2003) yielded a R_{free} of 25.26%; the heptose model as reported in the literature (**Clark et al., 2016**; **Wang et al., 2008**) yielded a R_{free} of 25.50%. The novel orientation of heptose, where the O6' and O7' hydroxyls were interchanged and rotated by 180° yielded 25.21% in model 1 and 25.22% in model 2. The difference between model 1 and model 2 lies on rotation of the C5-C6 bond, with the minimal difference seen in the R_{free} suggesting more than one possible orientation of the ring rotation could have bound. Therefore, taking into consideration the R_{free} and direct visualisation of the difference map, the best model tested suggested that the *H. influenzae* CA7 PS bound to the rfhSP-D through a heptose by the O6' and O7' hydroxyls. Surprisingly the heptose bound to Ca1 assumes a different conformation to that in the Eagan 4A PS, with the heptose rotated such that the positions of the O6' and O7' hydroxyls are interchanged. Discriminating between model 1 and model 2, see **Fig 3.13** for reference, as the best model solely using the R_{free} is not possible, however supporting model 1 (i.e the best model) are two novel orientations of two residues; the Arg343 flanking residue and the nearby Glu347. Both residues assume a novel orientation and conformation in order to accommodate the novel orientation of heptose.

The Glu347-Arg343 interaction is clearly defined by strong density (**Fig. 3.16** and **Fig. 3.17**) suggesting that in order to accommodate the novel orientation of heptose, the flanking Arg343 residue needs to move away from heptose C2-OH. If Arg343 did not assume this novel orientation the heptose could not bind using the orientation reported. Asp325, the other flanking residue, is directed towards heptose C4-OH, again within distance for hydrogen bond potential.

Considering that the *H. influenzae* Eagan CA7 PS contains two heptoses, HepI and HepII (if the PEA group of HepII was lost during hydrolysis) that could bind through

the O6' and O7' hydroxyls, it was important to determine the most likely heptose binding. According to the electron density map and best model, the heptose with the C2-OH facing the Arg343 group fits well with the electron density seen around heptose C3-OH. The C3-OH position can be explained only for HepI binding to Ca1. The HepII-HepI linkage is α1-3, indicating that the electron density seen around C3-OH represents the linkage between HepI-HepII with HepI bound to rfhSP-D. Rotating the main ring structure in order to fit C1-OH into the density, where C3-OH was previous located, failed to provide a good fit, again indicating HepI rather than HepII binding. A second point to note when considering if HepII is bound is the presence of the PEA group at position O6' of the HepII. The hydrolysis procedure has been reported to impact on the presence of phosphate groups and derivatives attached to the Kdo however, it remains unclear if it could also affect phosphate groups attached to the inner core heptoses (**Danan et al., 1982**; **Auzanneau et al., 1991**; **Masoud et al., 1997**). Therefore, at present, the full impact of mild acid hydrolysis on the PEA group present on HepII remains unknown, suggesting that the PEA may or may not be present. HepI is the same heptose reported for the shorter *H. influenzae* Eagan 4A PS mutant (**Clark et al., 2016**), which raises the questions why has HepI assumed this novel orientation, rotated by 180°, and how does this correlate with the higher binding affinity for Eagan 4A LPS (1 heptose) as compared to Eagan CA7 LPS (2 heptoses) (**Clark et al., 2016**).

In order to attempt to answer both questions, four ligand models were constructed; **Model A**- *H. influenzae* Eagan 4A PS model containing the Kdo-HepI-Glc with HepI bound as reported in literature (**Clark et al., 2016**); **Models B** and **C**- *H. influenzae* Eagan CA7 PS also containing the Kdo-HepI-Glc, but with an extra HepII bound to HepI by an α1-3 linkage and phosphorylated at position 6 by a PEA (Model B- HepI O6' and O7' hydroxyls positioned similarly to those reported for 4A PS; Model C- the HepI rotated by 180° with respect to Ca1, i.e the best model for CA7 PS); **Model D**- *H. influenzae* Wild-type Eagan PS (incomplete) containing the CA7 core plus a phosphorylated HepIII at

position 4 bound to HepII by α1-2 and a GlcI bound to HepII by α1-3 (HepI O6' and O7' bound and in similar orientation to model C for CA7 PS). Models A, B and C were used in attempting to answer questions 1, while models A, B and D were used in attempting to answer question 2.

3.3.6.1 Question 1 – why is the Eagan CA7 HepI rotated by 180° compared to the Eagan 4A HepI?

Question 1 explains why the novel orientation of HepI Eagan CA7 PS has occurred, with HepI rotated by 180° in comparison to the reported *H. influenzae* Eagan 4A PS HepI. It is important to understand that the rfhSP-D and CA7 complex was formed by co-crystallisation which implies no crystal packing constraints with the CA7 ligand free to assume its preferred and most stable orientation, with subsequent co-crystallisation. The fact that the CA7 is only found in the B subunit, and that when bound it does not mimic the high affinity binding shown by the Eagan 4A mutant (**Clark et al., 2016**) either here or presumably in solution, suggests that this is not quite so straightforward as it first appears. Three models were used in the attempt to answer the above question – model A – containing the 4A PS as reported in literature (**Clark et al., 2016**); model B – containing the CA7 PS HepI in similar orientation seen in 4A and model C- containing the CA7 PS HepI rotated by 180° so that the positions of hydroxyls O6' and O7' are interchanged with respect to Ca1.

Model A and model B were constructed according to the position of the anhydro Kdo and HepI reported in the literature for the Eagan 4A (**Clark et al., 2016**). In this orientation it was possible to rotate the Glc and HepII (for CA7, Model B) to positions where they did not clash internally or with protein. Model C containing the rotated *H. influenzae* Eagan CA7 PS was placed in a similar orientation to HepI as shown in **Fig. 3.15**, while ensuring that all the sugars of model C were within acceptable distances from each other and from the protein. While rotating the sugars of model C into acceptable

distances, a patch of density was observed located in the vicinity of Arg349, near where the PEA group off HepII would be located. Model C was fitted in similar orientation to that reported in this chapter for HepI, with the PEA in the density and refined, see **Fig. 3.20A** for the PEA fitting in the initial density and **Fig. 3.20B** for the density seen after refinement.

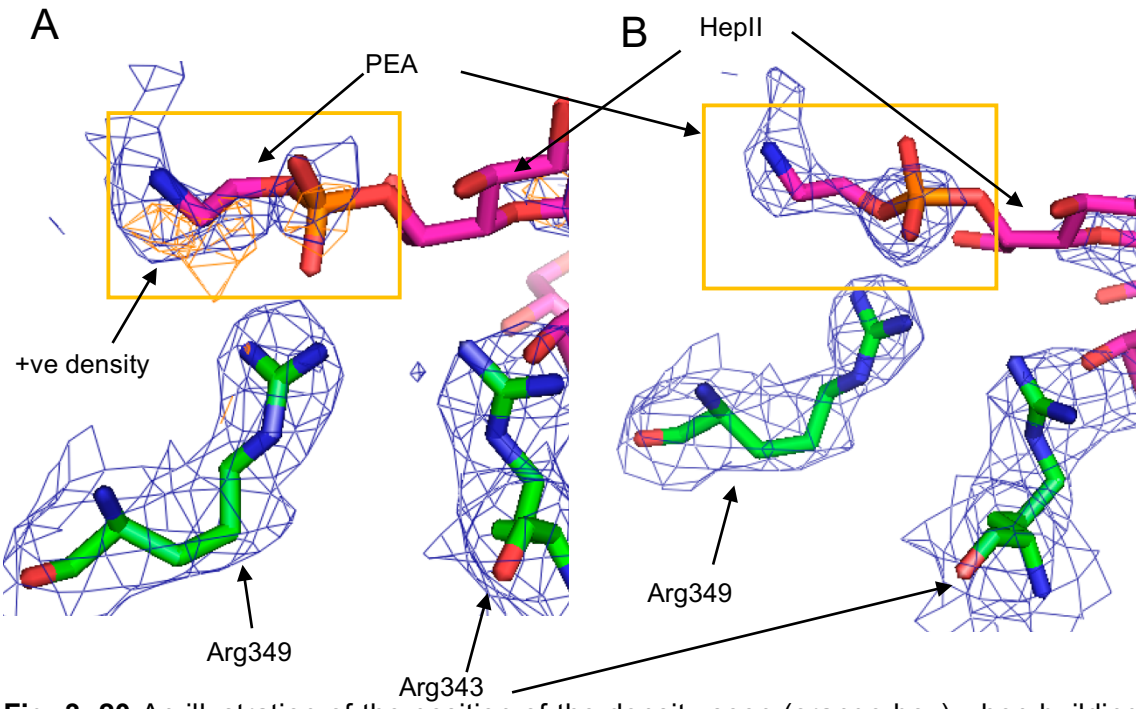


Fig. 3. 20 An illustration of the position of the density seen (orange box) when building model C containing the CA7 PS with the novel orientation of HepI; the PEA group seems to be located near Arg349. For both figures the direct map (2Fobs-Fcalc) is in blue and contoured 1σ. **A-** The density seen located in the vicinity of the PEA group before refinement, with the difference map (Fobs-Fcalc) positive density in orange and contoured at 2.5σ. **B-** The density seen after refinement, suggesting that the PEA could be present in this orientation. Figures A and B generated using *PYMOL.*

The density seen around the putative PEA bound to HepII following refinement suggested a possible orientation of how HepII and PEA could be oriented in model C, with the PEA group within hydrogen bonding distance of Arg349. **Fig. 3.21** shows an illustration of the Eagan CA7 PS model C orientation in comparison with the Eagan 4A PS orientation reported in the literature (**Clark et al., 2016**).

When superimposing model A, B and C together (**Fig. 3.22**), no evident reason can be seen as to why the Eagan CA7 PS HepI model B did not bind as reported for the

Eagan 4A PS model A. In Model A (4A) and model B (CA7) the HepI orientation is pointing outwards, towards the long loop region (LLR) loop 3 (L3) and L4, with the model C (CA7-rotated) facing inwards and above chain B. Therefore, bearing in mind that model C arises from this structural study while Model B represents a manual placement to avoid clashes, in terms of spatial organisation both model B and C could have enough space with no clashes and no steric restrictions.

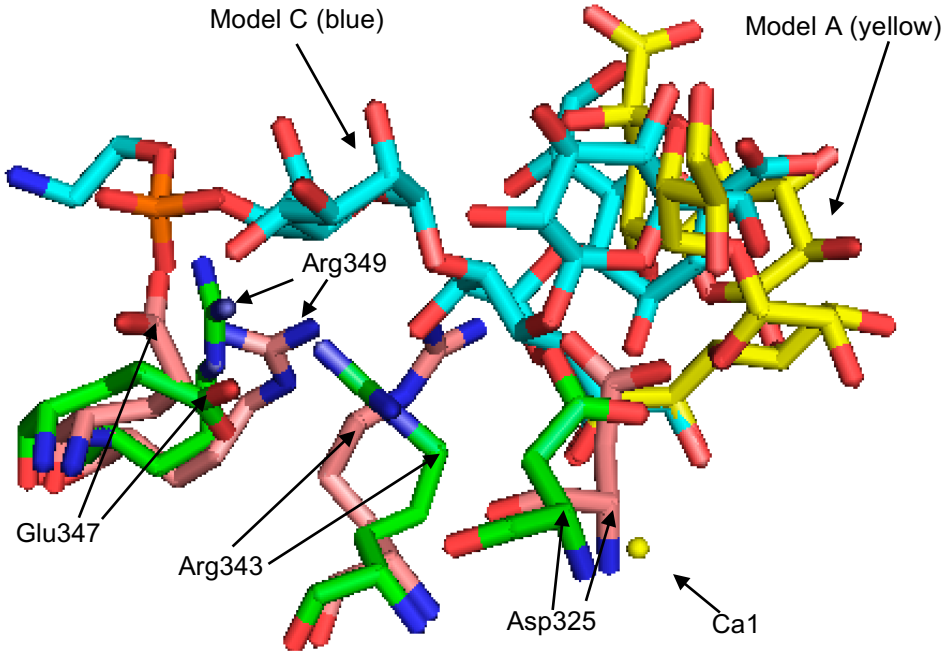


Fig. 3.21 An illustration of how model C (CA7 PS) would be orientated using the novel orientation of HepI rotated by 180° in comparison to Eagan PS model A (PDB: 4E52). Model C is displayed in blue (carbons), red (oxygens), orange (nitrogen) and dark blue (phosphate) with corresponding model residues Arg349, Glu347 and flanking residues Arg343 and Asp325 displayed in green (carbons), red (oxygens) and blue (nitrogen). Model A is displayed in yellow (carbons) and red (oxygens) with the corresponding model residues Asp325, Arg343, Glu347 and Arg349 displayed as reported in the 4A PS structure (**Clark et al., 2016**) in brown (carbons), red (oxygens) and blue (nitrogens). The position of Asp325, Arg343, Glu347 and Arg349 in the 4A structure (model A) differ significantly from the position of Asp325, Arg343, Glu347 and Arg349 seen for the CA7 best model (model C). No Ca1 coordinating residues displayed. Figure generated using *PYMOL*.

It is thus clear that the adoption of the novel conformation, which it seems reasonable to assume reflects the low affinity binding mode in solution, is due to additional constraints on the CA7 internal conformation, the overall interaction of CA7 with rfhSP-D in the region of the binding pocket, or both, preventing the adoption of the high affinity 4A

binding mode. This is consistent with binding of the CA7 by rfhSP-D both before and after the formation of the crystal lattice.

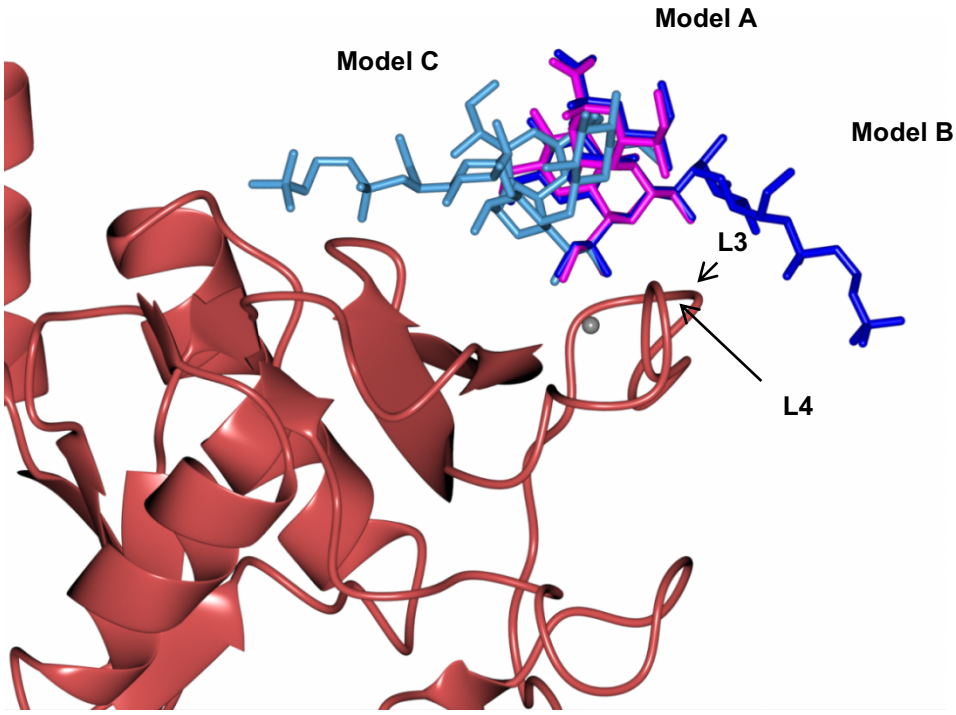


Fig. 3.22 An illustration of the superposition of models A, B and C. Model A shows the reported Eagan 4A PS (PDB: 4E52) displayed in magenta (**Clark et al., 2016**). Model B shows the Eagan CA7 PS with HepI in similar orientation to the HepI in 4A PS and with the PEA group above L3 and L4, displayed in dark blue. Model C shows the novel orientation reported here for the HepI with the PEA group attached to HepII in the vicinity of Arg349. Chain B is displayed using ribbons and in red. Image generated using *MOLECULAR GRAPHICS*.

When fitting model C, the PEA group attached to the HepII sits in the vicinity of the residue Arg349 offering the potential of extra protein-ligand interactions, not possible in model B where the PEA would face outwards towards the middle of L3 and L4. However, in order for HepI to assume this novel orientation reported here, and for the PEA group of HepII to be in vicinity of Arg349, Arg343 and Glu347 need to assume the novel and previously unreported orientation by forming a salt bridge. The positive indication of the presence of PEA attached to HepII suggests that the hydrolysis procedure has no effect on this particular phosphate substitution.

Therefore, when answering the question as to why the novel orientation of HepI is adopted, on the basis that binding occurs before crystal lattice formation by those subunits with *H. influenzae* Eagan CA7 bound in chain B, it is proposed to be the consequence of an inability to adopt the high affinity 4A mode of binding accompanied by the loss of the preferred ligand to binding pocket fit, and the associated interactions with the binding site flanking residues, compensated for in part by extra protein-ligand interactions additional to those possible in the standard HepI orientation.

As the Eagan CA7 is seen only in subunit B, it is not clear whether this binding occurs before or after formation of the crystal lattice. The alternative to binding before crystal lattice formation is that the protein forms a lattice from ligand-free rfhSP-D molecules with subsequent binding of the Eagan CA7. When superimposing model A and model B and using the crystal constraints and symmetry related molecules, it is clear that the Eagan CA7 PS in the same orientation as the Eagan 4A (model B) could not fit in the space between the chain B binding pocket and symmetry related rfhSP-D molecules. The CA7 PS entering the binding site of ligand-free $P42_12$ crystals could not assume the model B (similar to 4A) orientation due to the HepII-PEA of the CA7 PS clashing with Ala278-288Asn in chain A of a symmetry related rfhSP-D molecule (**Fig 3.23A**), while PEA is located near to Thr255-257 of chain B of a symmetry related rfhSP-D molecule (**Fig. 3.23B**). Therefore, it appears that the Eagan CA7 PS HepI could not be introduced into this crystal form ($P42_12$) with the same orientation as the Eagan 4A PS. Rotating the HepII in order to not clash with symmetry related molecules, causes the HepII to clash with the Kdo in one direction and with HepI O2' when rotated towards the other side.

The weight of the evidence suggests that *H. influenzae* Eagan CA7 PS is unable to adopt the standard high affinity preferred binding orientation shown by Eagan 4A, with crystals forming from rfhSP-D -CA7 complexes with a single subunit occupied by CA7 in the novel orientation. These discussions are revisited in Chapter 5 where the results of this study are considered alongside those for the *S. enterica minnesota* LPS/PS in

Chapter 4 as well as the previous work on Eagan 4A (**Clark et al., 2016**) and heptose itself (**Wang et al., 2008**).

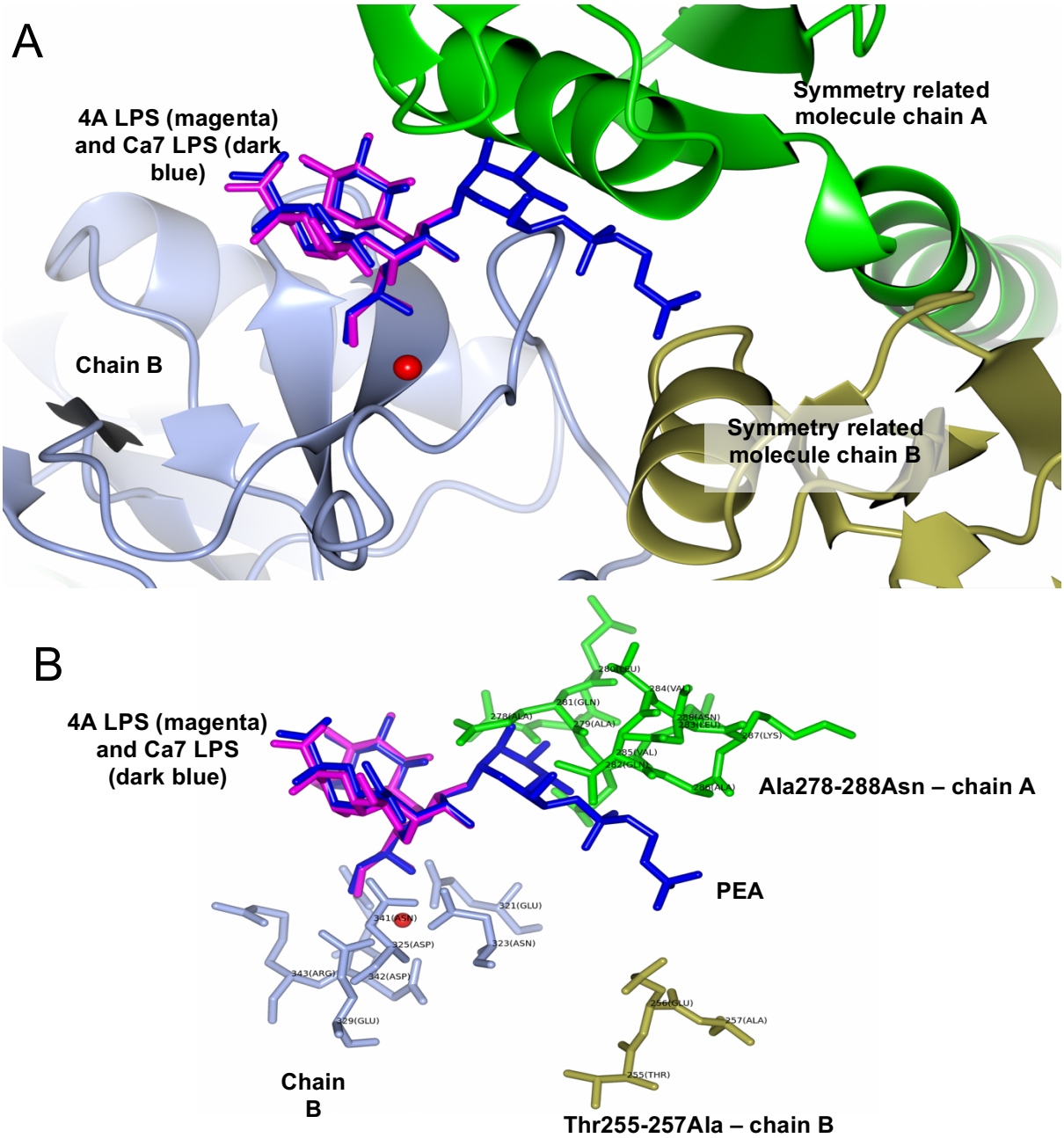


Fig. 3. 23 Superimposed *H. influenzae* Eagan 4A PS model A (PDB: 4E52) (**Clark et al., 2016**) structure as reported in the literature and represented in magenta, with the *H. influenzae* Eagan CA7 PS model B using a similar orientation to the 4A LPS and represented in dark blue. In light blue chain B is shown containing the Ca1 coordinating residues and the flanking residues Arg343 and Asp325. Ca1 is modelled in red. In green, chain A of a symmetry related molecule and in gold chain B of a symmetry related molecule. **A.** using ribbons, the display of how the CA7 PS (dark blue) could not bind in the same orientation as the 4A PS (magenta) due to clashes with the symmetry related molecule chain A (green). **B.** using cylinders and similar display to **A**, HepII-PEA (dark blue) of the CA7 PS could not bind through the same orientation as 4A LPS, as it clashes with the symmetry related molecule of chain A sequence Ala278-288Asn, with the PEA group in proximity to the symmetry related molecule chain B Thr255-Ala257. **A** and **B** displayed using *MOLECULAR GRAPHICS*.

3.3.6.2 Question 2 – How does the novel orientation of HepI shown for Eagan CA7 PS relate to the decreasing affinity of SP-D and rfhSP-D for Eagan wild type, (3 Heps), CA7 (2 Heps) and 4A (1 Hep)?

The current knowledge of the bound orientation seen for the HepI of Eagan 4A PS (**Clark et al., 2016**) and the novel orientation and interactions reported here for HepI of the Eagan CA7 PS suggests an explanation of the reduction in binding reported in the literature as the number of heptoses increases in the Eagan mutants; 4A, CA7 and Wild-Type.

The novel orientation reported in this chapter for CA7 (model C) HepI includes additional interactions between the HepII-PEA group and the Arg349 residue. Wild-type Eagan PS contains the conserved CA7 PS structure, offering the possibility for the HepII-PEA to be located in a similar place to that for CA7. This suggests that the wild-type could also assume the novel orientation of HepI shown by the Ca7 PS.

Using HepI as the binding preference for all mutants and superimposing the Eagan 4A PS as reported (model A), and the Eagan CA7 (model C) and wild-type PS (model D) rotated with respect to the 4A PS, the wild-type Eagan can be modelled without any clashes between carbohydrate and protein and with favourable steric interactions (**Fig. 3.24**). It is important to mention that CA7 and wild-type models can only occur if Arg343 and Glu347 assume the novel orientations as reported here for the CA7 structure. For practical reasons the complete wild-type PS could not be modelled, especially the extensions coming off the Glc bound to HepII; see **Fig. 3.1** for a representation of the wild-type Eagan LPS structure.

HepI was considered to be the preferred binding site where rfhSP-D could recognise *H. influenzae* PS mutants, in part due to the 4A PS (**Clark et al., 2016**) CA7 PS (reported here) complexes however, other recognition sites are possible; 4A PS offers two recognition sites for SP-D, Glc (O3’ and O4’) and HepI (O6’ and O7’); CA7 PS offers three

recognition sites for SP-D, Glc (O3' and O4'), HepI (O6' and O7') and HepII (O3' and O4'); Wild-type PS offers four recognition sites for SP-D, Glc (O3' and O4'), HepI (O6' and O7'), HepII (O3' and O4') and HepIII (O6' and O7') (**Shrive et al., 2003**; **Wang et al., 2008**; **Clark et al., 2016**).

The length of the LPS structures has been shown to be inversely correlated with binding affinities, meaning that longer LPS ligands bind more weakly than shorter LPS ligands to SP-D and rfhSP-D. Eagan 4A PS (1 Hep) has been shown to bind much more strongly than CA7 PS (2 Heps), which in turn shows a stronger binding than Wild-type (3 Heps) (**Clark et al., 2016**). Wild-type Eagan LPS contains a larger chain, 3 Heps and four available recognition sites, but actually shows weaker binding than CA7 and 4A, suggesting that chain branching can hinder recognition by SP-D.

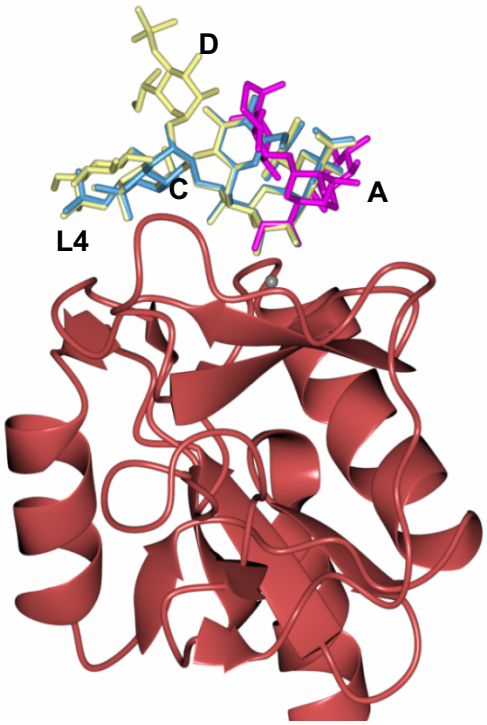


Fig. 3.24 Superposition of the reported Eagan 4A PS (PDB: 4E52, A and displayed in magenta), the Eagan CA7 PS (C and displayed in light blue) reported here and the wild-type Eagan (D and displayed in yellow) positioned in a similar manner to the CA7 PS. Chain B of rfhSP-D, including L4, is displayed in red using ribbons. Image produced using *MOLECULAR GRAPHICS* (**Clark et al., 2016**).

The three-dimensional structure and conformations of oligosaccharides such as *H. influenzae* 4A, CA7 and Wild-type Eagan LPS are dependent on the orientation of the glycosidic torsion angles and steric interactions with water or solvent molecules and protein, especially in aperiodic chains such as LPS (**Kirschner & Woods, 2001**). *H. influenzae* Eagan 4A HepI contains only two linkages to other sugars; HepI-Kdo α 1,4 and Glc-HepI β 1,4. The CA7 and Wild-type HepI contain three linkages; HepI-Kdo α 1,5, Glc-HepI a β 1,4 and HepII-HepI α 1,3. The extra linkage seen in the CA7 and Wild-type has been suggested to introduce steric constraints which might affect the binding by hSP-D to HepI (**Clark et al., 2016**).

Fig. 3.24 reveals a particular and interesting feature. The LLR region, specifically L4 of the LLR (see **Fig 1.4 A)** seems to sit very close to the Wild-type PS (model C) Glc bound to the HepII and where Hex3, Hex4 and Hex5 are present, suggesting a possible space restriction for the wild-type Eagan (**Fig. 3.25 C**)**.** When taking a closer look into the organisation of the LLR L4, it seems to sit very close to the binding pocket, suggesting that it could play a role in the binding affinity of longer ligands. Superimposing the present CA7 model and the model reported for the 4A PS (**Clark et al., 2016**) a nice fit between the models can be seen. However, when examining the L4 for both models, the L4 in the 4A PS bound structure seems to be much closer to the Ca1 than that in the CA7 structure, suggesting that L4 could have some flexibility, see **Fig. 3.25 A** and **B** for an illustration. The position of the L4 in the CA7 structure suggests movement away from the binding pocket, perhaps to accommodate the longer CA7 PS ligand, which contains an extra HepII potentially phosphorylated at position 6 by PEA, in comparison to the 4A PS. This could also suggest that L4 could prevent longer ligands binding in the preferred orientation. L4, containing the Asn323-Glu329 residues, has been reported to show conformational changes upon the loss of Ca2 (**Ng et al., 1998**), as in the present CA7-bound structure reported here. At present, it is not possible to determine if the extra flexibility seen is the consequence of the loss of Ca2 or the presence of the CA7 ligand.

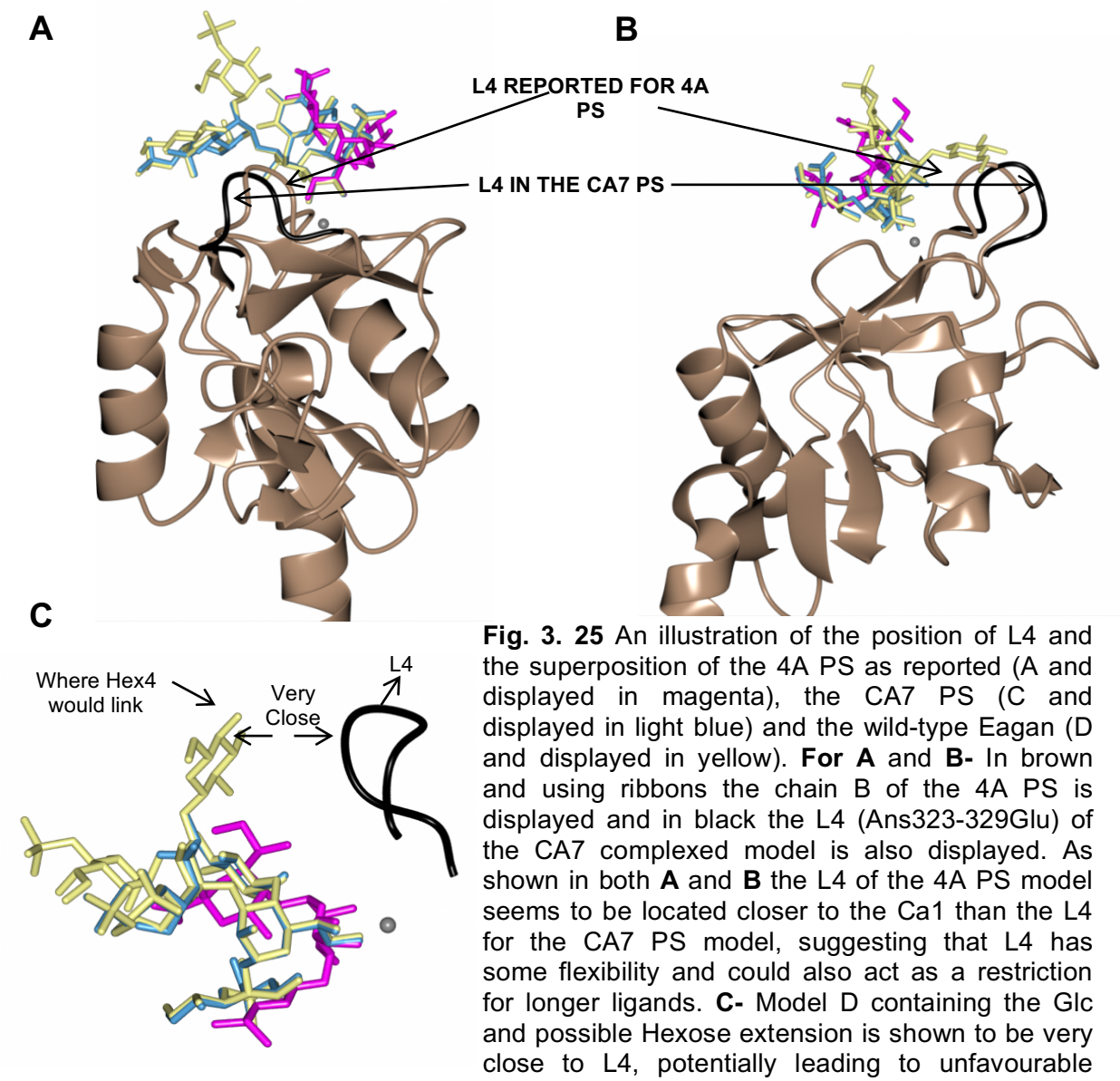


Fig. 3. 25 An illustration of the position of L4 and the superposition of the 4A PS as reported (A and displayed in magenta), the CA7 PS (C and displayed in light blue) and the wild-type Eagan (D and displayed in yellow). **For A** and **B-** In brown and using ribbons the chain B of the 4A PS is displayed and in black the L4 (Ans323-329Glu) of the CA7 complexed model is also displayed. As shown in both **A** and **B** the L4 of the 4A PS model seems to be located closer to the Ca1 than the L4 for the CA7 PS model, suggesting that L4 has some flexibility and could also act as a restriction for longer ligands. **C-** Model D containing the Glc and possible Hexose extension is shown to be very close to L4, potentially leading to unfavourable steric contacts.

Achieving the preferred binding mode via HepI, while assuming favourable steric interactions and carbohydrate-protein interactions would appear to be the main reason why a reduction in binding as branching increases is observed. The Eagan CA7 HepI LPS novel orientation seems to offer a higher binding potential than if the CA7 HepI was orientated in similar orientation the 4A. Wild-type HepI can assume a similar orientation to the CA7 HepI, however the Glc extension coming off HepII is in close proximity with L4 suggesting limited space to accommodate subsequent hexose additions and leading to unfavourable steric interactions and poorer binding. The last consideration to make is that

as part of the novel orientation of the CA7 HepI PS, Arg343 and Glu347 need to also assume the novel orientation reported here, with all other structures reported displaying a different orientation for both Arg343 and Glu347. This suggests that although the novel orientation for Arg343 and Glu347 can occur, it may be a less favourable conformation.

3.4 Summary

The work performed in this chapter undertook the co-crystallisation of rfhSP-D in complex with *H. influenzae* CA7 PS and followed the work performed and reported for the shorter *H. influenzae* 4A PS (**Clark et al., 2016**). Co-crystallisation was achieved, as previous soaks of CA7 PS into ligand-free rfhSP-D crystals failed to reveal any binding. The co-crystallisation of rfhSP-D and CA7 PS was highly successful in producing a myriad of protein crystals, with two data sets collected that provided sufficient detail to be used in the model building and characterisation of the recognition and binding of Eagan CA7 PS by rfhSP-D. While the electron density maps for both data sets displayed somewhat ambiguous density, the identification of novel conformations for the flanking residue Arg343 and for Glu347, and the novel salt bridge interaction between the two residues, provides additional evidence for the orientation of HepI in the binding pocket of rfhSP-D chain B. HepI shows a completely novel mode of interaction, with HepI rotated by 180° about C5-C6 which switches the positions of the coordinating hydroxyls with respect to those in the Eagan 4A PS complex. Similar to the reported structure for the rfhSP-D – Eagan 4A PS complex (**Clark et al., 2016**), HepI seems to be the preferred binding determinant for rfhSP-D, even when Glc and HepII are available to bind via O3' and O4' equatorial groups (**Shrive et al., 2003**; **Wang et al., 2008**). When attempting to explain the reasons behind the novel orientation of HepI, a full model of the CA7 PS was built and positioned with favourable steric interactions and within acceptable distances to the rfhSP-D chain B. The expected position of the PEA group attached to the HepII in the full model of the CA7 PS was putatively located in the electron density vicinity of Arg349, suggesting a possible interaction between PEA and Arg349. Finally, when superimposing all three *H.*

influenzae models and investigating the reasons why increased branching in *H. influenzae* mutants has been reported to lead to poor binding to rfhSP-D and SP-D, the position of hSP-D L4 was seen to be located in close proximity to Ca1 suggesting a potential role in binding longer ligands, but also as a possible constraint in terms of favourable spatial orientation for longer LPS ligands.

Chapter 4 - Structural study of co-crystallised rfhSP-D and *S. enterica* minnesota LPS and PS strains.

4.1 Introduction

The use of LPS in structural studies with rfhSP-D creates a series of difficulties, including the insolubility of the intact LPS and the possibility of structural changes in the polysaccharide as a result of mild acid hydrolysis to remove the lipid A and hence increase solubility. The hydrolysis of LPS can be easily performed and is a fairly efficient method of removing the lipid A, yielding a purified PS component which can be dissolved in a solution containing the rfhSP-D. Even though the hydrolysis method can efficiently separate the lipid A from the PS, it can also impact on structural components of the PS, by causing conformational and structural changes to the Kdo (**Danan et al., 1982**; **Auzanneau et al., 1991**; **Clark et al., 2016**). Of particular importance is the β elimination of phosphate components when present on the Kdo resulting in significant rearrangement of the Kdo (**Danan et al., 1982**; **Auzanneau et al., 1991**; **Clark et al., 2016**). It could be argued that this structural change in the Kdo reduces the physiological validity of the binding interactions determined through structural studies, although the key interaction has been shown to be mediated through Hep rather than Kdo (**Clark et al, 2016**). Solubilising the intact LPS would allow further characterisation of how hSP-D interacts and recognises the various glycoconjugates present on the surface of pathogens but there are few reports in the literature of methods to achieve this.

Therefore, the initial aim of the study of *Salmonella enterica* (serotype minnesota) LPS was to develop a suitable method of dissolving the full LPS for the R7 rough strain. These attempts yielded promising results, suggesting the potential to partially or fully solubilise full LPS to a concentration which can be used in crystallography.

The initial experiments evaluated the potential of Triton X-100 and absolute ethanol in aiding in the partial or full solubilisation of full LPS, and the production of co-crystals of rfhSP-D in complex with *S. enterica* minnesota LPS R7. *E. coli* O111:B4 LPS was initially used due to its abundant availability in our laboratory. The co-crystals produced were used to investigate the binding of rfhSP-D to solubilised *S. enterica* minnesota R7 LPS by means of X-ray diffraction. A third LPS (PS) ligand was also used, with *S. enterica* minnesota R5 PS (RcP^-), a slightly longer strain than R7 LPS, being hydrolysed and used by soaking into the co-crystals of rfhSP-D and R7 LPS. All the crystallisation attempts occurred at Huxley Building, Keele University, Staffordshire. The data collection took place at Diamond Light Source (DLS), Diamond House, Harwell Science and Innovation Campus, Fermi Ave, Didcot, Oxfordshire. rfhSP-D supply, purification and preparation were performed as previously described (see Chapter 2). The rfhSP-D batch used in combinations with full LPS was Batch 6 and the structure/industrial screens were from Molecular Dimensions.

4.1.1 Lipopolysaccharide ligands

The LPS used were *Escherichia coli* O111:B4 (Sigma-Aldrich – L4391), *Salmonella enterica* minnesota R5 (Enzo Lifesciences- ALX-581-017-L002) and R7 (Enzo Lifesciences ALX-581-018-L002).

4.1.1.1 *Escherichia coli* O111:B4 smooth LPS

The *E. coli* O111:B4 smooth LPS strain is an R3 core type (**Fig. 4.1**). This serotype can produce rough strain LPS, such as the *E. coli* J-5 (Rc) rough strain LPS. The O111:B4 LPS is composed of the lipid A linked to a core region rich in heptoses, glucoses, galactoses and the O-antigen, which is normally composed of GlcNac, GalNac and Gal (**Stenutz et al., 2006**). The O111:B4 LPS structural components were first proposed in 1975 to produce two different fractions, where LPS fraction 1 (12324.7 Da) contained 11 repeats of the O-antigen, while LPS fraction 2 (6022.3 Da) contained 3

repeats of the O-antigen (**Morrison et al., 1975**). HepI and HepII in the inner core can be phosphorylated leading to increased heterogeneity. The high heterogeneity seen in smooth LPS creates a problem when investigating specific strains (**Freudenberg et al., 2008**). At present the characterisation of the O111:B4 LPS and different fractions remains incomplete.

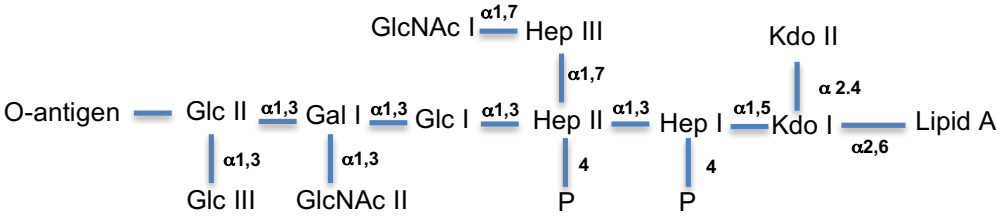


Fig. 4.1 The proposed structure of O111:B4 LPS, with the inner and outer core structures represented. At present due to the high heterogeneity seen in the LPS strains the full structure of the O111:B4 *E. coli* LPS remains unknown. Glc- Glucose; Hep- L-D-Heptose, KDO- 3-deoxy-D-manno-octulosonic acid; Gal- Galactose; GlcNAc - N-acetyl-glucosamine; P- Phosphate; PEA- Phosphate ethanolamine. Image adapted from **Raetz & Whitfield, 2002** and **Morrison et al., 1975**

4.1.1.2 *Salmonella enterica* minnesota R7 (Rd_1P^-) rough mutant LPS

Salmonella enterica minnesota R7 full LPS was the rough strain chosen to be investigated in more detail. This particular LPS was of high interest, as the core includes a double Hep and a triple Kdo, adding to the knowledge base of hSP-D – LPS interactions in addition to the *H. influenzae* Eagan 4A (1 Hep, 1 Kdo) and Eagan Ca7 (2 hep, 1 Kdo) structures reported here and by **Clark et al., (2016)**. The co-crystallisation of the full R7 LPS with rfhSP-D was dependent on the novel solubility method reported here (and chapter 2). Structural studies of rfhSP-D bound to full LPS have never been reported in the literature, suggesting that if the co-crystallisation of rfhSP-D and R7 LPS was successful, it would offer the opportunity to investigate and characterise in detail the binding between rfhSP-D and full LPS for the first time.

Salmonella R7 rough strain LPS was first reported in 1966 and occurs by means of mutations in the incorporation of the hexose GlcI to HepII, followed by a mutation in the phosphorylation of HepI and subsequent inactivation of phosphorylation of HepII. Both mutations produce a rough LPS containing the Lipid A, three KDOs and two Heps, where

the Heps are phosphate free (**Fig. 4.2**) (**Holst & Brade, 1991**; **Jousimies & Makela, 1974**; **Schnaitman & Klena, 1993**; **Raetz & Whitfield, 2002**; **Luderitz et al., 1966**).

It is important to note that during the biosynthesis of the LPS, the order of gene activation and enzyme transfers have been suggested to follow sequentially. This suggests that mutations leading to the non-addition of GlcI and phosphate groups would impact on the addition of, for example, HepIII to HepII or GalI or GalII to GlcI, yielding an incomplete core region (**Parker et al., 1992**; **Yethon et al., 1998**).

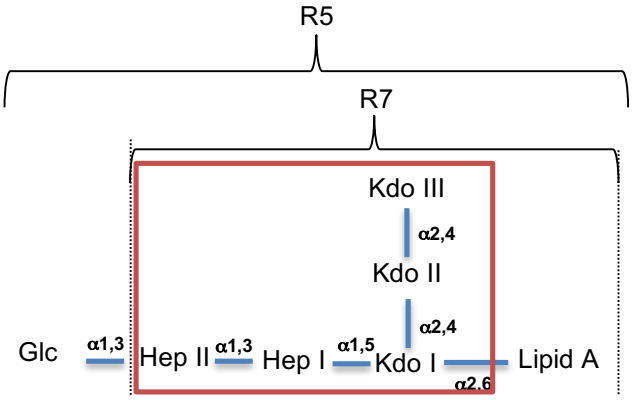


Fig. 4.2 The proposed structure of *S. enterica* (minnesota) R7 LPS containing a HepII-HepI-KdoI-KdoII-KdoIII pentasaccharide bound to the Lipid A and the R5 LPS containing an extra Glc. Both LPSs contain the same KdoIII-KdoII-KdoI-HepI-HepII pentasaccharide (orange box). Hep- L-D-Heptose, Kdo- 3-deoxy-D-manno-octulosonic acid, Glc- glucose. Image produced by using information in **Holst & Brade, 1991**; **Jousimies & Makela, 1974**; **Luderitz et al., 1966**.

4.1.1.3 *Salmonella enterica* minnesota R5 (RcP^-) rough mutant LPS

R5 LPS exhibits a similar structure to R7 LPS (**Fig. 4.2**), containing just an extra glucose (GlcI) bound to HepII. This strain arises from a single mutation in the phosphorylation of HepI, inhibiting and impacting on the phosphorylation of HepII and the addition of GalI and GalII to the GlcI. If the phosphorylation of HepI and HepII occurs, this initiates the extension from GlcI where GalI and GalII are added, producing an Rb2 strain **Muhlradt et al., 1967**; **Schnaitman & Klena, 1993**; **Jousimies & Makela, 1974**; **Raetz & Whitfield, 2002**.

4.2 Ligand preparation

4.2.1 Solubility of *E. coli* O111:B4 LPS in Triton X-100 and Ethanol

Triton X-100 and absolute ethanol were used in three separate experiments, in order to investigate their potential as solubilising agents of full LPS. The solubility of O111:B4 LPS in absolute ethanol and/or Triton X-100 was performed as previously described in chapter 2, where 24 well Linbro plates were used, except that no protein was added to the micro-bridge and instead the solution contained 1:1 rfhSP-D storage buffer (50 mM Tris base, 150 mM NaCl, 10 mM $CaCl_2$ pH 7.4) and O111:B4 LPS at varying concentrations and in combination with various concentrations of Triton X-100 and/or, absolute ethanol. An initial investigation was performed to evaluate the solubility limit of O111:B4 *E. coli* LPS in rfhSP-D storage buffer, with 6 mg/ml (0.65 mM) of O111:B4 LPS in rfhSP-D storage buffer representing the solubility limit. The impact of the varying concentrations of Triton X-100 and absolute ethanol on specific concentrations of O111:B4 were investigated by means of light microscopy visualisation and by comparison to solutions containing just O111:B4 LPS and rfhSP-D storage buffer.

The molecular weight of O111:B4 *E. coli* LPS was calculated to be approximately 9173.5 Da based on the two separate fractions previously reported to be present in the O111:B4 LPS. Since LPS is rather heterogeneous and no other information has been reported regarding the % of each fraction, it was assumed that both fractions existed in equal quantities (**Morrison et al., 1975**).

4.2.1.1 *E. coli* O111:B4 LPS solubilised in Triton X-100

Different concentrations of the *E. coli* O111:B4 LPS solution were prepared; 8 mg/ml (0.87 mM), 10 mg/ml (1.09 mM), 12 mg/ml (1.31 mM), 14 mg/ml (1.53 mM) and 16 mg/ml (1.74 mM). Six solutions of Triton X-100 were added to the LPS solutions (0.5 µl) producing a final concentration of 1 mM, 0.5 mM, 0.4 mM, 0.3 mM, 0.2 mM, and 0.1 mM of Triton X-100 + LPS (final volume 2 µl), with the aim of investigating the impact of Triton

X-100 above and below the CMC (**Table 4.1**). This produced 30 different mixtures (LPS + Triton X-100), which were compared against LPS solutions at 8 mg/ml, 10 mg/ml, 12 mg/ml, 14 mg/ml and 16 mg/ml with no Triton X-100 present.

The various different concentrations of Triton X-100 tested on specific O111:B4 LPS concentrations demonstrated that a concentration of Triton X-100 at 0.2-0.3mM can improve the solubilisation of O111:B4 LPS in storage buffer at 16 mg/ml, (**Fig. 4.3**).

LPS concentrations at 8 mg/ml, 10 mg/ml, 12 mg/ml and 14 mg/ml + Triton X-100 at 0.2 mM or 0.3 mM also demonstrated similar results to the 16 mg/ml mixture. The addition of 0.2-0.3 mM Triton X-100 to the various solutions produced a clear and homogeneous solution. When the concentration of Triton X-100 used was above 0.3 mM, the solution seemed slightly cloudy and at 0.1 mM it appeared precipitated.

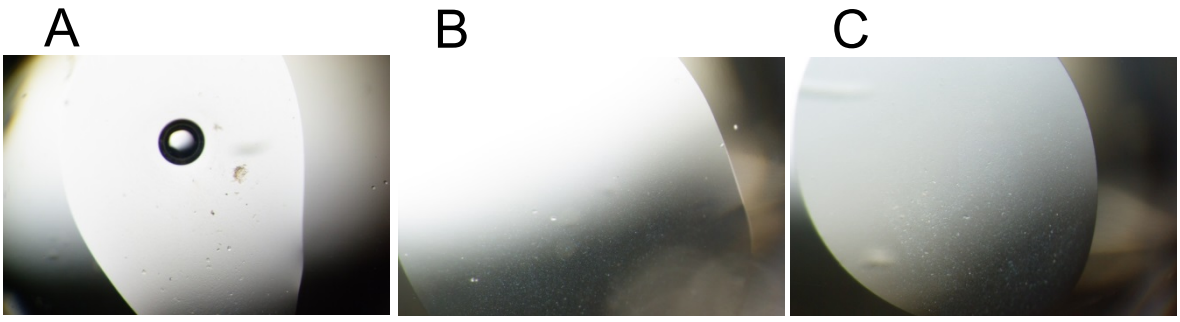


Fig. 4.3 A- 16 mg/ml O111:B4 LPS in rfhSP-D storage buffer, no detergent added. Typically, the LPS does not mix in solution, with eye visible "chunks" of LPS in suspension. **B-** 16 mg/ml O111:B4 LPS in rfhSP-D storage buffer + 0.3mM Triton X-100. No LPS "chunks" were seen and the solution looked homogenous. **C-** 16 mg/ml in O111:B4 LPS rfhSP-D storage buffer + 0.2 mM Triton X-100. Similar to **B,** at 0.2 mM Triton X-100 seemed to solubilise the LPS producing a fairly homogeneous solution.

4.2.1.2 O111:B4 LPS solubilised in absolute ethanol

Different concentrations of the *E. coli* O111:B4 LPS solution were prepared; 15 mg/ml (1.64 mM), 20 mg/ml (2.18 mM), 25 mg/ml (2.73 mM), 30 mg/ml (3.27 mM) and 35 mg/ml (3.82 mM) in combination with 30% (v/v) absolute ethanol (**Table 4.1**) as reported in the literature (**Sharma et al., 2013**). 0.6 µl of ethanol was added to each of the different concentrations of LPS, producing five different mixtures (LPS + 30% (v/v) absolute

ethanol) with 2 µl final volume. This was compared against solutions of LPS at 15 mg/ml, 20 mg/ml, 25 mg/ml, 30 mg/ml and 35 mg/ml with no ethanol added.

E. coli LPS (mg/ml)	Just Triton X-100						Just absolute ethanol	Triton X-100 and absolute ethanol in combination			
	Triton X-100 (mM)						Absolute ethanol (%) (v/v)	Solution Number	Absolute Ethanol (%) (v/v)	Triton X-100 (mM)	*E.coli* LPS (mg/ml)
8	1	0.5	0.4	0.3	0.2	0.1	N/A	1	10	0.25	40
10	1	0.5	0.4	0.3	0.2	0.1	N/A	2	10	0.25	35
12	1	0.5	0.4	0.3	0.2	0.1	N/A	3	10	0.25	30
14	1	0.5	0.4	0.3	0.2	0.1	N/A	4	20	0.25	40
15	N/A	N/A	N/A	N/A	N/A	N/A	30	5	20	0.25	35
16	1	0.5	0.4	0.3	0.2	0.1	N/A	6	20	0.25	30
20	N/A						30	7	30	0.25	40
25	N/A						30	8	30	0.25	35
30	N/A						30	9	30	0.25	30
35	N/A						30				

Tablle 4.1 The illustration of the three experiments performed to investigate the solubility potential of *E. coli* O111:B4 LPS using Triton X-100 alone (shaded in yellow), absolute ethanol (v/v) (shaded in green) and using a combination of Triton X-100 and absolute ethanol together (shaded in blue). In the Triton X-100 experiment 30 different mixtures (LPS + Triton X-100) were investigated, with a concentration of Triton X-100 at 0.2-0.3mM (in red in the table) improving the solubility of the LPS at 8, 10, 12, 14 and 16mg/ml. The second experiment used only absolute ethanol 30% (v/v), with a concentration of LPS up to 20mg/ml (in red in the table) showing signs of improved solubility. At 25, 30 and 35mg/ml absolute ethanol 30% (v/v) seemed inefficient in improving the solubility of the LPS with the three solutions looking saturated. The final experiment used a combination of Triton X-100 at 0.25mM (the CMC limit of Triton X-100) and absolute ethanol at 10, 20 and 30% (v/v). In this experiment solution 8 and 9 (LPS at 35mg/ml and 30mg/ml respectively) (displayed in red in the table) were both successful in partially or fully solubilising the *E. coli* LPS, suggesting that up to 35mg/ml a combination of Triton X-100 and absolute ethanol (v/v) could aid in the solubilisation of the LPS.

At 30% (v/v) absolute ethanol can aid in partially or fully solubilising O111:B4 LPS up to 20 mg/ml. Both15 mg/ml and 20 mg/ml of O111:B4 LPS produced a visibly clearer solution, but at 25 mg/ml or above the solutions seemed saturated with visible LPS “chunks” in suspension (**Fig. 4.4**).

4.2.1.3 O111:B4 LPS solubilised in triton X-100 and absolute ethanol

Three ethanol concentrations (10%, 20% and 30% (v/v)) and three LPS concentrations were tested (30 mg/ml (3.27 mM), 35 mg/ml (3.82 mM) and 40 mg/ml (4.36 mM). In the earlier 30% ethanol trials the LPS seemed saturated at 25 mg/ml, and the effect on this saturated solution of adding Triton X-100 was investigated. As Triton X-100 alone at 0.2 mM and 0.3 mM concentrations showed a positive effect on the solubilisation

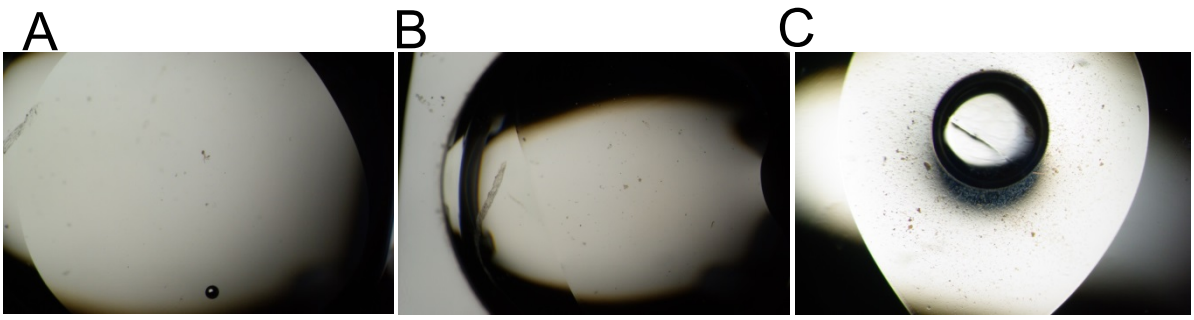


Fig. 4.4 A- 30% (v/v) absolute ethanol seemed to dissolved O111:B4 LPS up to 20 mg/ml, with the solution looking homogenous **B-** At 25 mg/ml the solution started to look saturated with bits of LPS in suspension. **C-** At 30 mg/ml the solution looked saturated with many bits of LPSs in solution.

of the LPS, a solution midway between, at ≈0.25 mM (the CMC limit for triton X-100) was used. In total 9 mixtures were made up as shown in **Table 4.1**.

Of all the mixtures tested, solutions 8 and 9 seem to hold the most potential for the partial or full solubilisation of the LPS (**Fig. 4.5**), displaying a homogenous solution with no signs of large LPS "chunks". Solution 8 contained a higher concentration of LPS, in comparison to solution 9 and this was visible as it showed some precipitation. All the solutions containing 10 or 20% ethanol were less effective at solubilising the LPS. At 40 mg/ml of LPS all the solutions seemed full of precipitation and debris.

Solutions 8 and 9 both present themselves as suitable mixtures for the solubilisation of O111:B4 LPS, suggesting that up to 35 mg/ml of O111:B4 LPS can be solubilised by using a mixture of 0.25 mM triton X-100 and 30% ethanol. Solution numbers 8 and 9 were performed a further three times to ensure the feasibility and reliability of the results and in all three attempts the same results were produced, a clear and

homogeneous solution with no big “chunks” of LPS in suspension.

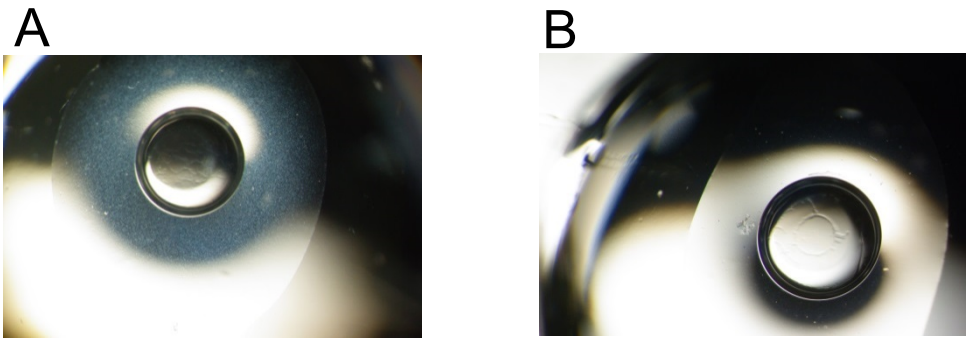


Fig. 4.5 Using a combination of 0.25 mM Triton X-100 and 30% (v/v) absolute ethanol seemed to have dissolved the O111:B4 LPS up to 35 mg/ml, with both solutions 8 (image **B**) and 9 (image **A**) being fairly homogenous. **A-** Solution 9 - 30% (v/v) absolute ethanol + 0.25 mM of Triton X-100 and 30 mg/ml LPS. **B-** Solution 8 - 30% (v/v) absolute ethanol + 0.25 mM of Triton X-100 and 35 mg/ml LPS.

4.2.2 Solubility limit of *S. enterica* minnesota R7 LPS in Ethanol and Triton X-100

The conclusions from the O111:B4 LPS experiments were used to investigate the solubility of R7 *S. enterica* minnesota LPS in combination with both Triton X-100 and absolute ethanol. The molecular weight of R7 *S. enterica* minnesota LPS was calculated by addition of the individual weights of the glycoconjugates and lipid A, yielding a molecular weight of approximately 2880.36 Da, with the lipid A accounting for 62% of the total molecular weight of the LPS (**Holst & Brade, 1991**; **Jousimies & Makela, 1974**; **Schnaitman & Klena, 1993**; **Raetz et al., 2002**; **Que et al., 2000**).

R7 *S. enterica* minnesota LPS solubility was investigated using the combination of 0.25 mM Triton X-100 and 30% (v/v) absolute ethanol. The estimated maximum R7 *S. enterica* minnesota LPS concentration was defined by calculating the concentration (mM) corresponding to the maximum O111:B4 LPS solubility observed in 0.25 mM and 30% (v/v) absolute ethanol (35 mg/ml - 3.67 mM). This equates to 10.57 mg/ml = 3.67 mM for the R7 *S. enterica* minnesota LPS.

Errors in this estimate are likely to arise when considering that O111:B4 *E. coli* LPS is smooth (longer), while R7 *S. enterica* minnesota LPS is a rough strain (shorter)

with a different ratio of hydrophilic to hydrophobic components (**Park et al., 2009**; **Kabanov & Prokhorenko, 2011**). A solubility trial was performed using 0.25 mM Triton X-100 and 30% absolute ethanol and four different concentrations of R7 *S. enterica* minnesota LPS around the estimated maximum achievable; 7.69 mg/ml (2.67 mM), 10.57 mg/ml (3.67 mM), 13.48 mg/ml (4.67 mM) and 16.33 mg/ml (5.67 mM). At 7.69 mg/ml (2.67 mM), 10.57 mg/ml (3.67 mM) and 13.48 mg/ml (4.67 mM) the R7 *S. enterica* minnesota LPS produced a homogenous and clear solution, while at 16.33 mg/ml (5.67 mM) the solution appeared saturated and with LPS in suspension. This experiment suggested that a maximum concentration of approximately 13.48 mg/ml (4.67 mM) of the R7 *S. enterica* minnesota LPS could be solubilised in rfhSP-D storage buffer, by using a mixture of 0.25 mM Triton X-100 and 30% (v/v) absolute ethanol.

4.2.3 Hydrolysis of *S. enterica* minnesota R5

The hydrolysis of *S. enterica* minnesota R5 LPS was performed according to the method previously described in Chapter 2. The molecular weight of *S. enterica* minnesota R5 LPS was calculated as ≈ 3030.36 Da by the addition of all glycoconjunagtes and Lipid A. The lipid A portion accounted for ≈ 1800 Da (**Que et al., 2000**), making up 58.8% of the LPS. 2 mg of R5 LPS were initially hydrolysed, yielding ≈0.7 mg of PS representing 84% hydrolysis efficiency. The 0.7 mg of R5 PS were dissolved in ≈35 µl of deionised water, producing a concentration of ≈50 mM for storage and for use in crystal soaking.

4.3 Co-crystallisation of rfhSP-D in combination with solubilised R7 LPS

The co-crystallisation of rfhSP-D in combination with R7 LPS was performed according to the methods described in Chapter 2. The final drop was composed of a 1:1 ratio of specific precipitant buffers and the protein/LPS solution containing rfhSP-D at 7.5 mg/ml (0.42 mM), R7 *S. enterica* minnesota LPS at 4.67 mM and 30% (v/v) absolute ethanol + 0.25 mM of Triton X-100. The Ca^{2+} concentration was ≈ 5 mM in the drop.

Structure screens were initially used to establish growth conditions. A total of 2 µl was placed in each micro-bridge with R7 LPS at 11.1-fold excess over the rfhSP-D monomer binding site. Six 24-well Linbro plates containing the 100 structure screen buffer precipitants of SS1 and SS2 were set down on 6/10/2014, with the following tray codes; RS43, RS44, RS45, RS46, RS47 and RS48. Positive crystal growth was seen in RS43 C3 (SS1 C.15), RS44 A2 (SS1 C.20), RS44 B4 (SS1 C.28), RS45 A2 (SS1 C.38), RS46 B6 (SS2 C.12) and RS47 A1 (SS2 C.19) (**Table 4.2**).

Tray + Well ID	Comments and conditions
RS43 C3	This well was set down on 6/10/2014. The condition used in this well comes from SS1 condition 15. This well contained 0.2M Mg acetate, 0.1M MES pH 6.5 and 20% PEG 8K.
RS44 A2	This well was set down on 6/10/2014. The condition used in this well comes from SS1 condition 20. This well contained 0.2M Calcium acetate, 0.1M MES pH 6.5 and 18% PEG 8K.
RS44 B4	This well was set down on 6/10/2014. The condition used in this well comes from SS1 condition 28. This well contained 0.1M Na Hepes pH 7.5, 0.8M Na/K dihydrogen Phosphate.
RS45 A2	This well was set down on 6/10/2014. The condition used in this well comes from SS1 condition 38. This well contained 0.1M Tris pH 8.5 and 8% PEG 8K.
RS46 B6	This well was set down on 6/10/2014. The condition used in this well comes from SS2 condition 12. This well contained 0.1M Na Hepes pH 7.5 and 10% PEG 6K and 5% MPD.
RS47 A1	This well was set down on 6/10/2014. The condition used in this well comes from SS2 condition 19. This well contained 0.1M Na Hepes pH 7.5 and 10% PEG 6K and 8% Ethylene Glycol.

Table 4.2 The six wells that produced positive crystal growth from the SS1 and SS2 buffer precipitants.

Eight follow ups were produced in a similar manner to that previously described in Chapter 3 (Eagan CA7), but failed to produce any crystal growth. As a as result the eight follow ups will not be discussed further. In total 14 24-well Linbro plates were produced and over 130 precipitant buffers and/or variations were investigated.

4.3.1 RS44 A2 (SS1-20), RS45 A2 (SS1-38), RS46 B6 (SS2-12), RS47 A1 (SS2-19)

These wells produced a variety of morphologies (**Fig. 4.6**) with a total of 15 crystals collected, cryo-protected and frozen for testing at Diamond Light Source. All four

wells produced similar shaped crystals, with rectangular, cubic, thin square plates and other irregular shapes seen.

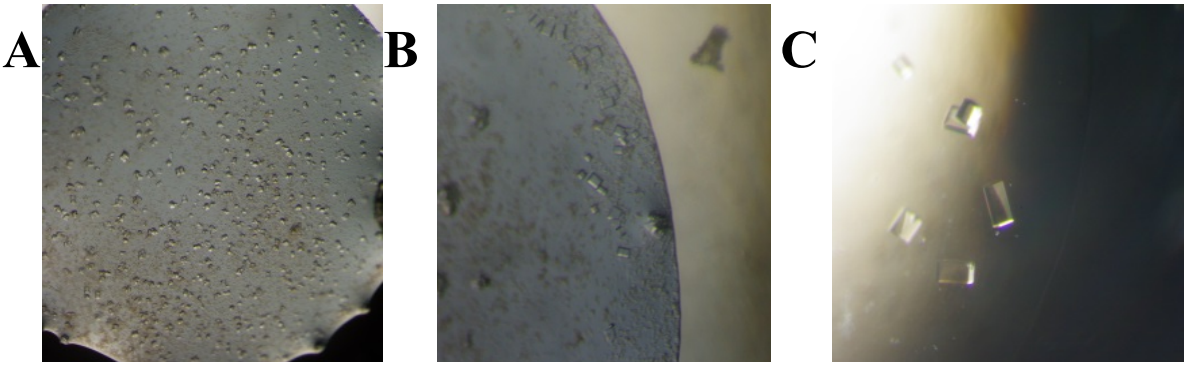


Fig. 4.6 Typical crystal morphologies. **A-** RS44 A2 well **B-** RS46 B6 well **C-** RS47 A1 well.

4.3.2 RS43 C3 (SS1- 15)

The RS43 C3 well produced various crystals with varied morphologies, such as squares, rectangles and odd-shaped crystals (**Fig. 4.7**).

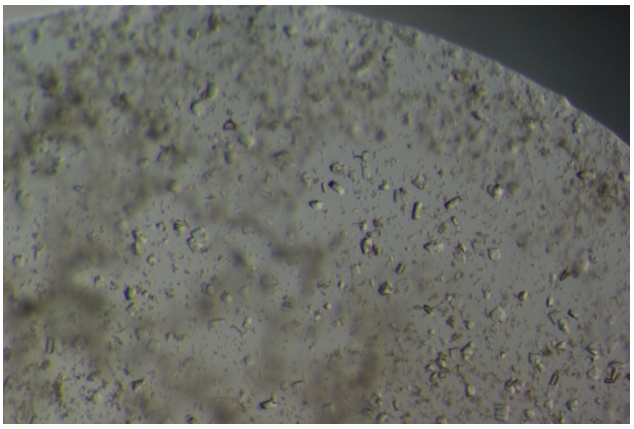


Fig. 4.7 The typical crystal morphology produced in the RS43 C3 well. As can be seen in the image a variety of morphologies were seen, mostly irregular shaped crystals. The brown precipitation in the solution is the *S. enterica* minnesota R7 LPS in solution.

4.3.3 RS44 B4 (SS1- 28)

The RS44 B4 well produced various crystals, mostly thin squares and rectangles, with a few multiple/layered crystals (**Fig. 4.8**).

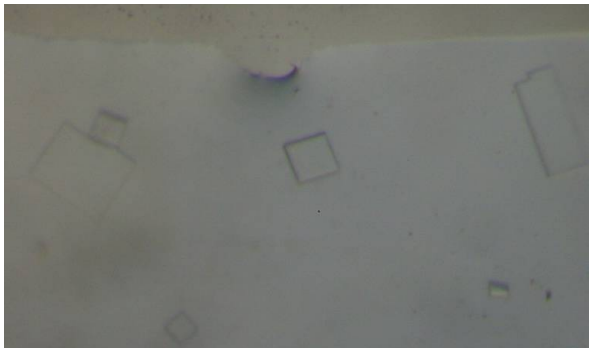


Fig. 4.8 The typical morphology of crystals produced in the RS44 B4 well is shown. Mostly thin squares were produced in this well.

4.4 RS44 A2 (SS1-20), RS45 A2 (SS1-38), RS46 B6 (SS2-12), RS47 A1 (SS2-19)

4.4.1 Cryo-protection

Cryo-protection was performed according to the method described in Chapter 2. A total of 33 co-crystals of rfhSP-D and R7 LPS were cryo-protected, flash-frozen and with X-ray data being collected in each case. Glycerol was chosen as the cryo-protectant of choice. Strategy 2 was initially used due to the experimental problems in mixing the R7 LPS + 30% ethanol + 0.25 mM of Triton X-100 with Glycerol. Strategy 2 was based on the addition of 0.5µl (twice) of 50% glycerol, with a 2-minute interval between additions. Strategy 2 produced a final concentration of Glycerol ≈ 16.6%, R7 LPS ≈ 3.11 mM and Ca^{2+} ≈ 3.33 mM in the micro-bridge. While using strategy 2, no ligand was present in the cryo solution. This strategy was used in the cryo-protection of crystals from the RS44 A2, RS45 A2, RS46 B6 and RS47 A11 wells (**Table 4.3**). The cryo-protection was successful in cryo-protecting the crystals, as no ice rings were detected during data collection and the crystals were able to sustain the addition of 50% glycerol.

Well Solutions +	Glycerol %	Volume added (µl)	Time in solution (minutes)	Comments
RS44 A2 - 0.2M Calcium acetate, 0.1M MES pH 6.5 and 18% PEG 8K. **RS45 A2** - 0.1M Tris pH 8.5 and 8% PEG 8K. **RS46 B6** - 0.1M Na Hepes pH 7.5 and 10% PEG 6K and 5% MPD. **RS47 A1** - 0.1M Na Hepes pH 7.5 and 10% PEG 6K and 8% Ethylene Glycol.	50	+0.5	≈2	No ligand present in all four cryo solutions.
	50	+0.5	≈2	

Table 4.3 The method of the cryo-protection performed for each of the RS44 A2, RS45 A2, RS46 B6 and RS47 A1 wells. No ligand was soaked in together with the cryo.

4.4.2 Data collection

Due to the high number of co-crystals tested only the crystals that yielded significant results will be discussed in detail. Successful data collection from crystals with

a resolution up to 4.0Å or above were further investigated by means of data analysis using the CCP4 software (see 2.5.5), with the aim of investigating the presence of *S. enterica* minnesota R7 LPS in the co-crystallised rfhSP-D crystals. Four crystals showed a medium to low resolution and were used in data analysis; RS44 A22 (crystal 2 from RS44 A2), RS45 A21, RS46 B63 and RS47 A1, shown in **Table 4.4**.

Crystal Code	Date Tested	Data Collection	Beamline	Notes
RS45 A21	16/1/15	DC- 2.01 Å	I04-1	These represented the best crystals in terms of resolution.
RS44 A22	16/1/15	DC- 3.74 Å	I04-1	
RS46 B63	30/1/15	DC- 2.02 Å	I04-1	
RS47 A11	30/1/15	DC- 2.52 Å	I04-1	

Table 4.4 The four crystals with best resolution are shown, including the date of testing, the resolution of the data collection and the DLS beamline where data was collected.

The RS44 A22, RS45 A21, RS46 B63 and RS47 A11 data sets were analysed by using the fast_dp for the subsequent production of each electron density map using the CCP4 suite as described in chapter 2. The four crystals were all in the ligand-free $P2_1$ space group, with similar unit cell dimensions as reported in the literature (**Shrive et al., 2003**; **Shrive et al., 2009**). For each data set the respective electron density map was produced in order to evaluate ligand presence. The production of each map revealed that the R7 LPS co-crystals with the $P2_1$ space group were ligand-free, with no sign of ligand occupancy in any of the subunits. The electron density maps showed a typical unliganded rfhSP-D trimer structure, where Ca1, Ca2 and Ca3 were present for subunit A, B and C and with no density for Ca4 (**Shrive et al., 2003**; **Shrive et al., 2009**). The electron density maps of the four data sets are not displayed in this thesis.

4.5 RS43 C3 (SS1–C5)

The attempts to co-crystallise rfhSP-D in combination with R7 *S. enterica* minnesota LPS Triton X-100 and ethanol described above produced ligand-free $P2_1$ crystals, where no binding between the rfhSP-D and R7 LPS occurred (**Shrive et al., 2003**; **Shrive et al., 2009**). This offers the opportunity to use the remaining ligand-free crystals to soak in other ligands. Therefore, once it had been shown that all the R7 LPS co-crystals tested were in fact ligand-free, it was decided to use the remaining wells

(RS43 C3 and RS44 B4) to investigate the binding of the *S. enterica* minnesota R5 PS. *S. enterica* minnesota R5 PS is very similar to R7, containing just an extra glucose, making it an ideal candidate for ligand binding studies.

The crystals in well RS43 C3 were cryo-protected using strategy 2, where 50% glycerol in combination with R5 PS at 10mM in the well precipitant buffer were added. The cryo-protection used for the RS43 C3 well is shown in **Table 4.5**. with a total of 5 cryo-protected crystals collected and frozen for testing at Diamond Light Source on beamline I04-1 on 13/4/15. The cryo-protection was successful as no ice rings were detected during data collection and the crystals sustained the addition of 50% glycerol.

Well Conditions +	Glycerol %	Volume added (µl)	Time in solution (minutes)	Comments
0.2M Mg acetate, 0.1M MES pH 6.5, 20% PEG 8K and R5 PS at 10mM	50	+0.5	≈2	The soaking of PS into an LPS co-crystallised rfhSP-D crystals was due to previous crystallisation attempts showing that $P2_1$ crystals were being produced, with no indication of LPS in the co-crystals.
	50	+0.5	≈2	

Table 4.5 The cryo-protection performed in the RS43 C3 well. Strategy 2 was used as the cryo-protection method and included R5 PS at 10mM.

The RS43 C35 data set (co-crystallised with solubilised R7 LPS, followed by addition of the R5 PS via cryo-protection) was analysed by using fast_dp**.** The crystal was, as had been observed for all previous co-crystals detailed above, in the typical ligand-free $P2_1$ space group, with similar unit cell dimensions as reported in the literature (**Shrive et al., 2003**; **Shrive et al., 2009**). The electron density map produced in order to evaluate ligand presence revealed no sign of ligand occupancy in any of the subunits. The electron density maps showed a typical unliganded rfhSP-D trimer structure, where Ca1, Ca2 and Ca3 were present for subunit A, B and C and with no density for Ca4 (**Shrive et al., 2003**; **Shrive et al., 2009**). The electron density map is not displayed in this thesis.

4.6 RS44 B4 (SS1-C28)

The RS44 B4 well represented one of the last wells with protein crystals produced by the unsuccessful attempts at co-crystallisation with the R7 LPS and Triton X-100 + absolute ethanol. As the previous attempt (RS43 C3) soaking the R5 PS at 10mM had been unsuccessful, a higher concentration of R5 PS (21mM) equating to ≈50x molecular excess of ligand to rfhSP-D monomer was used, with the aim of introducing the R5 PS in to the ligand binding site.

4.6.1 Cryo-protection

The cryo-protection of this well used strategy 1, with R5 PS at ≈21 mM present in the glycerol cryo solution which was added at increasing glycerol concentration to the well, with a final exchange performed (**Table 4.6**). The final Ca^{2+} concentration was ≈ 1.2 mM in the micro-bridge following exchange, with the R7 LPS from the original co-crystallisation also present in solution.

Well Conditions +	Glycerol %	Volume added (µl)	Time in solution (minutes)	Comments
0.1M Na Hepes pH 7.5, 0.8M Na/K dihydrogen Phosphate and R5 PS at 21mM.	5	+2	2	Similar to the RS43 C3 well cryo-protection details. Previous crystals produced a $P2_1$ space group suggesting that no R7 LPS was bound to the rfhSP-D and so R5 PS was added as a soak.
	10	+2	2	
	15	+2	2	
	20	+2	2	
	20	+2	2	
	20	-10 (solution reservoir) + 10 of 20% Glycerol– Exchanged	6	

Table 4.6 The cryo-protection method performed on the RS44 B4 well. This well was cryo-protected using Glycerol 5, 10, 15 and 20% where the R5 PS was present in each cryo solution at ≈21 mM.

In total 13 cryo-protected crystals were collected and flash frozen for testing at Diamond Light Source. The cryo-protection was successful in cryo-protecting the crystals, as no ice ring were detected during data collection.

4.6.2 Data collection

Crystal number 12 from RS44 B4 (RS44 B412) was exposed on 9/7/15 on the I04-1 beamline. This crystal had been co-crystallised with rfhSP-D with R7 *S. enterica* minnesota LPS at 4.67 mM + 30% (v/v) absolute ethanol + 0.25 mM of Triton X-100, and then R5 PS at ≈ 21 mM was soaked in during the cryo-protection stages. This suggests that this particular crystal could have interacted with two similar ligands. The crystal morphology resembled a tiny square (**Fig. 4.9A**), with one of corners cracked. The initial exposure of this crystal (SNAPS), showed positive protein diffraction spots with no sign of ice rings in the diffraction images. Following the snap EDNA suggested a maximum resolution of 3.0 Å. The parameters used in the snap and data collection are shown in **Table 4.7**.

SNAP or DC	Start Angle (˚)	φ (˚)	Total φ (˚)	Number of Images	Exposure per Image (s)	Total Exposure (s)	Maximum Resolution (Å)	Detector Distance (mm)	Transmission (%)
SNAP	0	0	90	3	0.5	1.5	3.0	578.5	100
DC	96	0.2	125	625	0.1	62.5	2.5	474	100

Table 4.7 Parameters used for the initial SNAP and for the RS44 B412 data collection. SNAP- first diffraction; DC- data collection; φ (phi) – angle of diffraction.

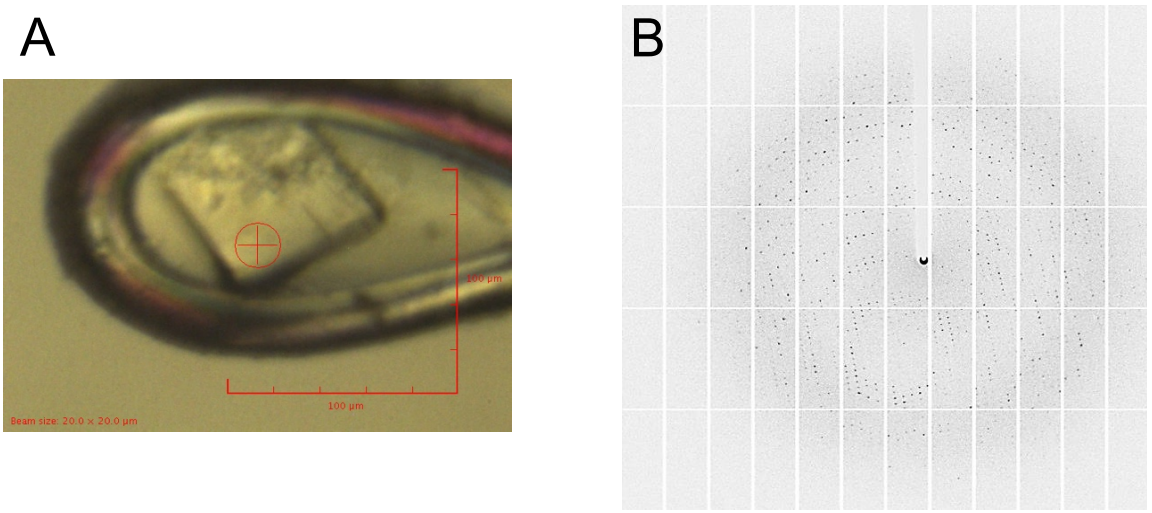


Fig 4.9 Images collected at DLS for the RS44 B412 crystal **A-** On the left, the picture of the crystal before a snap was taken to evaluate the quality of the crystal. **B-** On the right, the diffraction pattern produced following one of the snaps with the characteristics Braggs spots, proving that the crystal was protein.

4.6.3 Data Processing

The RS44 B412 diffraction spots and images were indexed and integrated in the $P42_12$ space group using *iMOSFLM*, not the usual space seen for rfhSP-D crystals, perhaps influenced by the presence of the R7 ligand in the co-crystallisation. Using *AIMLESS* various different runs were investigated, where batches of images containing a high R_{merge} were unselected and the resolution of the data limited (decreased). The data resolution was eventually limited to 3.3 Å, with a great improvement in the R_{merge}. *TRUNCATE* was used to investigate the twinning of the data set, with no twinning reported. *UNIQUEIFY* was used to add the free R-flag, while *RE-INDEX* was used to re-index the space group into $P42_12$. See **Table 4.8** for the data processing parameters following in house processing.

Data collection Parameters	RS44 B412		
Overall, InnerShell and OuterShell	Overall	InnerShell	OuterShell
Wavelength (Å)	0.92000		
Low Resolution limit (Å)	64.28	64.28	3.56
High Resolution limit (Å)	3.3	8.73	3.3
Space Group	P4 2_1 2		
Unit Cell lengths (Å)	a=113.70, b=113.70, c=107.00		
α, β, γ (°)	90, 90, 90		
Number of reflections	54188	2822	11577
Unique reflections	10886	641	2227
Average I/σ	8.7	14	4.8
CC (1/2)	0.987	0.998	0.906
Completeness %	98.9	96.8	99.5
$R_{merge\ \%}$	12.9	3.9	30.8
Multiplicity	5.0	4.4	5.2
Fast_dp (DLS) or processed in house	Processed in house		

Table 4.8 The data processing parameters for RS44 B412. This data set produced the $P42_12$ tetragonal space group, also reported in chapter 3, suggesting that R7 LPS was perhaps influencing crystal contact formation.

4.6.4 Rigid Body Refinement

The molecular replacement was performed by Dr. *Annette Shrive* using the ligand-free rfhSP-D structure (PDB: accession code 1PW9) and using *AMoRe* program (CCP4). The R-factors for the *REFMAC5* rigid body refinement for the RS32 A12 data were 38.27% for the R_{free} and 39.74% for the R-factor. This model contained the coordinates for the three chains of rfhSP-D minus the calcium and waters. The rigid body refined model allowed a map to be calculated for the inspection of the binding site in terms of the presence of the ligand and Ca^{2+} and the investigation of the overall structural organisation of key residues in each of three subunits.

4.6.5 Initial Electron Density map

The direct electron density map ($2F_{obs}$ - F_{calc}) and the difference map (F_{obs} - F_{calc}) were initially created by using the rigid body refined ligand-free model for the phases. The electron density maps for RS44 B412 following rigid body refinement showed the usual overall structure of rfhSP-D, previously reported in the literature (**Shrive et al., 2003**; **Shrive et al., 2009**; **Wang et al., 2008**). The electron density map and the superimposed model corresponded well. The electron density maps showed the three chains A, B and C and the overall structure seen in the trimeric aggregate with the three C-terminal globular domains (236-355) and the three chains linked through the α-helical coiled-coil neck region (203-235). As for the Eagan CA7 PS bound structure reported in Chapter 3, significant differences were noted in all three chains, particularly in terms of the Ca2 and Ca3 occupancy (**Shrive et al., 2003**; **Shrive et al., 2009**; **Wang et al., 2008**).

As for the rfhSP-D - Eagan CA7 PS structure, only chain B showed density around the binding pocket to suggest bound ligand. The RS44 B412 crystal was in contact with two separate *S. enterica* minnesota strains R7 LPS (initial co-crystallisation) and R5 PS (added during cryo-protection) (see **Fig. 4.2**).

Chain A showed density for Ca1, shown in **Fig. 4.10,** which represented the first co-crystal $P42_12$ electron density map showing the presence of the Ca1, however no density for Ca2 and Ca3 was seen. As for the rfhSP-D - Eagan CA7 structure (Chapter 3) a symmetry related molecule sits above the Chain A binding pocket within distance for potential intermolecular contact. Chain B is the only subunit showing suitable density for ligand building (**Fig. 4.11**), with density seen for Ca1, but no density for Ca2 or Ca3. Chain C, as seen in all $P42_12$ crystals, resembled the typical unliganded rfhSP-D binding pocket, with density for Ca1 and a symmetry related molecule in close proximity to the binding pocket, potentially impacting on the binding of the large ligands. Finally, no density for the presence of Ca4 could be seen, as reported for the Eagan CA7 complex in Chapter 3.

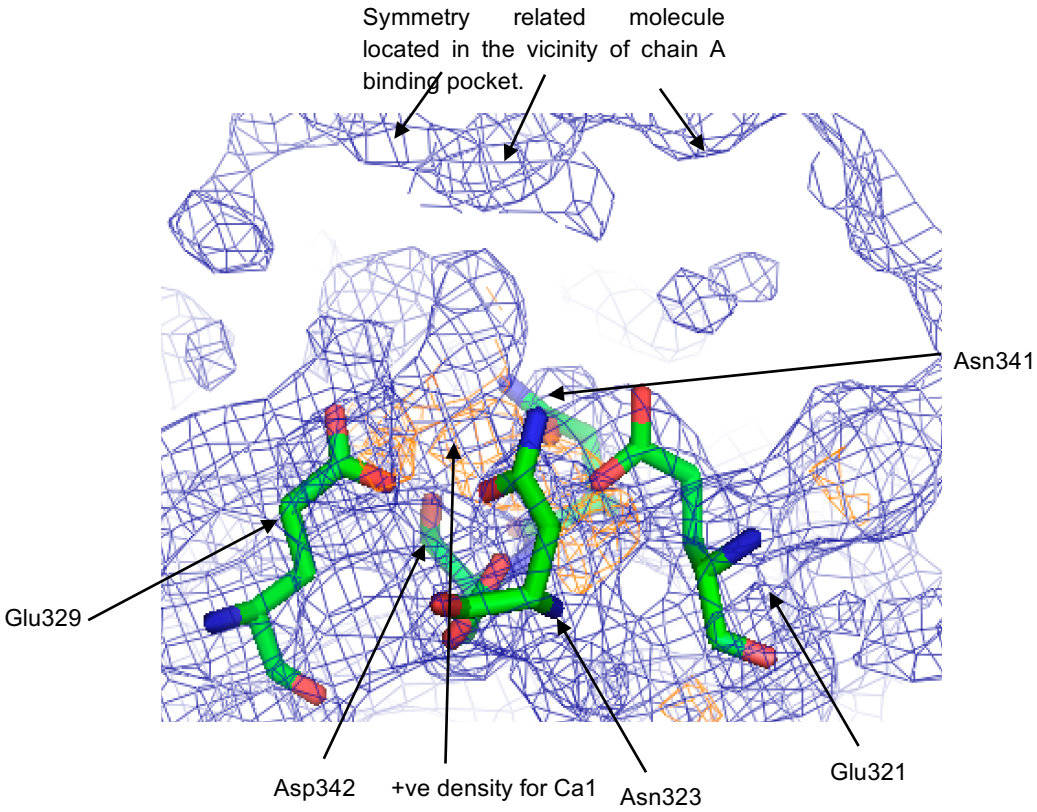


Fig. 4.10 Electron density map of RS44 B412 showing the chain A binding pocket, with density for Ca1 and the nearby symmetry related molecule. The location of the symmetry related molecule suggests no ligand could fit in in the tight space between the chain A binding pocket and the symmetry related molecule. In the figure Glu321, Asn323, Glu329, Asn341 and Asp342 are labelled; the direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ; the difference map (Fobs-Fcalc) is displayed in orange for positive (+ve) density and contoured at 2.5σ. In this image the density has been expanded (radius of contour) to display the symmetry related molecule. Figure generated using *PYMOL*.

4.6.6 Restrained Refinement

Restrained refinement was carried out for RS44 B412 as this offered the only co-crystal with suggested ligand presence. The RS44 B21 best model was completed using *REFMAC5* and *COOT*. R-factors improved from the initial rigid body refinement for the final model of RS44 B21 and were 25.01% for the R-factor (or R_{work}) and 29.46% for the R_{free}. The Ramachandran analysis showed that 95.29% (*MOLPROBITY*) of the residues modelled were in the preferred regions. No waters were modelled as this resolution makes it very difficult to determine the waters location. Some positive density in the map was not modelled, such as for weakly defined solvent molecules.

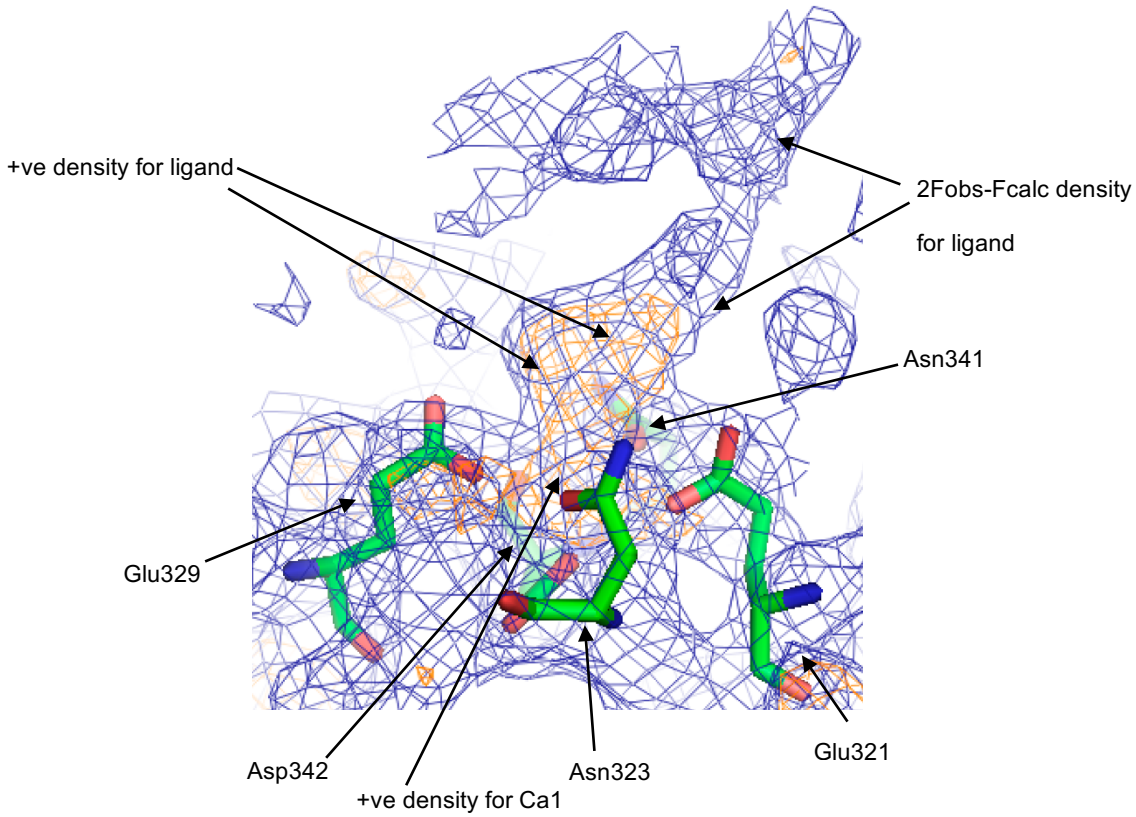


Fig. 4.11 RS44 B412 electron density map produced following the rigid body for the binding pocket in chain B. In the figure the direct map (2Fobs-Fcalc) in blue and countered at 1σ and the difference map (Fobs-Fcalc) positive (+ve) density in orange and countered at 2.5σ are both displayed. In the figure positive density can be seen for Ca1 (not modelled at this stage) and density for the ligand important Ca1 coordinating residues are displayed, including Glu321, Asn323, Glu329, Asn341 and Asp342 and these are labelled in the image. Figure generated using *PYMOL*.

The decision not to model some of the density seen between particular residues or sequences was made due to the low resolution of the data set and the high number of molecules that may be present in the crystal including; Na Hepes, Na/K dihydrogen Phosphate, Ethanol, Triton X-100, Glycerol, water, *S. enterica* minnesota R7 LPS and R5 PS strains. This omission may contribute to relatively high values for R_{work} and R_{free}.

4.6.7 Molecular Model Building

The electron density seen in binding pocket of chain B suggests a heptose, however, because this particular crystal was exposed to both *S. enterica* minnesota R7 LPS and R5 PS, created some difficulties in the characterisation of the correct *S. enterica* minnesota fragments. A heptose carbohydrate was initially modelled into the electron density suggesting a good fit between model and density and evidence that heptose was bound to rfhSP-D. The electron density of the refined heptose model clearly suggested further density at the C1 and C3 linkages of the heptose (**Fig. 4.12**).

Due to the fact that the RS44 B412 crystal had been exposed to two similar LPS of *S. enterica* minnesota it was necessary to further model the linkages in order to attempt a characterisation of the bound ligand, in this case *S. enterica* minnesota R7 LPS or R5 PS. *S. enterica* minnesota R7 LPS contains the LipidA-KdoIII-KdoII-KdoI-HepI-HepII structure, while *S. enterica* minnesota R5 PS contains the KdoIII-KdoII-KdoI-HepI-HepII-Glc structure, with both strains containing the same KdoIII-KdoII-KdoI-HepI-HepII core pentasaccharide (**Fig. 4.2**). In order to differentiate which ligand had bound to the rfhSP-D chain B, the electron density surrounding the binding pocket would have to suggest the presence of a Glc off HepII, meaning that R5 PS was bound, or the presence of the lipid A, suggesting that the R7 LPS was bound. During the extensive and long refinement process no convincing evidence could be found for the presence of Glc or lipid A. Therefore, no suggestion can be made about which ligand has bound.Due to the fact that the RS44 B412 crystal had been exposed to two similar mutants of *S. minnesota* it was necessary to further model the linkages in order to attempt a characterisation of the bound

ligand, in this case *S. minnesota* R7 LPS or R5 PS. *S. minnesota* R7 LPS contains the LipidA-KdoIII-KdoII-KdoI-HepI-HepII structure, while *S. minnesota* R5 PS contains the KdoIII-KdoII-KdoI-HepI-HepII-Glc structure, with both mutants containing the same KdoIII-KdoII-KdoI-HepI-HepII core pentasaccharide. In order to differentiate which ligand had bound to the rfhSP-D chain B, the electron density surrounding the binding pocket would have to suggest the presence of a Glc off HepII, meaning that R5 PS was bound, or the presence of the lipid A, suggesting that the R7 LPS was bound. During the extensive and long refinement process no convincing evidence could be found for the presence of Glc or lipid A. Therefore, no suggestion can be made about which mutant has bound.

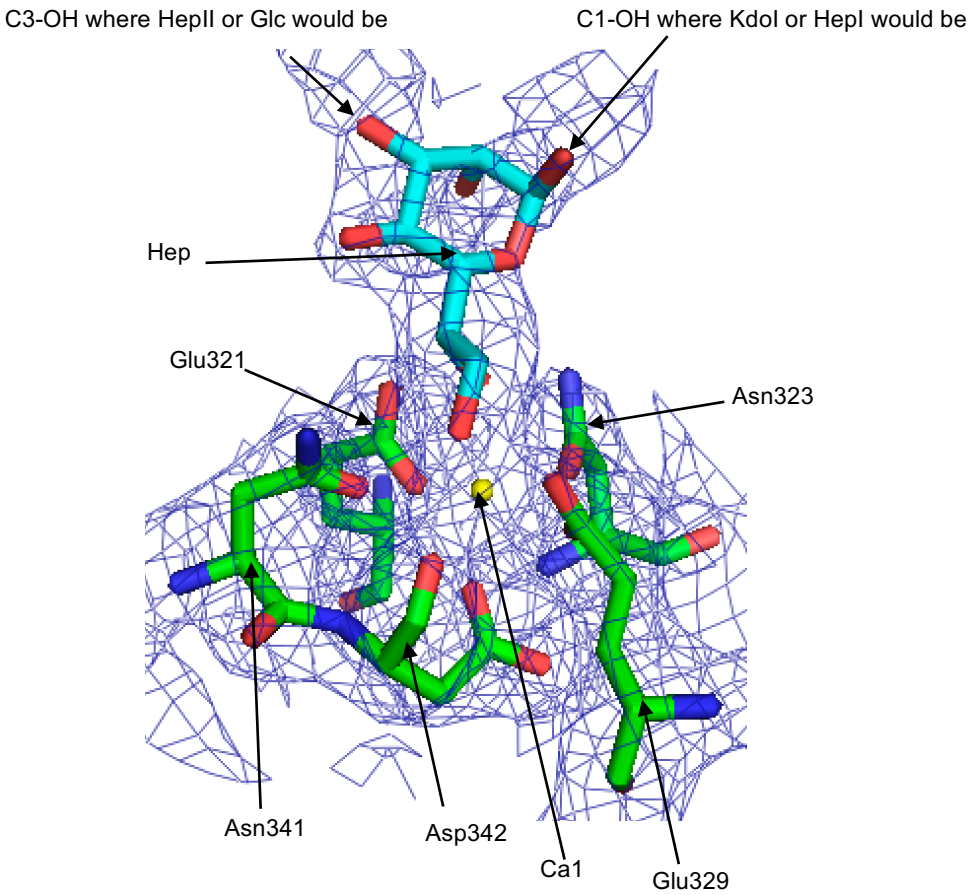


Fig. 4.12 The electron density for the RS44 B412 crystal with L-D-heptose orientated as reported in the literature (**Clark et al., 2016**) and after refinement PDB: 4E52. The density suggests linkages and likely orientations of other linked carbohydrates. In the figure Ca1 is displayed as a yellow sphere, with the Ca1 coordinating residues also displayed (Glu321, Asn323, Glu329, Asn341, Asp342); Heptose is displayed with the carbons in blue and with the oxygens in red. The direct map (2Fobs-Fcalc) is displayed in blue and contoured at 1σ. Figure generated using *PYMOL.*

The electron density (**Fig 4.12**) suggests that a heptose is bound at the Ca1 site and there are indications of linkages at Hep C1 and C3. Two possible ligand models could explain the density at the linkages of heptose C1 and C3; Model 1- KdoI-HepI-HepII (R7 LPS or R5 PS), with HepI bound or Model 2- HepII-HepI-Glc (R5 PS), with HepII bound. Both models were tested and refined. For the trisaccharide Kdo-HepI-HepII, the atom types, topology and chemical parameters (bond lengths and angles) were produced using *PRODRG*, a CCP4 supported program and part of the CCP4 suite. *PRODRG* produced an output pdb containing the ligand and a crystallographic information file (CIF) containing all the information specific to the trisaccharide. The CCP4 monomer library did not include description of Kdo-Hep linkages and chemical parameters ("**CCP4 Program Suite: *PRODRG"***). Upon examination of the refined electron density maps, model 1- KdoI-HepI-HepII, with HepI bound to rfhSP-D, is suggested as the best model to explain the electron density. In addition, Model 2 was discarded as a possible model for two separate reasons; Firstly, for Model 2 the HepII L-glycero group containing the O6' and O7' hydroxyls sit on the opposite site of the binding pocket to the density seen after refinements (**Fig. 4.13A**). Any attempts to place the L-glycero hydroxyls groups into the density seen were unsuccessful, as the HepII is bound by C3 to C1 of HepI with the L-glycero group located on the opposite site of the ring to where density was present. In order to place the L-glycero hydroxyls in the observed density the HepI-HepII linkage would have to change from the typical α1-3 to a α 1-4, as illustrated in **Fig. 4.13**. Secondly, for Model 2 there is electron density present in the vicinity of O4' of HepII (**Fig. 4.13A**) whereas there is no linkage off HepII at position 4 for *S. enterica* minnesota R7 LPS or R5 LPS (**Fig. 4.2**). Model 1 was refined to R_{free} 29.46%, while Model 2 R_{free} 29.54%. Model 2 was rejected in favour of Model 1 by taking into consideration the R_{free}, direct visualisation of the electron density maps and the potential linkage seen near where HepI (model 2) or KdoI (model 1) would be situated, with only model 1 KdoI explaining the potential linkage to KdoII. The above considerations suggest that model 1 offers a better fit than model 2 and so it is suggested as the most likely model, with HepI bound to rfhSP-D chain B binding pocket.

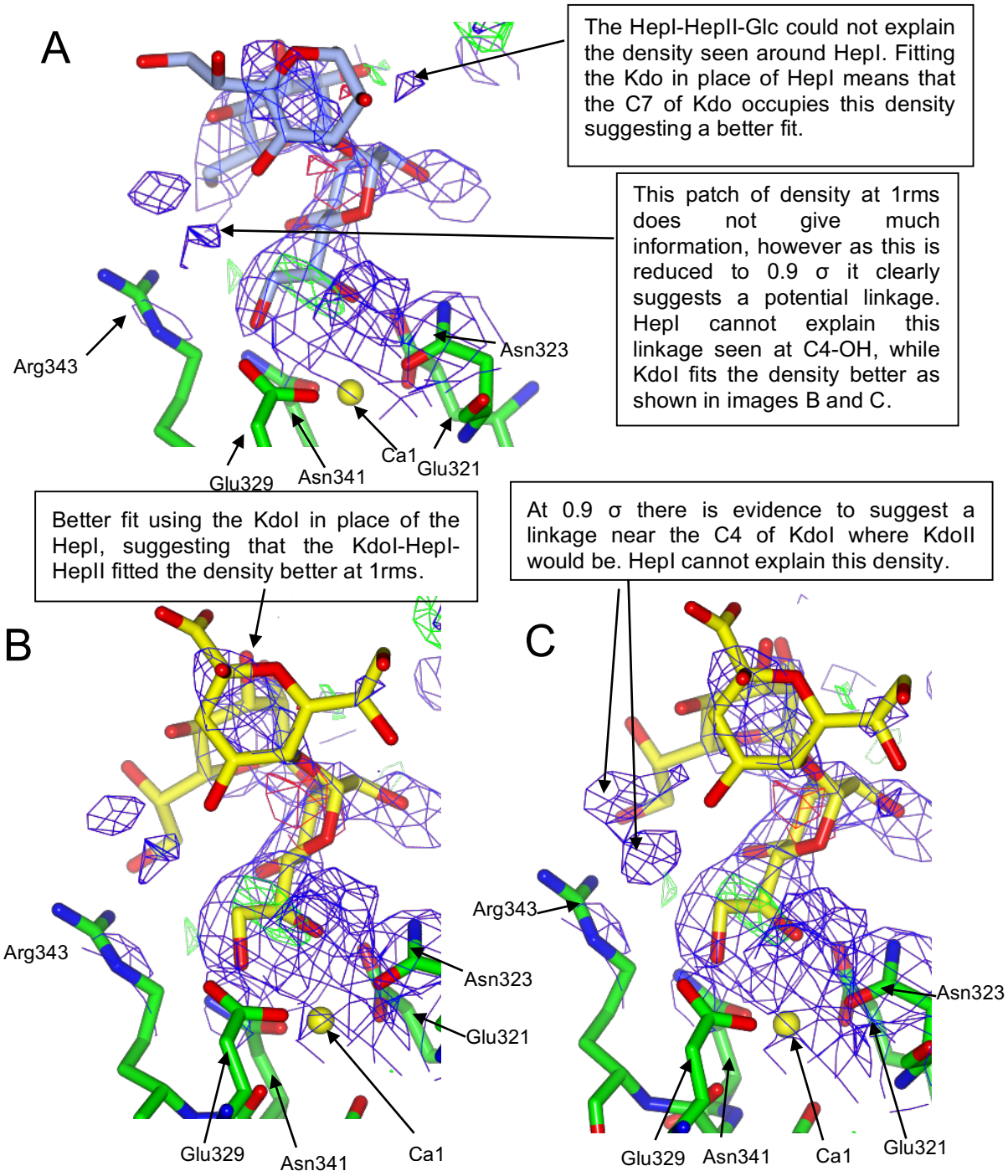


Fig. 4.13 The electron density seen for both Model 1 Kdo-HepI-HepII with HepI bound and Model 2 HepII-HepI-Glc with HepII bound to the rfhSP-D binding site. Model 1 displayed with carbons in yellow and oxygens in red (shown in images **B** and **C**) and model 2 displayed with carbons in light blue and oxygens in red. In images **A** and **B** 2Fobs-Fcalc map displayed and contoured 1σ, while image **C** contoured at 0.9σ, displayed in blue; Fobs-Fcalc contoured 3σ in all 3 images, with negative density displayed in red and positive density displayed in green **A-** Model 2 HepII cannot explain the density seen, with Model 1 offering a better fitting when the Kdo occupies the same place. **B-** Model 1 showing KdoI in place of HepII (image **A**). KdoI can explain the density seen on the right of HepII in image A, offering a better fitting than Model 2. **C-** Viewing the map at 0.9σ gives evidence to suggest another linkage, however HepI in image **A** cannot explain this density as there are no linkages coming off HepI in position 4; instead KdoI can explain this possible linkage as KdoI is linked to KdoII through position 4 of KdoI. Model 1 R_{free} 29.46% and Model 2 R_{free} 29.54%. Images generated using *CCP4mg.*

4.6.8 Final model

The electron density map of the binding pocket in chain B gives evidence to support Model 1 - KdoI-HepI-HepII, see **Fig. 4.14**, with HepI bound in the binding pocket through the O6' and O7' hydroxyl groups. The refinement statistics for the best model (model 1) are given in **Table. 4.9.**

Refinement statistics	
Data Range (Å)	64.28 - 3.3
No. of protein atoms	3457
No. of residues chain A and B	205-355
No. of residues chain C	206-355
No. of water molecules	N/A
No. of calcium ions	3
No. of ligand atoms	42
R_{work} (%)	25.01
R_{free} (%)	29.46
RMSD bond length (Å)	0.013
RSMD bond angle (°)	1.6
Average B-values ($Å^2$)[a]	
Protein main chain	44.6
Water	N/A
Ligand	63.282
Ramachandran plot values (%)[b]	
Favoured	95.29
Outliers	0.22

Table 4.9 The refinement statistics of the best model. [a] defined according to *Baverage*. [b] defined according to *MOLPROBITY*.

The orientation of the HepI hydroxyls agrees with previous reported structural studies showing L-glycero-D-manno-heptose bound in the binding pocket (**Wang et al., 2008**; **Clark et al., 2016**). This orientation is different from that reported here for *H. influenzae* CA7 PS (Chapter 3) where the hydroxyls sit interchanged through a 180° rotation of HepI (in relation to Ca1). The proposed binding of Hep1 follows the orientation

reported in the literature allowing for a direct comparison of the Ca1 coordination to both protein and ligand (HepI O6' and O7' hydroxyls) (**Table. 4.10**).

	Ca1 coordinating residues/ligand	Chain A (Å)	Chain B (Å)	Chain C (Å)	4A Chain B (Å)	4A Chain C (Å)
Ca1	Glu321 OE1	2.68	2.59	2.68	2.57	2.55
Ca1	Asn323 OD1	2.42	2.51	2.24	2.42	2.37
Ca1	Glu329 OE1	2.79	2.64	2.70	2.44	2.31
Ca1	Asn341 OD1	2.79	2.55	2.18	2.40	2.42
Ca1	Asp342 OD1	2.59	2.18	2.70	2.27	2.25
Ca1	Asp342 O	2.69	2.51	2.49	2.56	2.59
Ca1	Hep Hydroxyl - O6'	N/A	2.50	N/A	2.42	2.44
Ca1	Hep Hydroxyl - O7'	N/A	2.67	N/A	2.58	2.53

Table. 4.10 Ca1 coordination distances for RS44 B412 chains A, B and C. For chain B the distances between Ca1 and the Hep hydroxyls O6' and O7' are the distances calculated from the best model and shown in **Fig. 4.15** and **4.16**. The best model shows a similar Hep orientation to reported structures with the Hep hydroxyls O6' and O7' located in similar locations to those in the literature (**Wang et al., 2008**; **Clark et al., 2016**). Chain A and C showed no ligand occupancy. For comparison the *H. influenzae* Eagan 4A bound structure (PDB: 4E52) chain B and C is shown (**Clark et al., 2016**).

The O6' hydroxyl is interacting with Glu321 OE2 and Asn341 ND2 and the O7' is interacting with Glu329 OE2 in similar manner to reported in literature (**Wang et al., 2008**; **Clark et al., 2016**) (**Table 4.11**). The refinement of HepI showed good density to allow modelling of glyosidic bond extensions at C1 and C3. Hep1 at C1 is bound to KDO by α1–5, with HepII bound to HepI by α1–3. The model containing KdoI-HepI-HepII was further refined (R_{free} 29.46%), with the resulting density giving evidence for a likely orientation and best fit of KdoI-HepI-HepII. The density surrounding the Kdo is weak, with only a part of the sugar ring fitting the density at 1σ but the density surrounding C7 of the Kdo, and density seen near O4' where KdoII is expected to link, made it possible to suggest a likely orientation for KdoI. KdoI is modelled in the normal six membered ring structure, with no evidence to suggest structural change following hydrolysis (**Clark et al., 2016**).

Following various runs of refinement there was clear density to suggest the best orientation of HepII. For the KdoI and HepII no other interactions with the protein were

seen, however the density does suggest that both saccharides have little conformational freedom about the glycosidic linkage. This ordered orientation for both KdoI and HepII suggests that protein constraints or crystal contacts, through KdoIII-KdoII or the Glc if present, are holding both saccharides in place. In terms of the flanking residues Arg343 does not form any interactions with the KdoI-HepI-HepII. Asp325, located on the opposite site of the binding pocket, is within hydrogen bonding distance of O5' of HepI (2.67 Å).

		L-D-Heptose (**Wang et al., 2008**)			4A L-D-Heptose (**Clark et al., 2016**)			Final model, with HepI bound		
L-D-Heptose	Atom2	Chain A (Å)	Chain B (Å)	Chain C (Å)	Chain A (Å)	Chain B (Å)	Chain C (Å)	Chain A (Å)	Chain B (Å)	Chain C (Å)
6-OH	Glu321 OE2	2.50	2.56	2.52	N/A	2.50	2.59	N/A	2.80	N/A
	Asn341 ND2	3.19	3.00	3.22	N/A	2.92	3.06	N/A	2.88	N/A
	Asn341 OD1	2.98	3.30	3.27	N/A	3.03	3.05	N/A	3.70	N/A
7-OH	Asn341 ND2	3.16	3.02	3.02	N/A	2.85	2.84	N/A	3.37	N/A
	Glu329 OE2	2.69	2.68	2.63	N/A	2.69	2.62	N/A	2.49	N/A
	Asp342 O	3.68	3.32	3.10	N/A	3.00	2.94	N/A	3.13	N/A

Table 4.11 Distances between the HepI hydroxyls O6' and O7' and the Ca1 coordinating residues in the proposed Model 1 KdoI-HepI-HepII, with HepI bound in the binding pocket. The reported distances of the HepI of the *H. influenzae* Eagan 4A PS bound structure and in the disaccharide Hep bound structure, PDB: 4E52 and 2RIB respectively (**Clark et al., 2016**; **Wang et al., 2008**) are included for comparison. N/A implies no ligand presence or occupancy.

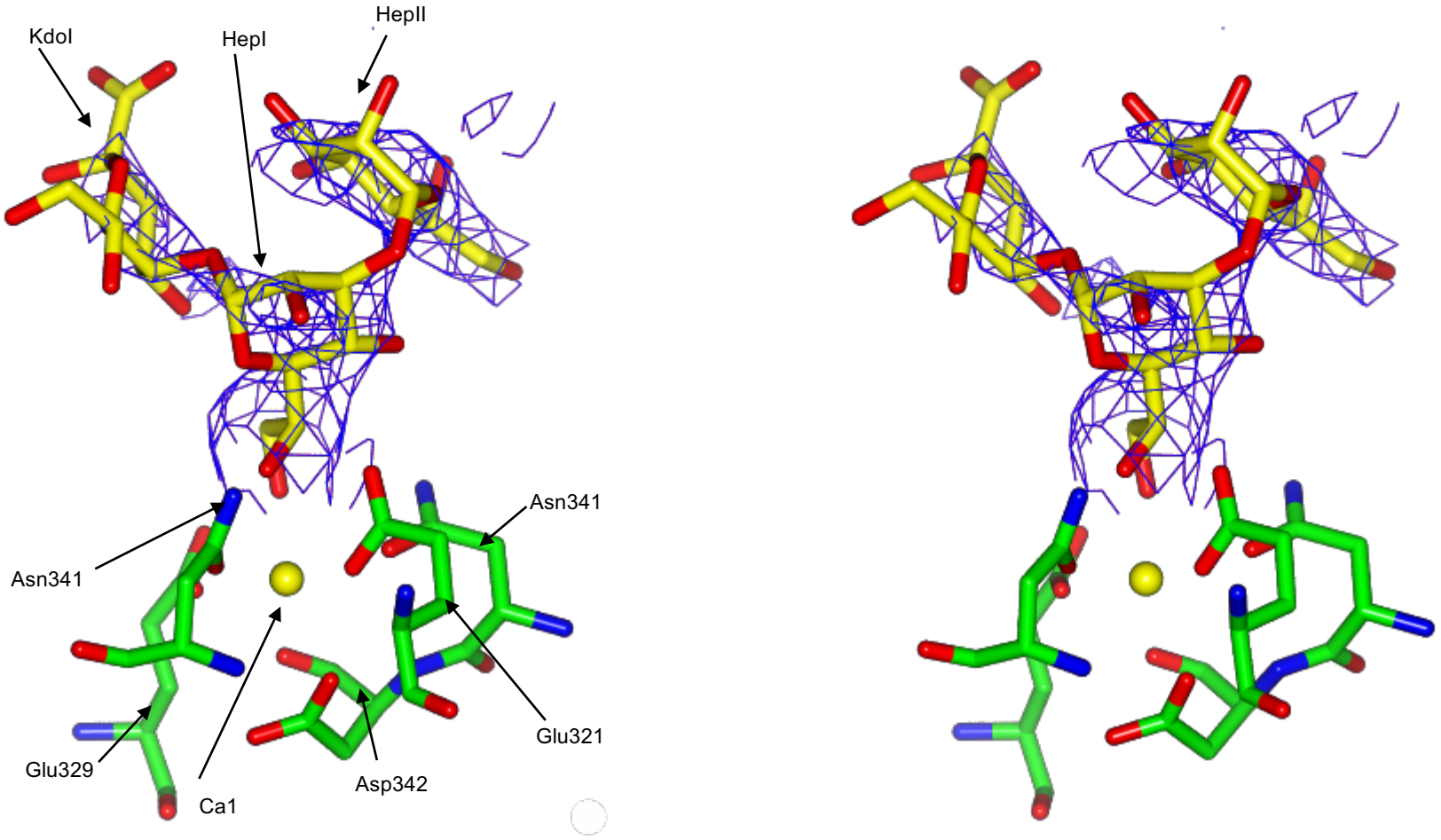


Fig. 4.14 The density in the electron density map from the RS44 B421 gives evidence to suggest that the KdoI-HepI-HepII (carbons in yellow) can be modeled, with HepI bound to the rfhSP-D. The HepI hydroxyls orientation is similar to reported in the literature (**Clark et al., 2016**; **Wang et al., 2008**) but different from that reported for the CA7 PS in chapter 3. Density for the main ring KdoI is relatively poor at 1σ, however at 0.9σ there is clear indications to suggest that the KdoI should be positioned in this orientation. Density for the HepII main ring is relatively strong, with the positions of the glycerol O6' and O7' a bit more ambiguous as these may have free rotation. Ca1 coordinating residues are displayed (Glu321, Asn323, Glu329, Asn341 and Asp342) displayed with the carbons in green. Model 1 is displayed with the carbons in yellow and oxygens in red. 2Fobs-Fcalc map contoured at 1σ and displayed in blue over model 1. Stereo Image generated using *CCP4mg.*

4.7 Summary and Conclusions

Triton X-100 and absolute ethanol have been used in combination with LPS (**Weiser et al., 1968**; **Li et al., 2004**; **Jang et al., 2009**; **Sharma et al., 2013**) to increase the solubility of LPS for use in structural studies. The initial experiments described here showed firstly that a 0.2-0.3 mM concentration of Triton X-100 can aid in the solubilisation of *E. coli* O111:B4 LPS up to 16 mg/ml, secondly that 30% (v/v) absolute ethanol can improve the solubility of O111:B4 LPS up to 20 mg/ml. By combining the two solvents, 0.2-0.3 mM Triton X-100 in combination with 30% (v/v) absolute ethanol, the *E. coli* O111:B4 LPS solubility was further improved, up to 35 mg/ml.

The co-crystallisation of rfhSP-D in combination with R7 LPS solubilised in 30% absolute ethanol (v/v) and 0.25mM Triton X-100 was initially carried out using the SS1 and SS2 100 buffer precipitants, with RS43 C3 (SS1-15), RS44 A2 (SS1-20), RS44 B4 (SS1-28), RS45 A2 (SS1-38), RS46 B6 (SS2-12) and RS47 A1 (SS2-19) wells producing positive crystal growth. The six conditions produced a variety of crystal morphologies including thin and thick square plates, triangles and some irregular shapes. Eight follow ups were produced from the initial SS1 and SS2 positive crystal growth conditions, but these did not produce any crystals.

Out of the 33 crystals collected and tested at DLS the majority of the crystals were of protein which diffracted to medium to low resolution (>4.0Å). Initially four crystals were analysed further as these offered the best overall data; RS44 A22 - 3.74 Å, RS45 A21 – 2.01 Å, RS46 B63 – 2.02 Å and RS47 A11 – 2.52 Å. The maps produced for the RS44 A22, RS45 A21, RS46 B63 and RS47 A11 crystals, all of which were in the $P2_1$ space group, showed no signs of ligand occupancy, with a typical ligand-free rfhSP-D trimer as reported in the literature (**Hakansson et al., 1999**; **Shrive et al., 2003**). As all these co-crystallised R7 LPS – rfhSP-D crystals did not contain bound ligand, a second ligand (*S. enterica* minnesota R5 PS) was soaked into the remaining two R7 LPS co-crystallisation wells RS44 B4 and RS43 C3. Crystal RS43 C35 diffracted to 3.00 Å in the space group

$P2_1$ but again the electron density map did not indicate bound ligand (R7 LPS or R5 PS). Crystal RS44 B412 diffracted to 3.30 Å and was indexed as space group $P42_12$. The soaked R5 PS was at a very high molecular excess, ≈ 50 x PS to rfhSP-D chain. Whether the crystal had formed in $P42_12$ due to the influence of the solubilised R7 LPS or for other reasons associated with the crystallisation conditions is not clear.

Despite the limited resolution of the data, the electron density in the binding pocket of chain B gave a strong indication of ligand and its orientation; however, as two different *S. enterica* minnesota strains with a common trisaccharide core were in contact with the crystal, discriminating between the two did not prove possible and no evidence can be given here as to whether the bound ligand is R7 LPS or R5 PS as both have the same core pentasaccharide KdoIII-KdoII-KdoI-HepI-HepII with an additional terminal glucose off HepII in the R5 strain.

Two models were initially built with Model 1 refined to R_{free} 29.46%, while Model 2 refined to R_{free} 29.54%. The final model 1 containing the KdoI-HepI-HepII trisaccharide in chain B, with HepI bound to Ca1 is suggested as the best model. The final R_{free} achieved is somewhat high, however considering that the data resolution was limited to 3.3Å and there was significant electron density which could not be modelled in, the final R_{free} of 29.46% does represent an acceptable and meaningful model. The additional electron density may be due to a variety of molecules or ions associated with the RSB44 B4 well including, Na Hepes, Na/K dihydrogen Phosphate, Ethanol, Triton X-100, Glycerol and Water (**Read et al., 2011**).

The orientation of HepI was similar to that previously reported in literature for Hep bound in the rfhSP-D CRD binding site, with the O6' hydroxyls positioned in the vicinity of Glu321 and Asn323 and O7' in the vicinity of Glu329 and Asn341 (**Wang et al., 2008**; **Clark et al., 2016**). The HepI ring orientation is very similar to the ring orientation reported by Wang and co-workers (PDB: 2RIC), and by Shrive and co-workers (PDB: 4E52) with

KdoI sitting in similar position to the reported anhydro Kdo in Eagan 4A LPS (**Wang et al., 2008**; **Clark et al., 2016**). Further interactions between protein and HepI include the Asp325 flanking residue, which is within distance to hydrogen bond with O5' of HepI (2.67 Å).

The electron density seen for the glycosidic extensions of the bound HepI suggest a particular orientation of KdoI and HepII. The density surrounding the linkage and ring of HepII is fairly strong, while the density surrounding the KdoI is poorer. The density seen for KdoI is rather ambiguous, with some of the main ring not fitting the density at 1 rms, however density can be seen to explain the glycerol groups at C7-OH and C8-OH, which may be fixed by interaction with non-modelled water molecules, while the density in the vicinity of the C4-OH of KdoI could arise from the α2-4 linkage between KdoII and KdoI. The KdoI reported here best fits the electron density as a normal six membered ring and not the 5-membered anhydro Kdo reported for the hydrolysed Eagan 4A LPS (**Clark et al., 2016**). In terms of Ca^{2+} occupancy, and contrary to the Eagan CA7 study reported in Chapter 3, the Ca1 in chain A was present in the typical unliganded binding pocket (**Shrive et al., 2003**; **Shrive et al., 2009**), with no Ca2, Ca3 and Ca4 present.

To date no structural studies of rfhSP-D in complex with *S. enterica* minnesota rough strains have been reported, the results presented here offering for the first time an insight into the recognition of *S. enterica* minnesota LPS by hSP-D. Ligand binding studies using *S. enterica* minnesota strains complexed with hSP-D and rfhSP-D have reported that *S. enterica* minnesota Re strains, containing just a Kdo disaccharide bound to the lipid A and lacking any heptose and glucose, did not bind to either SP-D or rfhSP-D (**Wang et al., 2008**). This suggests that the presence of inner core heptoses, or terminal glucoses, is a key factor in the recognition of *S. enterica* minnesota LPS by SP-D. The results presented here demonstrate that the preferred mode of binding is via the inner core heptose HepI.

Chapter 5 - General discussion, Conclusions and Future Work

The work presented in this thesis aims to increase our knowledge of the recognition of gram-negative bacteria by surfactant protein D through structural studies of rfhSP-D complexed with *H. influenzae* Eagan CA7 PS and *S. enterica* minnesota R7 LPS and R5 PS. The co-crystals reported in this work were an essential advance as all previous attempts to soak Eagan CA7 PS in to ligand-free crystals had failed and this LPS represents an important step in the characterisation of hSP-D recognition of *H. influenzae* following on from the Eagan 4A structural study (**Clark et al., 2016**). Unfortunately, while ligand-free rfhSP-D crystals soaked with ligand generally diffract to high resolution (**Shrive et al., 2003** and **2009**; **Clark et al., 2016**), weak diffraction and low maximum resolution were a recurrent problem for many of the crystals of rfhSP-D grown in the presence of LPS ligands.

5.1 Co-crystallisation and crystal form

The idea of using the co-crystallisation approach was firstly in order to utilize longer ligands in combination with rfhSP-D, initially for the Eagan CA7 PS and *S. enterica minnesota* R7 LPS, which on all previous occasions had led either to crystals degrading and dissolving rapidly or there being no bound ligand seen in the structure (for those that didn't dissolved). Secondly, this approach might be applicable as a more general method for other ligands which could not be soaked in to the ligand-free crystals. If the right ligand is chosen (known to bind to SP-D) and the right conditions are in place (buffer precipitant and concentration of protein and ligand), co-crystallisation could produce a different crystal form for rfhSP-D with bound ligand in place.

5.2 Calcium binding

The absence of Ca2 and Ca3 in all the structures arising from the crystallisation conditions which produce the tetragonal crystal form (see **Table 5.1**) is somewhat puzzling. Ca2 and Ca3 have been suggested to be important in the stability of the CRD,

with the loss of Ca2 reported to lead to movement of loop L1 and changes in loop L4 (**Ng et al., 1998**). The potential impact of the loss of Ca2 and Ca3 leads to L4 being located further away from Ca1.

	Chain	**A** CA7 PS Final Model	**B** R7 LPS/R5 Final Model	**C** Other $P42_12$ Co-crystals	**D** Other $P2_1$ Co-crystals	**E** 4A PS
Ca1	A	X	**P**	X	**P**	**P**
Ca2	A	X	X	X	**P**	**P**
Ca3	A	X	X	X	**P**	**P**
Ca1	B	**P**	**P**	**P**	**P**	**P**
Ca2	B	X	X	X	**P**	**P**
Ca3	B	X	X	X	**P**	**P**
Ca1	C	**P**	**P**	**P**	**P**	**P**
Ca2	C	X	X	X	**P**	**P**
Ca3	C	X	X	X	**P**	**P**
Ca4	Neck-CRD interface	X	X	X	X	X
Ligand	A	X	X	X	X	X
	B	**P**	**P**	**V**	X	**P**
	C	X	X	X	X	**P**

Table. 5.1 Direct comparison of the different Ca^{2+} occupancy. **A-** The final model for the CA7 PS, where HepI is bound in the novel orientation (Chapter 3). **B-** The final model for the R7 LPS/R5 PS model containing the Kdol-HepI-HepII trisaccharide with HepI bound (Chapter 4). **C-** Includes RS34 A22, RS32 A31, RS32 A12 and RS32 B27 (Chapter 3) and $P42_12$ crystals of rfhSP-D in complex with *E. coli* J-5 PS (data not shown). **D-** Includes RS43 C35, RS44 A22, RS45 A21, RS46 B63 and RS47 A11. **E-** for comparison the *H. influenzae* Eagan 4A bound structure (PDB: 4E52) is shown (**Clark et al., 2016**). **P-** present, X- not present, **V-** variable.

At first sight it may be proposed that the loss of the Ca2 and Ca3 could be caused by insufficient Ca^{2+} during crystallisation and cryo-protection. However, the concentration of Ca^{2+} used during the crystallisations here is in line with previous work performed over many years in our research group, where no incomplete Ca^{2+} occupancy in any of the chains has been seen. In addition, the unliganded crystals of *S. enterica* minnesota R7 LPS ($P2_1$) where co-crystallisation with rfhSP-D had been unsuccessful were produced using standard Ca^{2+} concentrations in the crystallisation and cryo-protection stages and all showed the presence of Ca2 and Ca3 for all the chains of rfhSP-D (**Table 5.1**).

Incomplete Ca2 and Ca3 occupancy has been previously reported in the literature in the structural study of a trimeric recombinant fragment of rat SP-A, containing neck and CRD domains, demonstrating the presence of Ca1 but with Ca2 and Ca3 absent (**Head et**

al., 2003). Additional investigations performed by Head and co-workers, using Samarium ions during MAD phasing experiments, reported a metal ion binding site located in the vicinity of the Ca2 site in rfhSP-D (**Shrive et al., 2003**; **Head et al., 2003**). Head and co-workers suggested that the Ca1 site is filled before the Ca2 in the recombinant fragment of rat SP-A and this may also be the case for rfhSP-D co-crystals, particularly where the crystallisation buffer and additives compete with rfhSP-D for calcium ions.

A possible contributory reason for the loss of Ca2 and Ca3 in rfhSP-D molecules may be the effect of the PS and LPS on the Ca^{2+}, as LPS is known to be highly reactive with aqueous metal cations such as calcium (**Selvarengan et al., 2010**). LPS-LPS interactions when membrane bound have been reported to be modulated by Na^{+}, Ca^{2+} and Mg^{2+}, with these cations playing key roles in the organisation of LPS, by providing bridging potential between negatively charged groups and reducing the strong negative electrostatic repulsion between LPS molecules (**Kucerka et al., 2008**). Divalent cations such as Ca^{2+} can induce changes in terms of hydration, fluidity and aggregate structure of the LPS membranes (**Nascimento et al., 2013**), leading to Ca^{2+} compacting the LPS molecules and preventing water from penetrating the outer/inner core of the membrane bound LPS (**Kucerta et al., 2008**). Furthermore, LPS phosphate and carbonyl groups have been reported to bind to Ca^{2+} while remaining hydrated by H_2O, achieving a double layer around the phosphorylated carbohydrates (**Dias et al., 2014a**). This suggests that the LPS and PS molecules used here can chelate Ca^{2+} present in the surrounding environment and that this may be a contributor to the loss of Ca2 and Ca3. The crystallisation method used for both Eagan CA7 and *S. enterica* minnesota R7 LPS introduces rfhSP-D molecules, the Ca^{2+} and the PS or LPS in solution together, with the rfhSP-D and PS or LPS in competition for the Ca^{2+}. Therefore, the lack of Ca2 and Ca3 can be suggested in part to be a consequence a direct competition for Ca^{2+} in solution by rfhSP-D molecules and PS or LPS molecules, with the PS or LPS clustering the Ca^{2+} and in effect impacting on the occupation of the Ca2 and Ca3 sites (**Dias et al., 2014**; **Dias et**

al., 2014a). This may also extend to an influence on the loss of Ca1 in some subunits where ligand is not seen to bind, for example Chain A in the final Eagan CA7 PS structure where spatial restraints appear to prevent ligand binding.

Alongside the impact of the ligand, the crystallisation and cryo-protection conditions will certainly play a major role in determining the occupancy of all the calcium sites and it does appear that the conditions which drive the crystallisation in to the tetragonal space group are also associated with the loss of Ca2 and Ca3, and the loss of Ca1 in those chains where ligand cannot, or does not, bind. The exception to this appears to be the rfhSP-D crystal which was crystallised in the presence of *S. enterica* minnesota R7 LPS and subsequently soaked with R5 PS through the cryo-buffer. This reinforces the influence of not only the crystallisation conditions but also the cryo-protection buffers.

5.3 Recognition and binding of LPS by rfhSP-D

The structures reported here provide evidence to suggest a significant advance in our understanding of how the recognition and binding of bacterial LPS by collectins occurs. Previous work reported in the literature revealed not only the bound sugar interacting with Ca1 and binding site residues through a pair of hydroxyls positioned in a similar manner to the mannose equatorial pair O3' and O4', but also that the orientation of the bound sugar in the binding site, interactions of the linked sugars with the binding site flanking residues, and spatial and steric constraints also play major roles in determining binding affinity. Both *H. influenzae* Eagan CA7 PS (Chapter 3) and *S. enterica* minnesota R5/R7 (Chapter 4) display HepI bound to the Ca1, with secondary protein-ligand interactions also playing key roles in determining binding affinity.

5.3.1 Orientation of the bound Heptose

The binding orientation of the bound Hep in the *H. influenzae* Eagan 4A PS structure (**Clark et al., 2016**) was reported to be similar to the Hep binding reported by Wang and co-workers (**Wang et al., 2008)**, suggesting a preferred binding mode and

orientation of bound Hep. *S. enterica* minnesota R5/R7 HepI (Chapter 4) assumes the preferred binding mode with evidence from the model supporting the view that the alternative binding (180° rotation about C5-C6 which interchanges the hydroxyls which coordinate Ca1) was also possible, with no spatial or steric constraints. The *H. influenzae* Eagan CA7 structure reported here does however show the alternative orientation of the bound Hep, with a significant shift in terms of the angle of entry of the Hep ring with respect to the binding pocket. It is reasonable to assume that the different orientations make different overall interactions with the binding pocket as a whole, and that the affinity for the preferred orientation is greater than that for the alternative (the O6' and O7' heptose hydroxyl positions interchanged).

The structural data presented here and in the literature (**Clark et al., 2016**) suggest a preferred mechanism of inner core heptose recognition via the O6' and O7' heptose hydroxyls, even though a terminal Glc may also be available to bind through the O3' and O4' equatorial hydroxyls. This HepI recognition and binding is supplemented by additional interactions between the remainder of the LPS inner core and protein, when possible. However, the novel orientation of HepI for the *H. influenzae* Eagan CA7 mutant demonstrates that SP-D can also interact with LPS ligands through different HepI orientations, giving rise to increased flexibility in the recognition and binding of target pathogens.

5.4 Implications and Future Work

The innate immune system provides an efficient and rapid immunological response to a variety of pathogens. In humans and in vertebrates, upon infection the innate immune system has the ability to activate the adaptive immune system, leading to an enhanced and specialised response to the infection causing organism. The ability of the innate immune system to provide an effective and potent defence against pathogens is dependent in part on the recognition of the variety of PAMPs on the surface of the different pathogens, and DAMPs following necrosis, which is achieved by innate immune

proteins such as PRRs. Collectins are secreted PRRs with the ability to recognise a variety of glycoconjugates on the surface of bacteria, viruses and fungi.

rfhSP-D, a recombinant fragment of hSP-D, has been shown to be biologically and therapeutically active (**Madan et al., 2001**; **Hickling et al., 1999**; **Strong et al., 2003**; **Singh et al., 2003**; **Ikegami, 2006**); understanding the binding mechanisms of collectins to pathogens offers the potential to further develop chimeric or recombinant fragments that could one day be used in order to enhance the immune response towards particular pathogens. Gram negative bacterial LPS is a major inducer of sepsis, with increasing clinical complications in LPS induced infections. The problematic increase of antibiotic resistant bacteria is a growing public health threat, with current over use of antibiotics and the growing number of superbugs posing a significant health risk. Therefore, while it remains important to source and fund novel antibiotics, it is also important that the search continues for novel therapies that can be used in the recognition and clearance of bacteria.

The recombinant fragment rfhSP-D offers the potential to be used as a novel therapy in the recognition and activation of immune system towards bacteria/viral clearance. rfhSP-D has been widely investigated *in vitro* and *in vivo* with reported similar positive properties as the full length SP-D. However, before rfhSP-D can be considered as a therapeutic agent, it is important to understand how recognition occurs and what the roles of specific residues are in binding to pathogens. This is where structural studies come in, by elucidating at the atomic level the interactions between protein and the various pathogens and providing information on which residues are important. Knowledge of the residues that help achieve binding is key, as mutational studies have shown the impact of changing one or two amino acids in the binding pocket, leading to varied affinities towards different ligands.

The production of rfhSP-D with specific mutations to recognise specific pathogens, may then lead to a drug-free biological pathogen clearance specific treatment, currently with no known contraindications or side effects. Future structural studies will hopefully lead to a complete understanding of how best to modify and use rfhSP-D to target specific disease causing organisms, exploiting the untapped therapeutic potential of rfhSP-D (**Clark et al., 2010**).

References

Ahmed, M., Sing, A. K., Mondal, J. A. (2016). Hydrogen-bonding and vibrational coupling of water in a hydrophobic hydration shell as observed by Raman-MCR and isotopic dilution spectroscopy. *Physical Chemistry Chemical Physics.* **18:** 2767-2775.

Alcorn, J. F. & Wright, J. R. (2004). Degradation of Pulmonary Surfactant Protein D by *Pseudomonas aeruginosa* Elastase Abrogates Innate Immune Function. *The Journal of Biological Chemistry*. **279(29):** 30871-30879.

Alexander, C. & Rietschel, E. T. (2001). Bacterial lipopolysaccharide and innate immunity. *Journal of Endotoxin Research.* **7(3):** 167-202.

Allen, M. J., Laederach, A., Reily, P. J., Mason, R. J. (2001). Polysaccharide Recognition by Surfactant Protein D: Novel Interactions of a C-type Lectin with Nonterminal Glucosyl Residues. *Biochemistry.* **40:** 7789-7798.

Allen, M. J., Laederach, A., Reily, P. J., Mason, R. J., Voelker, D. R. (2004). Arg343 in human surfactant protein D governs discrimination between glucose and N-acetylglucosamine ligands. *Glycobiology.* **14(8):** 693-700.

Amor, K., Heinrichs, D., Frirdich, E., Ziebell, K., Johnson, R., Whitfield, C. (2000). Distribution of Core Oligosaccharides Types in Lipopolysaccharides from *Escherichia coli*. *Infection and Immunity*. **68(3):** 1116-1124.

Asherie, N. (2004). Protein crystallization and phase diagrams. *Methods.* **34:** 266-272.

Atochina-Vasserman, E. N. (2011). S-nitrosylation of surfactant protein D as a modulator of pulmonary inflammation. *Biochimica et Biophysica Acta.* **1820:** 763-769.

Auriti, C., Prencipe, G., Inglese, R., Azzari, C., Ronchetti, M. P., Tozzi, A., Seganti, G., Orzalesi, M., De Benedetti, F. (2010). Role of mannose-binding lectin in nosocomial sepsis in critically ill neonates. *Human Immunology*. **71:** 1084-1088.

Auzanneau, F-I., Charon, D., Szabo, L. (1991). Phosphorylated Sugars. Part27. 'Synthesis and Reactions, in Acid Medium of 5-O-Substitued Methyl 3-Deoxy-α-D-manno-oct-2-ulopyranosidonic acid 4-Phosphates. *Journal of Chemical Society.* **Trans 1:** 509-517.

Axelgaard, E., Jensen, L., Dyrlund, T. F., Nielsen, H. J., Enghild, J. J., Thiel, S., Jensenius, J. C. (2013). Investigations on Collectin Liver 1. *The Journal of Biological Chemistry*. **288:** 23407-23420.

Barreira, E. R., Precioso, A. R., Bousso, A. (2011). Pulmonary Surfactant in Respiratory Syncytial Virus Bronchiolitis: the Role of Pathogenesis and Clinical Applications. *Pediatric Pulmonology*. **46:** 415-420.

Barrow, A. D., Palarasah, Y., Bugatti, M., Holehouse, A. S., Byers, D. E., Holtzman, M. J., Vermi, W., Skodt, K. et al. (2015). OSCAR Is a Receptor for Surfactant Protein D That Activates TNF- α Release from Human $CCR2^{+}$ Inflammatory cytokines. *The Journal of Immunology.* **194(7):** 3317-3326.

Basset, C., Holton, J., O'Mahony, R., Roitt, I. (2003). Innate Immunity and pathogen-host interaction. *Vaccine*. **21:** S2/12-S3/23.

Beck, K. & Brodsky, B. (1998). Supercoiled Protein Motifs: The Collagen Triple-Helix and the α-Helical Coiled Coil. *Journal of Structural Biology.* **122:** 17-29.

Blanco, O. & Perez-Gil, J. (2007). Biochemical and pharmacological differences between preparations of exogenous natural surfactant user to treat Respiratory Distress Syndrome: Role of the different components in an efficient pulmonary surfactant. *European Journal of Pharmacology*. **568:** 1-15.

Borron, P. J., Crouch, E. C., Lewis, J. F., Wright, J. R., Possmayer, F., Fraher, L. J. (1998). Recombinant Rat Surfactant-Associated Protein D Inhibits Human T

Lymphocytes Proliferation and IL-2 Production. *The Journal of Immunology.* **161:** 4599-4603.

Botas, C., Poulain, F., Akiyama, J., Brown, C., Allen, L., Goerke, J., Clements, J., Carlson, E., et al. (1998). Altered surfactant homeostasis and alveolar type II cell morphology in mice lacking surfactant protein D. *PNAS.* **95:** 11869-11874.

Bouchet, V., Hood, D. W., Li, J., Brisson, J-R., Randle, G. A., Martin, A., Li, Z., Goldstein, R., et al. (2003). Host-derived sialic acid is incorporated into *Haemophilus influenzae* lipopolysaccharides and is a major virulence factor in experimental otitis media. *PNAS.* **100(15):** 8898-8903.

Brade, H., Moll, H., Rietschel, E. T. (1985). Structural Investigations on the Inner Core Region of Lipopolysaccharides from *Salmonella Minnesota* Rough Mutants. *Biomedical Mass Spectrometry.* **12(10):** 602-609.

Brade, H., Zahringer, U., Rietxhel, E. (1984). Spectroscopic analysis of a 3-deoxy-D-manno-2-octulosonic acid (KDO)-disaccharide from the lipopolysaccharide of a *Salmonella godesberg* Re mutant. *Carbohydrate Research*. **134:** 157:166.

Brinker, K. G., Martin, E., Borron, P., Mostaghel, E., Doyle, C., Harding, C. V., Wright, J. R. (2001). Surfactant Protein D enhances bacterial antigen presentation by bone marrow-derived dendritic cells. *American Journal of Physiology.* **281:** 1453-1463.

Bufler, P., Schmidt, B., Schikor, D., Bauernfeind, A., Crouch, E. C., Griese, M. (2003). Surfactant Protein A and D Differently Regulate the Immune Response to Nonmucoid *Pseudomonas aeruginosa* and Its Lipopolysaccharide. *American Journal of Physiology.* **28:** 249-256.

Caroff, M., Karibian, D., Cavaillon, J-M., Haeffner-Cavaillon, N. (2002). Structural and functional analyses of bacterial lipopolysaccharides. *Microbes and Infection*. **4:** 915-926

"CCP4 Program Suite: Blend". http://www.ccp4.ac.uk/html/blend.html, 2016. Web. 2 May 2016.

"CCP4 Program Suite: Prodrg". http://www.ccp4.ac.uk/html/cprodrg.html, 2016. Web. 2 May 2016.

Chandler, D. (2005). Interfaces and the driving force of hydrophobic assembly. *Nature.* **437:** 640-647.

Chaplin, M. (2007). Water's hydrogen bond strength. *Condensed Matter.* 0706.1355

Chessa, D., Spiga, L., De Riu, N., Delaconi, P., Mazzarello, V., Ganau, G., Rubino, S. (2014). Lipopolysaccharides Belonging to Different *Salmonella* Serovars are Differentially capable of Activating Toll-Like receptor 4. *Infection and Immunity.* **82(11):** 4553-4562.

Clark, H. W. (2010). Untapped Therapeutic Potential of Surfactant Proteins: Is There a Case for Recombinant SP-D Supplementation in Neonatal Lung Disease? *Neonatology.* **97:** 380-387.

Clark, H. W., Mackay, R-M., Deadman, M. E., Hood, D. W., Madsen, J., Moxon, E. R., Townsend, J. P., Reid, K. B. M., Ahmed, A., Shaw, A. J., Greenhough, T. J., Shrive, A. K. (2016). Crystal Structure of a Complex of Surfactant Protein D (SP-D) and *Haemophilus influenzae* Lipopolysaccharide Reveals Shielding of Core Structures in SP-D- Resistant Strains. *Journal of Infection and Immunity*. **84(5):** 1585-1592.

Clark, S. E., Elchelberger, K. R., Weiser, J. N. (2013). Evasion of killing by human antibody and complement through multiple variations in the surface oligosaccharide of *Haemophilus influenzae. Molecular Microbiology.* **88(3):** 603-618.

Clark, S. E., Snow, J., Li, J., Zola, T. A., Weiser, J. N. (2012). Phosphorylcholine Allows for Evasion of Bactericidal Antibody by *Haemophilus influenzae. PLoS Pathogens.* **8(3):** 1-14.

Corn, P. G., Anders, J., Takala, A. K., Kayhty, H., Hoiseth, S. K. (1993). Genes Involved in *Haemophilus influenzae* Type b Capsule Expression are Frequently Amplified. *The Journal of Infectious Diseases.* **167:** 356-364.

Cox, A. D., Li, J., Brisson, J-R., Moxon, E. R., Richards, J. C. (2002). Structural analysis of the lipopolysaccharide *Neisseria meningitidis* strain BZ157 galE: localisation of two phosphoethanolamine residues in the inner core oligosaccharide. *Carbohydrate Research.* **337:** 1435-1444.

Crisel, R. M., Baker, R. S., Dorman, D. E. (1975). Capsular Polymer of *Haemophilus influenzae,* Type b. I. Structural Characterization Of The Capsular Polymer of Strain Eagan. *The Journal of Biological Chemistry.* **250 (13):** 4926-4930.

Crouch, E. (2000). Surfactant protein-D and pulmonary host defense. *Respiratory Research.* **1(2):** 93-108.

Crouch, E., Hartshorn, K., Horlacher, T., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Seeberger, P. H., Head, J. (2009). Recognition of Mannosylated Ligands and Influenza A Virus by Human Surfactant Protein D: Contributions of an Extended Site and Residue 343. *Biochemistry.* **48:** 3335-3345.

Crouch, E., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Head, J. (2006). Contributions of Phenylalanine 335 to Ligand Recognition by Human Surfactant Protein D. *The Journal of Biological Chemistry.* **281(26):** 18008-18014.

Crouch, E., McDonald, B., Smith, K., Roberts, M., Mealy, T., Seaton, B., Head, J. (2007). Critical Role of Arg/Lys343 in the Species-Dependent Recognition of Phosphatidylinositol by Pulmonary Surfactant Protein D. *Biochemistry.* **46:** 5160-5169.

Crouch, E., Parghi, D., Kuan, S. F., Persson, A. (1992). Surfactant protein D: subcellular localization in nonciliated bronchiolar epithelial cells. *American Journal of Physiology.* **263(1):** 60-66.

Crouch, E., Persson, A., Chang, D., Heuser, J. (1994). Molecular Structure of Pulmonary Surfactant Protein (SP-D). *The Journal of Biological Chemistry.* **269(25):** 17311-17319

Cummings, R. D. & McEver, R. P. (1999). Chapter 31. C-type Lectins – Essentials of Glycobiology.

Danan, A., Mondange, M., Sarfati, S. R., Szabo, P. (1982). Synthesis and Behaviour under Acidic 2-Deoxy-α-DD-arabino-hexopyranose and 3-Deoxy-2-ketoaldonic Acids Bearing O-phosphono or O-glucosyl Substituents at Position B to the Carbonyl Function. *Journal of Chemical Society.* **Trans 1:** 1275-1282.

De Castro, C., Parrilli, M., Holst, O., Molinaro, A. (2010). Microbe-Associated Molecular Patterns in Innate Immunity: Extraction and Chemical Analysis of Gram-Negative Bacterial Lipopolysaccharides.

Dec, M. & Wernicki, A. (2006). Conglutinin, CL-43 and CL-46 - three bovine collectins. *Polish Journal of Veterinary Sciences.* **9(4):** 265-275. *Methods in Enzymology.* **480:** 89-115.

Di Tommaso, D., Ruiz-Agudo, E., de Leeuw, N. H., Putnis, A., Putnis, C. V. (2014). Modelling the effects of salt solutions on the hydration of calcium ions. *Physical Chemistry Chemical Physics.* **16:** 7772-7785.

Dias, R. P., Hora, G. C. A., Ramstedt, M., Soares, T. A. (2014a). Outer Membrane Remodelling: The Structural Dynamics and Electrostatics of Rough Lipopolysaccharide Chemotypes. *Journal of Chemical Theory and Computation.* **10:** 2488-2497.

Dias, R. P., Li, L., Soares, T. A., Alexov, E. (2014). Modelling the electrostatic Potential of Asymmetric Lipopolysaccharide Membranes: The MEMPOT Algorithm Implemented in DelPhi. *Journal of Chemical Theory and Computation.* **35:** 1418-1429.

Dommett, R. M., Klein, N., Turner, M. W. (2006). Mannose-binding lectin in innate immunity: past, present and future. *Tissue Antigens.* **68:** 193-209.

Doyle, I. R., Davidson, K. G., Barr, H. A., Nicholas, T. E., Payne, K., Pfitzner, J. (1998). Quantity and Structure of Surfactant Proteins Vary Among Patients with Alveolar Proteinosis. *American Journal of Respiratory Critical Care Medicine.* **157:** 658-664.

Drickamer, K. (1992). Engineering galactose-binding into a C-type mannose-binding protein. *Nature.* **360:** 183-186.

Drickamer, K. (1993). Evolution of (Ca^{2+})-dependent animal lectins. *Progress in Nucleic Acid Research and Molecular Biology*. **45:** 207-232.

Eckerle, I., Zimmermann, S., Kapaun, A., Junghanss, T. (2010). *Salmonella enterica* Serovar Virchow Bacteremia Presenting as Typhoid-Like Illness in an Immunocompetent Patient. *Journal of Clinical Microbiology.* **48(7):** 2643-2644.

Eckert, M. (2012). Then & Now. Max von Laue and the discovery of X-ray diffraction in 1912. *Annals of Physics.* **524(5):** 83-85.

Elhalwagi, B. M., Damodarasamy, M., McCormack, F. X. (1997). Alternate Amino Terminal Processing of Surfactant Protein A results in Cysteinyl Isoforms Required for Multimer Formation. *Biochemistry.* **36:** 7018-7025.

Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D.* **60:** 2126-2132.

Evans, G., Axford, D., Owen, R. L. (2011). The design of macromolecular crystallography diffraction experiments. *Acta Crystallographica Section D.* **67:** 261-270.

Evans, P. R. & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica Section D.* **69:** 1204-1214.

Evans, P. R. (2011a). An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D*. **67:** 282-292.

Feinberg, H., Torgersen, D., Drickamer, K., Weis, W. I. (2000). Mechanism of pH-dependent *N*-Acetylgalactosamine Binding by a Functional Mimic of the Hepatocyte Asialoglycoprotein Receptor. *The Journal of Biological Chemistry*. **275(45):** 35176-35184.

Ferguson, J. S., Voelker, D. R., McCormack, F. X., Schlesinger, L. S. (1999). Surfactant Protein D binds to *Mycobacterium tuberculosis* Bacilli and Lipoarabinomannan via Carbohydrate-Lectin Interactions Resulting in Reduced Phagocytosis of the Bacteria by Macrophages. *The Journal of Immunology.* **163(1):** 312-321.

Ferry, R. M. (1923). Studies In The Chemistry Of Hemoglobin. I- The Preparation of Hemoglobin. *Journal of Biological Chemistry.* **81:** 175-203.

Foadi, J., Aller, P., Alguel, Y., Cameron, A., Axford, D., Owen, R. L., Armour, W., Waterman, D. G., Iwata, S. Evans, G. (2013). Clustering procedures for the optimal selection data sets from multiple crystals in macromolecular crystallography. *Acta Crystallographica Section D.* **69:** 1617-1632.

Fox, K. L., Yildirim, H. H., Deadman, M. E., Schweda, E. K. H., Moxon, E. R., Hood, D. W. (2005). Novel lipopolysaccharide biosynthetic genes containing tetranucleotide

repeats in *Haemophilus influenzae,* identification of a gene for adding O-acetyl groups. *Molecular Microbiology.* **58(1):** 207-216.

Frecer, V., Ho, B., Ding, J. L. (2000). Molecular dynamics study on lipid A from *Escherichia coli:* insights into its mechanism of biological action. *Biochimica et Biophysica Acta.* **1466:** 87-104

Frecer, V., Ho, B., Ding, J. L. (2000a). Interpretation of biological activity data of bacterial endotoxins by simple molecular models of mechanism of action. *European Journal of Biochemistry.* **267:** 837-852.

Freudenberg, M. A., Tchaptchet, S., Keck, S., Fejer, G., Huber, M., Schutze, N., Beutler, B., Galanos, C. (2008). Lipopolysaccharides sensing an important factor in the innate immune response to Gram-negative bacterial infections: Benefits and hazards of LPS hypersensitivity. *Immunology.* **213:** 193-203.

Garcia-Verdugo, I., Wang, G., Floros, J., Casals, C. (2002). Structural Analysis and Lipid-Binding Properties of Recombinant Human Surfactant Protein A Derived from one or both genes. *Biochemistry*. **41:** 14041-14053.

Gardai, S. J., Xiao, Y-Q., Dickinson, M., Nick, J. A., Voelker, D. R., Greene, K. E., Henson, P. M. (2003). By Binding SIRPα or Calreticulin/CD91, Lung Collectins Act as Dual Function Surveillance Molecules to Supress or Enhance Inflammation. *Cell.* **115:** 13-23.

Garman, E. F. (2010). Radiation damage in macromolecular crystallography: what is it and why should we care? *Acta Crystallographica Section D.* **66:** 339-351.

Gingras, A. R., Girija, U. V., Keeble, A. H., Panchal, R., Mitchell, D. A., Moody, P. C. E., Wallis, R. (2011). Structural Basis of Mannan-Binding Lectin Recognition by Its Associated Serine Protease MASP-1: Implications for Complement Activation. *Cell-Structure.* **19:** 1635-1643.

Glasser, J. R. & Mallampalli, R. K. (2012). Surfactant and its role in the pathobiology of pulmonary infection. *Microbes and Infection*. **14:** 17-25

Goerke, J. (1998). Pulmonary surfactant: functions and molecular composition. *Biochimica et Biophysica Acta*. **1408:** 79-89

Goh, B. C., Rynkiewicz, M. J., Cafarella, T. R., White, M. R., Hartshorn, K. L., Allen, K., Crouch, E. C., Calin, O. et al. (2013). Molecular Mechanisms of Inhibition of Influenza by Surfactant Protein D Revealed by Large-Scale Molecular Dynamics Simulation. *Biochemistry.* **52:** 8527-8538.

Greene, K. E., King Jr, T. E., Kuroki, Y., Bucher-Bartelson, B., Hunninghake, G. W., Newman, L. S., Nagae, H., Mason, R. J. (2002). Serum surfactant proteins-A and –D as biomarkers in idiopathic pulmonary fibrosis. *European Respiratory Journal*. **19:** 439-446.

Gronow, S., & Brade, H. (2001). Lipopolysaccharides biosynthesis: which steps do bacteria need to survive. *Journal of Endotoxin Research*. **7(1):** 2-23

Gronow, S., Lindner, B., Brade, H., Muller-Loennies, S. (2009). Kdo-(2->8)-Kdo-(2->4)-Kdo but not Kdo-(2->4)-Kdo--(2->4)-Kdo is an acceptor for transfer of L-glycero-α-D-manno-heptose by *Escherichia coli* heptosyltransferase I (WaaC). *Innate Immunity.* **15(1):** 13-23.

Guo, C-J., Atochina-Vasserman, E. N., Abramova, E., Foley, J. P., Zaman, A., Crouch, E., Beers, M. F., Savani, R. C., et al. (2008). S-Nitrosylation of Surfactant Protein-D Controls Inflammatory Function. *PLOS Biology.* **6(11):** 2414-2423.

Gupta, G. & Surolia, A. (2007). Collectins: sentinels of innate immunity. *BioEssays.* **29:** 452-464.

Haagsman, H. P., Hogenkamp, A., van Eijk, M., Veldhuizen, E. J. A. (2008). Surfactant Collectins and Innate Immunity. *Neonatology.* **93:** 288-294.

Haagsman, H. P., Sargeant, T., Hauschka, P. V., Benson, B. J., Hawgood, S. (1990). *Biochemistry.* **29:** 8894-8900.

"*Haemophilus Influenzae* Type B (Hib) - NHS Choices" http://www.nhs.uk/conditions/Hib/Pages/Introduction.aspx, 2016. Web. 4 Apr. 2016.

Haczku, A. (2008). Protective role of the lung collectins surfactant protein A and surfactant protein D in airway inflammation. *Mechanisms of Allergic diseases.* **122(5):** 862-879.

Hakansson, K. & Reid, K. B. M. (2000). Collecting structure: A review. *Protein Science*. **9:** 1607-1617.

Hakansson, K., Lim, N, K., Hoppe, H., Reid, K, B, M. (1999). Crystal structure of the trimeric α-helical coiled-coil and the three lectin domains of the human lung surfactant protein D. *Structure.* **7(3):** 255-264.

Hambleton, G. & Davies, P. A. (1975). Bacterial Meningitis. Some aspects of diagnosis and treatment. *Archives of Disease in Childhood.* **50:** 674-684.

Han, Y., Han, X., Wang, S., Meng, Q., Zhang, Y., Ding, C., Yu, S. (2014). The waaL gene is involved in lipopolysaccharide synthesis and plays a role on the bacterial pathogenesis od avian pathogenic *Escherichia coli. Veterinary Microbiology.* **172:** 486-491.

Hansen, S. & Holmskov, U. (2002a). Lung Surfactant Protein (SP-D) and the Molecular Diverted Descendants: Conglutinin, CL-43 and CL-46. *Immunobiology*. **205:** 498-517.

Hansen, S., Holm, D., Moeller, V., Vitved, L., Bendixen, C., Reid, K. B. M., Skjoedt, K., Holmskov, U. (2002). CL-46, a Novel Collectin Highly Expressed in Bovine Thymus and Liver. *The Journal of Immunology*. **169:** 5726-5734.

Hansen, S., Holm, D., Moeller, V., Vitved, L., Bendixen, C., Skjoedt, K., Holmskov, U. (2003). Genomic and molecular characterisation of CL-43 and its proximal promoter. *Biochimica et Biophysica Acta*. **1625:** 1-10.

Hansen, S., Selman, L., Palaniyar, N., Ziegler, K., Brandt, J., Kliem, A., Jonasson, M., Skjoedt, M-O., et al. (2010). Collectin 11 (CL-11, CL-K1) Is a MASP-1/3- Associated Plasma Collectin With Microbial Activity. *The Journal of Immunology*. **185:** 6096-6104.

Hartshorn, K. L., Crouch, E., White, M. R., Colamussi, M. L., Kakkanatt, A., Tauber, B., Shepherd, V., Sastry, K. (1998). Pulmonary surfactant proteins A and D enhance neutrophil uptake of bacteria. *The American Journal of Physiology.* **274(6 Pt 1):** 958-969.

Hartshorn, K. L., Crouch, E., White, M. R., Eggleton, P., Tauber, A. I., Chang, D., Sastry, K. (1994). Evidence for a Protective role of Pulmonary Surfactant Protein D (SP-D) against *Influenza A* viruses. *Journal of Clinical Investigation.* **94:** 311-319.

Hartshorn, K. L., Holmskov, U., Hansen, S., Zhang, P., Meschi, J., Mogues, T., White, M. R., Crouch, E. C. (2002). Distinctive anti-influenza properties of recombinant collectin 43. *Biochemistry Journal*. **366**: 87-96.

Hartshorn, K. L., Sastry, K. N., Chang, D., White, M. R., Crouch, E. C. (2000). Enhanced anti-influenza activity of surfactant protein D and serum conglutinin fusion protein. *American Journal of Physiology Lung Cellular and Molecular Physiology*. **278:** 90-98.

Hartshorn, K. L., White, M. R., Shepherd, V., Reid, K., Jensenius, J. C., Crouch, E. C. (1997). Mechanisms of anti-influenza activity of surfactant proteins A and D:

comparison with serum collectins. *The American Journal of Physiology.* **273:** 1156-1166.

Haurum, J. S., Thiel, S., Haagsman, H. P., Laursen, S. B., Larsen, B., Jensenius, J. C. (1993). Studies on the carbohydrate-binding characteristics of human pulmonary surfactant-associated protein A and comparison with two other collectins: mannan-binding protein and conglutinin. *Biochemistry Journal*. **293:** 873-878.

Head, J. F., Mealy, T. R., McCormack, F. X., Seaton, B. A. (2003). Crystal Structure of Trimeric Carbohydrate Recognition and Neck Domains of Surfactant Protein A. *The Journal of Biological Chemistry*. **278(44):** 43254-43260.

Heerklotz, H., Binder, H., Lantzsch, G., Klose, G. (1997). Lipid/Detergent Interaction Thermodynamics as a Function of Molecular Shape. *The Journal of Physical Chemistry.* **101:** 639-645.

Heinrichs, D. E., Yethon, J. A., Whitefield, C. (1998). Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica. Molecular Microbiology.* **30(2):** 221-232.

Henriksen, M. L., Brandt, J., Andrieu, J-P., Nielsen, C., Jensen, P. H., Holmskov, U., Jorgensen, T. J. D., Palarasah, Y., et al. (2013). Heteromeric Complexes of Native Collectin Kidney 1 and Collectin Liver 1 Are Found in the Circulation with MASPs and Activate the Complement System. *The Journal of Immunology*. **191:** 6117-6127.

Hickling, T. P., Bright, H., Wing, K., Gower, D., Martin, S. L., Sim, R. B., Malhotra, R. (1999). A recombinant trimeric surfactant protein D carbohydrate recognition domain inhibits respiratory syncytial virus infection *in vitro* and *in vivo*. *European Journal of Immunology.* **29:** 3478-3484.

Hogenkamp, A., van Eijk, M., van Dijk, A., van Asten, A. J. A. M., Veldhuizen, E. J. A., Haagsman, H. P. (2006). Characterization and expression of newly identified chicken collectin. *Molecular Immunology*. **43:** 1604-1616.

Holmskov, U., Laursen, S. B., Malhotra, R., Wiedemann, H., Timpl, R., Stuart, G. R., Tornoe, I., Madesen, P. S., et al. (1995). Comparative study of the structural and functional properties of a bovine plasma C-type lectin, collectin-43, with other collectins. *Biochemistry Journal*. **305:** 889-896.

Holmskov, U., Lawson, P., Teisner, B., Tornoe, I., Willis, A. C., Morgan, C., Koch, C., Reid, K. B. M. (1997). Isolation and Characterisation of a New Member of the Scavenger Receptor Superfamily, Glycoprotein-340 (gp-340), as a lung Surfactant Protein-D binding Molecule. *The Journal of Biological Chemistry.* **272(21):** 13743-13749.

Holmskov, U., Teisner, B., Willis, A. C., Reid, K. B.M., Jensenius, J. C. (1993). Purification and characterisation of a Bovine Serum Lectin (CL-43) with Structural Homology to Conglutinin and SP-D and Carbohydrate Specificity Similar to Mannan-Binding Protein. *The Journal of Biological Chemistry.* **268(14):** 10120-10125.

Holmskov, U., Thiel, S., Jensenius, J. C. (2003). Collectins and Ficolins: Humoral Lectins of the Innate Immune Defense. *Annual Reviews of Immunology.* **21:** 547-578.

Holst, O. & Brade, H. (1991). Structural studies of the core region of the lipopolysaccharide from *Salmonella minnesota* strain R7 (rough mutant chemotype Rd_1). *Carbohydrate Research.* **219:** 247-251.

Holst, O. (2007). The structures of core regions from enterobacterial lipopolysaccharides- an update. *Federation of European Microbial Societies*. **271:** 3-11

Honda, Y., Takahashi, H., Kuroki, Y., Akino, T., Abe, S. (1996). Decreased Contents of Surfactant Proteins A and D in BAL Fluids of Healthy Smokers. *Chest.* **109(4):** 1006-1009.

Hood, D. W., Deadman, M. E., Allen, T., Masoud, H., Martin, A., Brisson, J. R., Fleischmann, R., Venter, J. C., Richards, J. C., Moxon, E. R. (1996a). Use of the complete genome sequence information of *Haemophilus influenzae* strain Rd to investigate lipopolysaccharides biosynthesis. *Molecular Microbiology*. **22(5):** 951-965.

Hood, D. W., Deadman, M. E., Jennings, M. P., Bisercic, M., Fleischmann, R. D., Venter, J. C., Moxon, E. R. (1996). DNA repeats identify novel virulence genes in *Haemophilus influenzae. Proceeding of the National Academy of Sciences.* **93:** 11121-11125.

Hood, D. W., Makepeace, K., Deadman, M. E., Rest, R. F., Thibault, P., Martin, A., Richards, J. C., Moxon, E. R. (1999). Sialic acid in the lipopolysaccharide of *Haemophilus influenzae:* strain distribution, influence on serum resistance and structural characterisation. *Molecular Microbiology.* **33(4):** 679-692.

Hoppe, H-J., Barlow, P. N., Reid, K. B. M. (1994). A parallel three stranded α-helical bundle at the nucleation site of collagen triple-helix formation. *FEBS LETTERS.* **344:** 191-195.

Hsieh, C-H. & Wu, W-G. (1996). Structure and Dynamics of Primary Hydration Shell of Phosphatidylcholine Bilayers at Subzero Temperatures. *Biophysical Journal.* **71:** 3278-3287.

Huber, M., Kalis, C., Keck, S., Jiang, Z., Georgel, P., Du, X., Shamel, L., Sovath, S., Mudd, S. et al. (2006). R-form LPS, the master key to the activation of TLR4/MD-2 positive cells. *European Journal of Immunology.* **36:** 701-711.

Ikegami, M., Carter, K., Bishop, K., Yadav, A., Masterjohn, E., Brondyk, W., Scheule, R. K., Whitsett, J. A. (2006). Intratracheal Recombinant Surfactant Protein D Prevents Endotoxin Shock in the Newborn Preterm Lamb. *American Journal of Respiratory Care Medicine.* **173:** 1342-1347.

Iwasaki, A. & Medzhitov, R. (2015). Control of adaptive immunity by the innate immune system. *Nature Immunology*. **16(4):** 343-353.

Jang, H., Kim, H-S., Moon, S-C., Lee, Y-R., Yu, K-Y., Lee, B-K., Youn, H. Z., Jeong, Y-J., et al. (2009). Effects of protein concentration and detergent on endototoxin reduction by ultrafiltration. *BMB reports.* **42(7):** 462-466.

Jiang, H. & Chess, L. (2009). How the Immune System Achieves Self-Nonself Discrimination During Adaptive Immunity. *Advances In Immunology.* **102**: 95-133.

Jousimies, H. & Makela, P. H. (1974). Genetic Analysis of *Salmonella minnesota* R mutants with Defects in the Biosynthesis of the Lipopolysaccharides Core. *Journal of Bacteriology*. **119(3):** 753-759.

Juers, D. H. & Matthews, B. W. (2001). Reversible Lattice Repacking Illustrates the Temperature Dependence of Macromolecular Interactions. *Journal of Molecular Biology.* **311:** 851-862.

Kabanov, D. S. & Prokhorenko, I. R. (2011). Relationships between Physiochemical Characteristics and Biological Activity of Lipopolysaccharides. *Biochemistry (Moscow) Supplement Series A*: *Membrane and Cell Biology.* **5(4):** 293-309.

Kato, N. (1993). Crystallization and Electron Microscopy of Bacterial Lipopolysaccharides (LPS). *Micron*. **24(1):** 91-114.

Katz. A. K., Glusker, J. P., Beebe, S. A., Bock, C. W. (1996). Calcium Ion coordination: A comparison with That of Beryllium, Magnesium and Zinc. *Journal of the American Chemical Society.* **118:** 5752-5763.

Kendrew, J.C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. (1958). A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. *Nature.* **181:** 662-666.

Kerr, M. H. & Paton, J. Y. (1999). Surfactant Protein Levels in Severe Respiratory Syncytial Virus Infection. *American Journal of Respiratory Care Medicine.* **159:** 1115-1118.

Keshi, H., Sakamoto, T., Kawai, T., Ohtani, K., Katoh, T., Jang, S-J., Motomura, W., Yoshizaki, T., et al. (2006). Identification and characterization of a novel human collectin CL-K1. Microbiology and Immunology. **50(12):** 1001-1013.

King, P. (2012). *Haemophilus influenzae* and the lung (*Haemophilus* and the lung). *Clinical and translational Medicine.* **1:** 10.

Kirschner, K. N & Woods, R. J. (2001). Solvent interactions determine carbohydrate conformation. *PNAS.* **98(19):** 10541-10545.

Kishore, U., Greenhough, T., Walters, P., Shrive, A., Ghai, R., Kamran M., Bernal, A., Reid, K. et al. (2006). Surfactant Proteins SP-A and SP-D structure, function and receptors. *Molecular Immunol*ogy. **43(9):** 1293-315.

Kishore, U., Wang, J-Y., Hoppe, H-J., Reid, K.B.M. (1996). The α-helical neck region of human lung surfactant protein D is essential for the binding of the carbohydrate recognition domains to lipopolysaccharides and phospholipids. *Biochemistry Journal*. **318:** 505-511.

Klein, G., Muller-Loennies, S., Lindner, B., Kobylak, N., Brade, H., Raina, S. (2013). Molecular and structural basis of inner core lipopolysaccharides alterations in *Escherichia coli:* incorporation of glucuronic acid and phosphoethanolamine in the heptose region. *Journal of Biological Chemistry*. **288(12):** 8111-8127.

Kolatkar, A. R. & Weis, W. I. (1996). Structural Basis of Galactose Recognition by C-type Animal Lectins. *The Journal of Biological Chemistry.* **27(12):** 6679-6685.

Korfhagen, T. R., Sheftelyevich, V., Burhans, M. S., Bruno, M. D., Ross, G. F., Wert, S. E., Stahlman, M. T., Jobe, A. H. et al. (1998). Surfactant Protein-D regulates Surfactant Phospholipid Homeostasis *in Vivo*. *The Journal of Biological Chemistry.* **273(43):** 28438-28443.

Kovacs, H., O'Donoghue, S. I., Hoppe, H-J., Comfort, D., Reid, K. B. M., Campbell, I. D., Nilges, M. (2002). Solution structure of the coiled-coil trimerization domain from lung surfactant protein D. *Journal of Biomolecular NMR.* **24:** 89-102.

Kramer, R. Z., Bella, J., Mayville, P., Brodsky, B., Berman, H. M. (1999). Sequence dependent conformational variations of collagen triple-helical structure. *Nature Structural Biology.* **6(5):** 454-457.

Kuan, S-F., Rust, K., Crouch, E. (1992). Interactions of Surfactant Protein D with Bacterial Lipopolysaccharides. *Journal of Clinical Investigation.* **90:** 97-106.

Kucerka, N., Papp-Szabo, E., Nieh, M-P., Harroun, T. A., Schooling, S. R., Pencer, J., Nicholson, E. A., Beveridge, T. J., et al. (2008). Effects of Cations on the Structure of Bilayers Formed by Lipopolysaccharides Isolated from *Pseudomonas aeruginosa* PAO1. *Journal of Physical Chemistry.* **112:** 8057-8062.

Kudo, K., Sano, H., Takahashi, H., Kuronuma, K., Yokota, S-I., Fujii, N., Shimada, K-I., Yano, I., Kumazawa, Y. (2004). Pulmonary Collectins Enhance Phagocytosis of

Mycobacterium avium through Increased Activity of Mannose Receptor. *The Journal of Immunology.* **172:** 7592-7602.

Lahti, M., Lofgren, J., Marttila, R., Renko, M., Klaavuniemi, T., Haataja, R., Ramet, M., Hallman, M. (2002). Surfactant Protein D Gene Polymorphism Associated with Severe Respiratory Syncytial Virus Infection. *Paediatric Research.* **51(6):** 696-699.

Lambrecht, B. N. & Hammad, H. (2012). The airway epithelium in asthma. *Nature Medicine.* **18(5):** 684-692.

Lazaridis, T. & Paulaitis, M. E. (1992). Entropy of Hydrophobic Hydration: A New statistical Mechanical Formulation. *Journal of Physical Chemistry.* **96:** 3847-3855.

le Marie, M., Champeil, P., Moller, J. V. (2000). Interaction of membrane proteins and lipids with solubilizing detergents. *Biochimica et Biophysica Acta.* **1508:** 86-111.

Leith-Larsen, R., Holmskov, U., Hojrup, P. (1999). Structural characterisation of human and bovine lung surfactant protein D. *Biochemistry Journal.* **343:** 645-652.

Leslie, A. G. W. (2006). The integration of macromolecular diffraction data. *Acta Crystallographica Section D.* **62:** 48-57.

LeVine, A. M., Whitsett, J. A., Gwozdz, J. A., Richardson, T. R., Fisher, J. H., Burhans, M. S., Korfhagen, T. R. (2000). Distinct Effects of Surfactant Protein A or D Deficiency During Bacterial Infection on the lung. *The Journal of Immunology.* **165:** 3934-3940.

LeVine, A. M., Whitsett, J. A., Hartshorn, K. L., Crouch, E. C., Korfhagen, T. R. (2001). Surfactant Protein D Enhances Clearance of Influenza A virus from the Lung in vivo. *The Journal of Immunology.* **167:** 5868-5873.

Li, J., Bauer, S., Mansson, M., Moxon, E. Richards, J., Schweda, E. (2001). Glycine is a common substituent of the inner core in *Haemophilus influenzae* lipopolysaccharide. *Glycobiology*. **11(12):** 1009-1015.

Li, P., Wohland, T., Ho, B., Ding, J. L. (2004). Perturbation of Lipopolysaccharide (LPS) Micelles by Sushi 3 (S3) Antimicrobial Peptide: THE IMPORTANCE OF AN INTERMOLECULAR DISULFIDE BOND IN S3 DIMER FOR BINDIN, DISRUPTION, AND NEUTRALIZATION OF LPS. *The Journal of Biological Chemistry.* **279:** 50150-50156.

Lichtenberg, D., Ahyayauch, H., Goni, F. M. (2013). The Mechanisms of Detergent Solubilisation of Lipid Bilayers. *Biophysical Journal.* **105:** 289-299.

Lim, B-L., Wang, J-Y., Holmskov, U., Hoppe, H-J., Reid, K. B. M. (1994). Expression of the Carbohydrate Recognition Domain of Lung Surfactant Protein D and Demonstration of its Binding to Lipopolysaccharides of Gram-Negative Bacteria. *Biochemical and Biophysical Research Communications.* **202(3):** 1674-1680.

Lim, B-S. & Holmskov, U. (1996). Expression of the Carbohydrate Recognition Domain of Bovine Conglutinin and Demonstration of Its Binding to iC3b and Yeast Mannan. *Biochemical and Biophysical Research Communications*. **218:** 260-266.

Liu, B., Knirel, Y. A., Feng, L., Perepelov, A. V., Senchenkova, S. N., Reeves, P. R., Wang, L. (2013). Structural diversity in *Salmonella O* antigens and its genetic basis. *FEMS Microbial Reviews.* **38:** 56-89.

Luderitz, O., Galanos, C., Risse, H. J., Ruschamann, E., Schelecht, S., Schmidt, G., Schulte-Holthausen, H., Wheat, R., Westphal, O. (1967). Structural relationships of Salmonella O and R antigens. *Annals New York Academy Sciences*. 349-374.

Luderitz, O., Schulte-Holthausen, H. J. R. H., Strominger, J. L., Sutherland, I. W., Westphal, O. (1965). Biochemical studies of the Smooth-Rough Mutation in *Salmonella minnesota. Journal of Bacteriology.* **89(2):** 343-354.

Madan, T., Eggleton, P., Kishore, U., Strong, P., Aggrawal, S. S., Sarma, P. U., Reid, K. B. M. (1997). Binding of Pulmonary Surfactant Proteins A and D to *Aspergillus fumigatus* Conidia Enhances Phagocytosis and Killing by Human Neutrophils and Alveolar Macrophages. *Infection and Immunity.* **65(8):** 3171-3179.

Madan, T., Kishore, U., Singh, M., Strong, P., Hussain, E. M., Reid, K. B. M., Sarma, P. U. (2001). Protective role of Lung Surfactant Protein D in a Murine Model of Invasive Pulmonary *Aspergillosis*. *Infection and Immunity.* **69(4):** 2728-2731.

Mahajan, L., Madan, T., Kamal, N., Sing, V. K., Sim, R. B., Telang, S. D., Ramchand, C. N., Waters, P., et al. (2008). Recombinant surfactant-D selectively increases apoptosis in eosinophils of allergic asthmatics and enhances uptake of apoptotic eosinophils by macrophages. *International Immunology.* **20(8):** 993-1007.

Malkin, A. J., Kuznetsov, Y. G., Land, T. A., DeYoreo, J. J., McPherson, A. (1995). Mechanisms of growth for protein and virus crystals. *Nature.* **2(11):** 956-959.

Mannerstedt, K., Segerstedt, E., Olsson, J., Oscarson, S. (2008). Synthesis of a common tetrasaccharide motif of *Haemophilus influenzae* LPS inner core structures. *Organic & Biomolecular Chemistry*. **6:** 1087-1091.

Mansfield, L. P. & Forsythe, S. J. (2001) Demonstration of the Rb_1 lipopolysaccharides core structure in Salmonella strains with the monoclonal antibody M105. *Journal of Medical Microbiology.* **50:** 339-344

Markart, P., Ruppert, C., Wygrecka, M., Colaris, T., Dahal, B., Walmrath, D., Harbach, H., Whilhelm, J., et al. (2007). Patients with ARDS show improvement but

not normalisation of alveolar surface activity with surfactant treatment: putative role of neutral lipids. *Thorax.* **62:** 588-594.

Masoud, H., Altman, E., Richards, J. C., Lam, J. S. (1994). General Strategy for Structural Analysis of the Oligosaccharide Region of Lipooligosaccharides. Structure of the Oligosaccharide Component of *Pseudomonas aeruginosa* IATS Serotype 06 Mutant R5 Rough-Type Lipopolysaccharide. *Biochemistry*. **33:** 10568-10578.

Masoud, H., Martin, A., Thibault, P., Moxon, E., Richards, J. (2003). Structure of Extended lipopolysaccharide glycoforms containing two Globotriose units in *Haemophilus influenzae* Serotype b Strain RM7004. *Biochemistry*. **42:** 4463-4475.

Masoud, H., Moxon, E. R., Martin, A., Krajcarski, D., Richards, J. C. (1997). Structure of the Variable and Conserved Lipopolysaccharides Oligosaccharide Epitopes Expressed by *Haemophilus influenza*e Serotype b Strain Eagan. *Biochemistry.* **36:** 2091-2103.

Matsushita, M., Ezekowitz, R. A. B., Fujita, T. (1995). The Gly-54 - Asp allelic form of human mannose-binding protein (MBP) fails to bind MBP-associated serine protease. *Biochemistry Journal.* **311:** 1021-1023.

Matthews, B. W. (1968). Solvent Content of Protein Crystals. *Journal of Molecular Biology.* **33:** 491-497.

Matthews, B. W. (1976). X-ray Crystallographic Studies of Proteins. Annual Reviews of *Physical Chemistry.* **27:** 493-523.

McCallum, K. L., Schoenhals, G., Laakso, D., Clarke, B., Whitfield, C. (1989). A High-Molecular-Weight Fraction of Smooth Lipopolysaccharide in *Klebsiella* Serotype O1:K20 Contains a Unique O-antigen Epitope and Determines Resistance to Nonspecific Serum Killing. *Infection and Immunity.* **57(12):** 3816-3822.

McCormack, F. X. & Whitsett, J. A. (2002). The pulmonary collectins, SP-A and SP-D, orchestrate innate immunity in the lung. *Journal of Clinical Investigation.* **109:** 707-712.

McCormack, F. X., King, T. E., Voelker, D. R., Robinson, P. C., Mason, R. J. (1991). Idiopathic Pulmonary Fibrosis: Abnormalities in the Bronchoalveolar Lavage Content of Surfactant Protein A. *American Review of Respiratory Disease*. **144(1):** 160-166.

McPherson, A. & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F.* **70:** 2-20.

McPherson, A., Malkin, A. J., Kuznetsov, Y. G. (2000). Atomic Force Microscopy In The Study of Macromolecular Crystal Growth. *Annual Reviews of Biophysical and Biomolecular Structure.* **29:** 361-410.

Meschi, J., Crouch, E. C., Skolnik, P., Yahya, K., Holmskov, U., Leth-Larsenm, R., Tornoe, I. Et al. (2005). Surfactant protein D binds to human immunodeficiency virus (HIV) envelope protein gp120 and inhibits HIV replication. *Journal of General Virology.* **86:** 3097-3107.

Mikerov, A. N., Umstead, T. M., Huang, W., Liu, W., Phelps, D. S., Floros, J. (2004). SP-A1 and SP-A2 variants differentially enhance association of *Pseudomonas aeruginosa* with rat alveolar macrophages. *American Journal of Physiology Lung Cellular and Molecular Physiology*. **288:** 150-158.

Mikerov, A. N., Wang, G., Umstead, T. M., Zacharatos, M., Thomas, N. J., Phelps, D. S., Floros, J. (2007). Surfactant Proteins A2 (SP-A2) Variants Expressed in CHO Cells Stimulate Phagocytosis of *Pseudomonas aeruginosa* More than Do SP-A1 Variants. *Infection and Immunity*. **75(3):** 1403.

Miyamura, K., Leigh, L. E. A., Lu, J., Hopkin, J., Bernial, A. L., Reid, K. B. M. (1994). Surfactant protein D binding to alveolar macrophages. *Biochemistry Journal.* **300:** 237-242.

Mogensen, T. H. (2009). Pathogen Recognition and Inflammatory Signalling in Innate Immune Defences. *Clinical Microbiology Reviews.* **22(2):** 240-273.

Moran, A. P., Khamri, W., Walker, M. M., Thursz, M. R. (2005). Role of surfactant protein D (SP-D) in innate immunity in the gastric mucosa: evidence of interaction with *Helicobacter pylori* lipopolysaccharide. *Journal of Endotoxin Research.* **11(6):** 357-362.

Morrison, D.C. & Leive, L. (1975). Fractions of lipopolysaccharides from *Escherichia coli* O111:B4 prepared by two extraction procedures. *The journal of Biological Chemistry.* **250(8):** 2911-2919.

Motomura, W., Yoshizaki, T., Ohtani, K., Okumura, T., Fukuda, M., Fukuzawa, J., Mori, K., Jang, S-J., et al. (2008). Immunolocalization of a Novel Collectin CL-K1 in Murine Tissues. *Journal of Histochemistry*. **56(3):** 243-252.

Mueller, M., Wang, M., Schulze-Briese, C. (2012). Optional fine φ-slicing for single-photon-counting pixel detectors. *Acta Crystallographica Section D*. **68:** 42-56.

Muhlradt, P., Risse, H. J., Luderitz, O., Westphal, O. (1968). Biochemical Studies on Lipopolysaccharides of *Salmonella minnesota. European Journal of Biochemistry.* **4:** 139-145.

Muller-Loennies, S., Lindner, B., Brade, H. (2002). Structural analysis of deacylated lipopolysaccharide of *Escherichia coli* strains 2513 (R4 core type) and F653 (R3 core type). *European Journal of Biochemistry*. **269:** 5982-5991.

Muller-Loennies, S., Lindner, B., Brade, H. (2003). Structural Analysis of Oligosaccharides from Lipopolysaccharides (LPS) of *Escherichia coli* K12 Strain W3100 Reveals a Link between Inner and Outer Core LPS Biosynthesis. *The Journal of Biological Chemistry.* **278(36):** 34090-34101.

Murzyn, K. & Pasenkiewicz-Gierula, M. (2015) Structural Properties of the Water/Membrane Interface of a Bilayer Built of the *E. coli* Lipid A. *Journal of Physical Chemistry B.* **119:** 5846-5856.

Nakamura, K., Ohya, W., Funakoshi, H., Sakaguchi, G., Kato, A., Takeda, M., Kudo, T., Nakamura, T. (2006). Possible Role of Scavenger Receptor SRCL in the Clearance of Amyloid- β in Alzheimer Disease. *Journal of Neuroscience Research*. **84:** 874-890.

Nascimento, A. Jr., Pontes, F. J. S., Lins, R. D., Soares, T. A. (2014). Hydration, ionic valence and cross-linking propensities of cations determine the stability of lipopolysaccharides (LPS) membranes. *Chemical Communications.* **50:** 231-233.

Nave, C. & Garman, E. F. (2005). Towards an understanding of radiation damage in cryocooled macromolecular crystals. *Journal of Synchrotron Radiation.* **12:** 257-260.

Newton, K. & Dixit, V. M. (2012). Signalling in Innate Immunity and Inflammation. *Cold Spring Harbor Perspective in Medicine*. **4:** 1-20.

Ng, K. K. -S., Drickamer, K., Weis, W. I. (1996). Structural analysis of Monosaccharide Recognition by Rat Liver Mannose-Binding Protein. *Journal of Biological Chemistry.* **271:** 663-674.

Ng, K. K.-S. & Weis, W. I. (1997). Structure of a Selectin-like Mutant of Mannose-Binding Protein Complexed with Sialylated and Sulfated Lewis^{x} Oligosaccharides. *Biochemistry.* **36(5):** 979-988.

Ng, K. K.-S., Kolatkar, A. R., Park-Snyder, S., Feinberg, H., Clark, D. A., Drickamer, K., Weis, W. I. (2002). Orientation of Bound Ligands in Mannose-binding Proteins. *The Journal of Biological Chemistry*. **277(18):** 16088-16095.

Ng, K. K.-S., Park-Snyder, S., Weis, W. I. (1998). Ca^{2+} -Dependent Structural Changes in C-type Mannose-Binding Proteins. *Biochemistry.* **37:** 17965-17976.

Nishikiori, H., Chiba, H., Ariki, S., Kuronuma, K., Otsuka, M., Shiratori, M., Ikeda, K., Watanabe, A. (2014). Distinct compartmentalization of SP-A and SP-D in the vasculature and lungs of patients with idiopathic pulmonary fibrosis. *BMC Pulmonary Medicine.* **14(196):** 1-10.

Nnalue, N. A. (1999). All Accessible Epitopes in the *Salmonella* Lipopolysaccharides Core Are Associated with Branch Residues. *Infection And Immunity.* **67(2):** 998-1003

Nordkamp, M. J. M. O., van Eijk, M., Urbanus, R. T., Bont, L., Haagsman, H. P., Meyaard, L. (2014). Leukocyte-associated Ig-like receptor-1 is a novel inhibitory receptor for surfactant protein D. *Journal of Leukocyte Biology.* **96:** 105-111

O'Reilly, M. A., Nogee, L., Whitsett, J. A. (1988). Requirement of the collagenous domain for carbohydrate processing and secretion of a surfactant protein, SP-A. *Biochimica et Biophysica Acta*. **969:** 176-184.

O'Riordan, D. M., Standing, J. E., Kwon, K-Y., Chang, D., Crouch, E. C., Limper, A. H. (1995). Surfactant Protein D Interacts with *Pneumocystis carinii and* Mediates Organism Adherence to Alveolar Macrophages. *The Journal of Clinical Investigation.* **95:** 2699-2710.

Oberley, R. E. & Snyder, J. M. (2003). Recombinant human SP-A1 and SP-A2 proteins have different carbohydrate-binding characteristics. *American Journal of Physiology- Lung Cellular and Molecular Physiology.* **284:** 871-881.

Oberley, R. E., Ault, K. A., Neff, T. L., Khubchandani, K. R., Crouch, E. C., Snyder, J. M. (2004a). Surfactant proteins A and D enhance the phagocytosis of *Chlamydia* into THP-1 cells. *American Journal of Physiology.* **287:** 296-306.

Oberley, R. E., Goss, K. L., Ault, K. A., Crouch, E. C., Snyder, J. M. (2004). Surfactant protein D is present in the human female reproductive tract and inhibits *Chlamydia trachomatis. Molecular Human Reproduction.* **10(12):** 861-870.

Ogasawara, Y., Kuroki, Y., Akino, T. (1992). Pulmonary Surfactant Protein D Specifically Binds to Phosphatidylinositol. *The Journal of Biological Chemistry.* **267(29):** 21244-21249

Ohtani, K., Suzuki, Y., Wakamiya, N. (2012). Biological Functions of the Novel Collectins CL-L1, CL-K1 and CL-P1. *Journal of Biomedicine and Biotechnology*. **Volume 2012**: Article ID 493945.

Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Keshi, H., Sakai, Y., Fukuoh, A., et al. (2001). The Membrane-type Collectin CL-P1 Is a Scavenger Receptor on Vascular Endothelial Cells. *The Journal of Biological Chemistry*. **276(47):** 44222-44228.

Ohtani, K., Suzuki, Y., Eda, S., Kawai, T., Kase, T., Yamazaki, H., Shimada, T., Keshi, H., et al. (1999). Molecular cloning of a Novel Human Collectin from Liver (CL-L1). *The Journal of Biological Chemistry.* **274(19):** 13681-13689.

Olsthoorn, M. M. A., Petersen, B. O., Schlecht, S., Haverkamp, J., Bock, K., Thomas-Oates, J. E., Holst, O. (1998). Identification of a Novel Core Type in Salmonella Lipopolysaccharide. *The Journal of Biological Chemistry*. **273(7):** 3817-3829.

Pace, C. N., Trevino, S., Prabahakaran, E., Scholtz, J. M. (2004). Protein Structure, stability and solubility in water and other solvents. *Philosophical Transactions of the Royal Society of B.* **359:** 1225-1235.

Pagowska-Klimek, I. & Cedzynski, M. (2014). Mannan-Binding Lectin in Cardiovascular Disease. *Biomedical Research International*. *Article ID- 616817- 13 pages.*

Palaniyar, N., Clark, H., Nadesalingam, J., Hawgood, S., Reid, K. B. M. (2003). Surfactant Protein D Binds Genomic DNA and Apoptotic cells and Enhances Their clearance, in vivo. *Annual New York Academy of Sciences.* **1010:** 471-475.

Palaniyar, N., Nadesalingam, J., Clark, H., Shih, M. J., Dodds, A. W., Reid, K. B. M. (2004). Nucleic Acid is a Novel Ligand for Innate, Immune Pattern Recognition Collectins Surfactant Proteins A and D and Mannose Binding Protein. *The Journal of Biological Chemistry.* **279(31):** 32728-32736.

Parikh, S. J. & Chorover, J. (2007). Infrared spectroscopy studies of cation effects on lipopolysaccharides in aqueous solution. *Colloids and Surfaces B: Biointerfaces.* **55:** 241-250.

Park, B. S., Song, D. H., Kim, H. N., Choi, B-S., Lee, H., Lee, J-O. (2009). The structural basis of lipopolysaccharides recognition by the TLR4-MD-2 complex. *Nature*. **458:** 1191-1196.

Parker, C. T., Kloser, A. W., Schnaitman, C. A., Stein, M. A., Gottesman, S., Gibson, B. W. (1992). Role of the *rfaG* and *rfaP* genes in Determining the Lipopolysaccharide Core Structure and Cell Surface Properties of *Escherichia coli* K-12. *Journal of Bacteriology.* **174(8):** 2525-2538.

Parker, C. T., Pradel, E., Schnaitman, C. A. (1992a). Identification and Sequences of the Lipopolysaccharide Core Biosynthetic Gene *rfaQ, rfaP* and *rfaG* of *Escherichia coli* K-12. *Journal of Bacteriology.* **174 (3):** 930-934.

Pattanajitvilai, S., Kuroki, Y., Tsenezawa, W., McCormack, F. X., Voelker, D. R. (1998). Mutational Analysis of Arg197 of Rat Surfactant Protein A: HIS197 CREATES

SPECIFIC LIPID UPTAKE DEFECTS. *The Journal of Biological Chemistry*. **273:** 5702-5707.

"Peptidoglycans". Sigma-Aldrich. http://www.sigmaaldrich.com/technical-documents/articles/biology/glycobiology/peptidoglycans.html., 2016. Web. 4 Mar. 2016.

Persson, A., Chang, D., Crouch, E. (1990). Surfactant Protein D is a Divalent Cation-dependent Carbohydrate-binding protein. *The Journal of Biological Chemistry.* **265(10):** 5755-5760.

Persson, A., Rust, K., Chang, D., Moxley, M., Longmore, W., Crouch, E. (1988). CP4: A pneumocyte-derived Collagenous Surfactant-Associated Protein. Evidence for Heterogeneity of Collagenous Surfactant Proteins. *Biochemistry.* **27:** 8576-8564.

Peterson, A. A. & McGroarty, E. J. (1985). High-Molecular-Weight Components in Lipopolysaccharides of *Salmonella typhimurium, Salmonella minnesota* and *Escherichia coli. Journal of Bacteriology.* **162(2):** 738-745.

Phillips, N. J., Apicella, M. A., Griffiss, J. M., Gibson, B. W. (1992). Structural Characterization of the Cell Surface Lipoooligosaccharides from Nontypable Strain of *Haemophilus influenzae*. *Biochemistry.* **31:** 4515-4526.

Picardi, M. V., Cruz, A., Orellana, G., Perez-Gil, J. (2011). Phospholipid packing and hydration in pulmonary surfactant membranes and films as sensed by LAURDAN. *Biochimical et Biophysica Acta*. **1808(3):** 696-705.

Pikaar, J. C., Voorhout, W. F., van Golde, L. M. G., Verhoef, J., Van Strijp, J. A. G., van Iwaardem, J. F. (1995). Opsonic Activities of Surfactant Protein A and D in Phagocytosis of Gram-Negative by Alveolar Macrophages. *The Journal of Infectious Diseases*. **172:** 481-489.

Postle, A. D., Mander, A., Reid, K. B. M., Wang, J-Y., Wright, S. M., Moustaki, M., Warner, J. O. (1999). Deficient Hydrophilic Lung Surfactant Proteins A and D with Normal Surfactant Phospholipid Molecular Species in Cystic Fibrosis. *American Journal of Respiratory Cell and Molecular Biology.* **20:** 90-98.

Privalov, P. L. & Gill, S. J. (1989). The hydrophobic effect: a reappraisal. *Pure & Applied Chemistry.* **61(6):** 1097-1104.

Que, N. L. S., Lin, S., Cotter, R. J., Raetz, C. R. H. (2000). Purification and Mass Spectrometry of Six Lipid A Species from the Bacterial Endosymbiont *Rhizobium etli:* DEMONSTRATION OF A CONSERVED DISTAL UNIT AND AVARIABLE PROXIMAL PORTION. *Journal of Biological Chemistry.* **275:** 28006-28016.

Raetz, C. & Whitfield, C. (2002). Lipopolysaccharides Endotoxins. *Annual Reviews Biochemistry*. **71:** 635-700.

Raetz, C. R. H., Ulevitch, R. J., Wright, S. D., Sibley, C. H., Ding, A., Nathan, C. F. (1991). Gram-negative endotoxin: an extraordinary lipid with profound effects on eukaryotic signal transduction. *The FASEB Journal.* **5:** 2652-2660.

Raschke, T. M. & Levitt, M. (2005). Nonpolar solutes enhance water structure within hydration shells while reducing interactions between them. *PNAS.* **102(19):** 6777-6782.

Read, R. J., Adams, P. D., Arendall, W. B., Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., et al. (2011). A New Generation of Crystallographic Validation Tools for the Protein Data Bank. *Structure - Cell.* **19:** 1395-1412.

Reading, P. C., Holmskov, U., Anders, E. M. (1998). Antiviral activity of bovine collectins against rotaviruses. *Journal of General Virology*. **79:** 2255-2263.

Reeves, P. (1995). Role of O-antigen variation in immune responses. *Trends in Microbiology.* **3(10):** 381-386.

Restrepo, C. I., Dong, Q., Savov, J., Mariencheck, W. I., Wright, J. R. (1999). Surfactant Protein D Stimulates Phagocytosis of *Pseudomonas aeruginosa* by Alveolar Macrophages. **21:** 576-585.

Ruelle, P., BuchMann, M., Kesselring, U. (1994). Hydrophobic Effect at the Origin of Low Solubility of Inert Solid Substances in Hydrogen-Bonded solvents. *Journal of Pharmaceutical Sciences.* **83(3):** 396-403.

Saevarsdottir, S., Oskarsson, O. O., Aspelund, T., Eiriksdottir, G., Vikingsdottir, T., Gudnason, V., Valdimarsson, H. (2005). Mannan binding lectin as an adjunct to risk assessment for myocardial infarction in individuals with enhanced risk. *Journal of Experimental Medicine*. **201(1):** 117-125.

Sahly, H., Ofek, I., Podschun, R., Brade, H., He, Y., Ullmann, U., Crouch, E. (2002). Surfactant Protein D Binds Selectively to *Klebsiella pneumoniae* Lipopolysaccharides Containing Mannose-Rich O-Antigens. *The Journal of Immunology.* **169:** 3267-3274.

Sano, H., Chiba, H., Iwaki, D., Sohma, H., Voelker, D. R., Kuroki, Y. (2000). Surfactant Proteins A and D bind CD14 by Different Mechanisms. *The Journal of Biological Chemistry.* **275(29):** 22442-22451.

Schelenz, S., Malhotra, R. Sim, R. B., Holmskov, U., Bangcroft, G. J. (1995). Binding of Host Collectins to the Pathogenic Yeast *Cryptococcus neoformans:* Human Surfactant Protein D Acts as an Agglutinin for Acapsular Yeast Cells. *Infection and Immunity*. **63(9):** 3360-3366.

Schlapbach, L. J., Mattman, M., Thiel, S., Boillat, C., Otth, M., Nelle, M., Wagner, B., Jensenius, J. C., Aebi, C. (2010). Differential Role of the Lectin Pathway of

Complement Activation in Susceptibility to Neonatal Sepsis. *Clinical Infectious Diseases.* **51(2):** 153-162.

Schmid, R. (2001). Recent Advances in the Description of the Structure of Water, the Hydrophobic effect, and the Like-Dissolves-Like rule. *Monatshefte fur Chemie*. **132:** 1295-1326.

Schnaitman, C. A. & Klena, J. D. (1993). Genetics of Lipopolysaccharide Biosynthesis in Enteric Bacteria. *Microbiology Reviews.* **57(3):** 655-682.

Schweda, K., Richards, J., Hood, D., Moxon, E. (2007). Expression and structural diversity of the lipopolysaccharide of *Haemophilus influenzae*. Implications in virulence. *International Journal of Medical Microbiology*. **297:** 297-306.

Sek, L., Boyd, B. J., Charman, W. N., Porter, C. J. H. (2006). Examination of the impact of a range of Pluronic surfactants on the in-vitro solubilisation behaviour and oral bioavailability of lipidic formulations of atovaquone. *Journal of Pharmacy and Pharmacology.* **58:** 809-820.

Selman, L. & Hansen, S. (2012). Structure and function of collectin liver 1 (CL-L1) and collectin 11 (CL-11, CL-K1). *Immunobiology.* **217:** 851-863.

Selvarengan, P., Kubicki, J. D., Guegan, J.-P., Chatellier, X. (2010). Complexation of carboxyl groups in bacterial lipopolysaccharides: Interactions of H^{+}, Mg^{2+}, Ca^{2+}, Cd^{2+}, and UO^{2+}_{2} with Kdo and galacturonate molecules via quantum mechanical calculations and NMR spectroscopy. *Chemical Geology.* **273:** 55-75.

Senft, A. P., Korfhagen, T. R., Whitsett, J. A., Shapiro, S. D., LeVine, A. M. (2005). Surfactant Protein-D regulates Soluble CD14 through Matrix Metalloproteinase-2. *The Journal of Immunology.* **174:** 4953-4959.

Shang, F., Rynkiewicz, M. J., McCormack, F. X., Wu, H., Cafarella, T. M., Head, J. F., Seaton, B. A. (2011). Crystallographic Complexes of Surfactant Protein A and Carbohydrates Reveal Ligand-induced Conformational Change. *Journal of Biological Chemistry.* **286(1):** 757-765.

Sharma, P., Dube, D., Sinha, M., Yadav, S., Kaur, P., Sharma, S., Sing, T. P. (2013). Structural Insights into the Dual Strategy of Recognition by Peptidoglycan Recognition Protein, PGRP-S: Structure of the Ternary Complex of PGRP-S with Lipopolysaccharide and Stearic Acid. *Plos One.* **8(1):** 1-8.

Sheriff, S., Chang, C. Y., Ezekowitz, A. B. (1994). Human mannose-binding protein carbohydrate recognition domain trimerizes through a triple α-helical coiled-coil. *Nature- Structural Biology*. **1(11):** 789-794.

Shrive, A. K., Martin, C., Burns, I., Paterson, J. M., Martin, J. D., Townsend, J. P., Waters, P., Clark, H. W., Kishore, U., Reid, K. B. M., Greenhough, T. J. (2009). Structural Characterisation of Ligand-Binding Determinants in Human Lung Surfactant Protein D: Influence of Asp325. *Journal of Molecular Biology.* **394:** 776-788.

Shrive, A., Tharia, H., Strong, P., Kishore, U., Burns, I., Rzikallah, P., Reid, K., Greenhough, T. J. (2003). High-resolution structural insights into ligand binding and immune cell recognition by human lung surfactant protein D. *Journal of Molecular Biology*. **331(2):** 509-23

Silveyra, P. & Floros, J. (2012). Genetic variant associations of human SP-A and SP-D with acute and chronic lung injury. Frontiers in Bioscience. **17:** 407-429.

Singh, M., Madan, T., Waters, P., Parida, S. K., Sarma, P. U., Kishore, U. (2003). Protective effects of a recombinant fragment of human surfactant protein D in a murine model of pulmonary hypersensitivity induced by dust allergens. *Immunology Letters.* **86:** 299-307.

Skarzynski, T. (2013). Collecting data in home laboratory: evolution of X-ray sources, detectors and working practices. *Acta Crystallographica Section D.* **69:** 1283-1288.

Skvarla, J. (2001). Hydrophobic interaction between macroscopic and microscopic surfaces. Unification using surface thermodynamics. *Advances in Colloid and Interface Science.* **91:** 335-390.

Sliwa-Dominiak, J., Tokarz-Deptula, B., Niedzwiedzka-Rystwej, P., Deptula, W. (2010). Conglutinin – an Important Element of Natural Immunity in Ruminants (a Review). *Acta Veterinaria Brno*. **79:** 99-104.

Smyth, M. S. & Martin, J. H. J. (2000). X-Ray Crystallography. *The journal of Clinical Pathology.* **53:** 8-14.

Syrogiannopoulos, G. A., Hansen, E. J., Erwin, A. L., Munford, R. S., Rutledge, J. Reisch, J. S., McCracken, G. H. Jr. (1988). *Haemophilus influenzae* Type b Oligosaccharide Induces Meningeal Inflammation. *The Journal of Infections Diseases.* **157(2):** 237-244.

Steinebrunner, N., Sandig, C., Zimmermann, S., Stremmel, W., Eisenbach, C., Mischnik, A. (2013). *Salmonella enterica serovar* Minnesota urosepsis in a patient with Crohn's disease in the absence of recent or current gastrointestinal symptoms. *Journal of Medical Microbiology.* **62:** 1360-1362.

Stenutz, R., Weintraub, A., Widmalm, G. (2006). The structures of *Escherichia coli* O-polysaccharide antigens. *Federation of European Microbial Societies*. **30:** 382-403.

Strong, P., Townsend, P., Mackay, R., Reid, K. B. M., Clark, H. W. (2003). A recombinant fragment of human SP-D reduces allergic responses in mice sensitized to house dust mice allergens. *Clinical and Experimental Immunology.* **134:** 181-187.

Strunk, T., Currie, A., Richmond, P., Simmer, K., Burgner, D. (2011). Innate immunity in human newborn infants: prematurity means more than immaturity. *The Journal of Maternal-Fetal and Neonatal Medicine.* **24(1)**: 25-31.

Sun, G. Y. & Sun, A. Y. (1985). Ethanol and Membrane Lipids. *Alcoholism: Clinical and Experimental Research.* **9(2):** 164-180.

Tagaram, H. R. S., Wang, G., Umstead, T. M., Mikerov, A. N., Thomas, N. J., Graff, G. R., Hess, J. C., Thomassen, M. J., Kavuru, M. S., Floros, J. (2007). Characterization of a human surfactant protein A1 (SP-A1) gene-specific antibody; SP-A1 content variation among individuals of varying age and pulmonary health. *American Journal of Physiology Lung Cellular and Molecular Physiology*. **292:** 1052-1063.

Taguchi, T., Bronnimann, C., Eikenberry, E. F. (2008). NEXT GENERATION X-RAY DETECTORS FOR IN-HOUSE XRD. *Cambridge University Press.* **23(2):** 101-105.

Takeuchi, O. & Akira, S. (2010). Pattern Recognition Receptors and Inflammation. *Cell.* **140:** 805-820.

Tanaka, S., Oda, Y., Ataka, M., Onuma, K., Fujiwara, S., Yonezawa, Y. (2001). Denaturation and Aggregation of Hen Egg Lysozyme in Aqueous Ethanol Solution Studied by Dynamic Light Scattering. *Biopolymers.* **59:** 370-379.

Tate, M. D., Job, E. R., Deng, Y-M., Gunalan, V., Maurer-Stroh, S., Reading, P. C. (2014). Playing Hide and Seek: How Glycosylation of Influenza Virus Hemagglutinin Can Modulate the Immune Response to Infection. *Viruses.* **6:** 1294-1316.

Taylor, G. L. (2010). Introduction to Phasing. *Acta Crystallographica Section D.* **66:** 325-338.

Tecle, T., Tripathi, S., Hartshorn, K. L. (2010). Defensins and Cathelicidins in lung immunity. *Innate Immunity*. **16(3):** 151-159.

Teng, T-Y. (1990). Mounting of Crystals for Macromolecular Crystallography in a Free-Standing Thin Film. *Journal of Applied Crystallography.* **23:** 387-391.

Terwilliger, T. C., Adams, P. D., Moriarty, N. W., Cohn, J. D. (2007). Ligand identification using electron-density map correlations. *Acta Crystallographica Section D.* **63:** 101-107.

Thiel, S. & Takahashi, Kazue (2013) Collectins. In: eLS. John Wiley & Sons, Ltd: Chichester. DOI: 10.1002/9780470015902.a0003659.pub3.

Thiel, S., Vorup-Jensen, T., Stover, C. M., Schwaeble, W., Laursen, S. B., Poulsen, K., Willis, A. C., Eggleton, P., et al. (1997). A second serine protease associated with mannan- binding lectin that activates the complement. *Letters to Nature.* **386:** 506-510.

Thormahelen, M., Marx, A., Sack, S., Mandelkow, E. (1998). The Coiled-Coil Helix in the Neck of Kinesin. *Journal of Structural Biology.* **122:** 30-41.

Tino, M. J. & Wright, J. R. (1999). Glycoprotein-340 Binds Surfactant Protein-A (SP-A) and stimulates Alveolar Macrophages Migration in an SP-A-Independent Manner. *American Journal of Respiratory Cell and Molecular Biology.* **20:** 759-768

Tomlinson-Phillips, J., Davis, J., Ben-Amotz, D. (2011). Structure and Dynamics of Water Dangling OH Bonds in Hydrophobic Hydration Shells. Comparison of Simulation and Experiment. *The Journal of Physical Chemistry A.* **115:** 6177-6183.

Torres, L. G., Velasquez, A., Brito-Arias, M. A. (2011). Ca-alginate spheres behaviour in presence of some solvents and water-solvent mixtures. *Advances in Bioscience and Biotechnology.* **2:** 8-12.

Tsang, R. S., Schelect, W. S., Aleksic, S., Chan, K. H., Chau, P. Y. (1991). Lack of the α-1,2-linked N-acetyl-D-glucosamine epitope in the outer core structures of

lipopolysaccharides from certain O serogroups and subspecies of *Salmonella enterica. Res. Microbiol.* **142:** 521-523.

Uemura, T., Sano, H., Katoh, T., Nishitani, C., Mitsuzawa, H., Shimizu, T., Kuroki, Y. (2006). Surfactant Protein A without the Interruption of Gly-X-Y Repeats Loses a Kink of Oligomeric Structure and Exhibits Impaired Phospholipid Liposome Aggregation Ability. *Biochemistry.* **45:** 14543-14551.

van de Wetering, J. K., van Eijk, M., van Golde, L. M. G., Hartung, T., van Strijp, J. A. G., Batenburg, J. J. (2001). Characteristics of Surfactant Protein A and D binding to Lipoteichoic Acid and Peptidoglycan, 2 Major Cell Wall Components of Gram-Positive Bacteria. *The Journal of Infectious Diseases.* **184:** 1143-1153.

van de Wetering, J. K., van Golde, L. M. G., Batenburg, J. J. (2004). Collectins: Players of the innate immune system. *FEBS*. **271:** 1229-1249

van de Wetering, J. K., van Remoortere, A., Vaandrager, A. B., Batenburg, J. J., van Golde, L. M. G., Hokke, C. H., van Hellemond, J. J. (2004a). Surfactant Protein D Binding to Terminal α-1-3-Linked Fucose Residues and to *Schistosoma mansoni*. *American Journal of Respiratory Cell and Molecular Biology.* **31:** 565-572.

van Eijk, M., van de Lest, C. H. A., Batenburg, J. J., Vaandrager, A. B., Meschi, J. Hartshorn, K. L., et al. (2002). Porcine Surfactant Protein D Is N-glycosylated in its Carbohydrate Recognition Domain and Is Assembled into Differently Charged Oligomers. *American Journal of Respiratory Cell and Molecular Biology.* **26:** 739-747.

van Rozendaal, B. A. W. M., van Spriel, A. B., van de Winkel, J. G. J., Haagsman, H. P. (2000). Role of Pulmonary Surfactant Protein D in Innate Defense against *Candida albicans.* **182:** 917-922.

Vazquez de Lara, L., Becerril, C., Motano, M., Ramos, C., Maldonado, V., Melendez, J., Phelps, D. S., Pardo, A., Selman, M. (2000). Surfactant components modulate

fibroblast apoptosis and type I collagen and collagenase-1 expression. *American Journal of Physiology Lung Cellular and Molecular Physiology.* **279:** 950-957.

Vekilov, P. G. & Vorontsova, M. A. (2014). Nucleation precursors in protein crystallization. *Acta Crystallographica Section F.* **70:** 271-282.

Vekilov, P. G. (2010). The two-step mechanism of nucleation of crystals in solution. *Nanoscale.* **2:** 2346-2357.

Veldhuizen, E. J. A., van Eijk, M., Haagsman, H. P. (2011). The carbohydrate recognition domain of collectins. *The FEBS journal.* **278:** 3930-3941.

Verez-Bencomo, V., Fernandez-Santana, V., Hardy, E., Toledo, M., Rodriguez, M., Heynngnezz, L., Rodriguez, A., Baly, A., Herrera, L., Izquierdo, M., Villar, A. et al. (2004). A Synthetic Conjugate Polysaccharide Vaccine Against *Haemophilus influenzae* type b. *Science*. **305:** 522-525.

Vinogradov, E., Van der Drift, K., Thomas-Oates, J., Meshkov, S., Brade, H., Holst, O. (1999). The structures of the carbohydrate backbones of the lipopolysaccharides from *Escherichia coli* rough mutants F470 (R1 core type) and F576 (R2 core type). *European Journal Biochemistry*. **261:** 629-639

Voss, T., Eistetter, H., Schafer, K. P. (1988). Macromolecular Organization of Natural and Recombinant Lung Surfactant Protein SP 28-36. Structural homology with the complement factor C1q. *Journal of Molecular Biology.* **201:** 219-227.

Voss, T., Melchers, K., Scheirle, G., Schafer, K. P. (1991). Structural comparison of recombinant pulmonary surfactant protein SP-A derived from two human coding sequences: implications for the chain composition of natural human SP-A. *American Journal of Respiratory Cell and Molecular Biology*. **4(1):** 88-94.

Voynow, J. A. & Kubin, B. K. (2009). Mucins, Mucus and Sputum. *Chest.* **135:** 505-512.

Vukajlovich, S. W. (1986). Antibody-Independent Activation of the Classical Pathway of Human Serum Complement by Lipid A Is restricted to Re-Chemotype Lipopolysaccharide and Purified Lipid A. *Infection and Immunity.* **53(3):** 480-485.

Wall, R., Powell, A., Sohn, E., Foster-Frey, J., Bannerman, D., Paape, M. (2009). Enhanced Host Immune Recognition of Mastitis Causing *Escherchia coli* in CD-14 Transgenic Mice. *Animal Biotechnology.* **20(1):** 1-14.

Wallis, R. & Drickamer, K. (1999). Molecular Determinants of Oligomer Formation and Complement Fixation in Mannose-binding proteins. *The Journal of Biological Chemistry.* **274(6):** 3580-3589.

Walsh, E. J. & Moran, A. P. (1997). Influence of medium composition on the growth and antigen expression of *Helicobacter pylori.* **83:** 67-75.

Wang, G., Myers, C., Mikerov, A., Floros, A. (2007). Effect of Cysteine 85 on Biochemical Properties and Biological Functions of Human Surfactant Protein A variants. *Biochemistry*. **46:** 8425-8435.

Wang, G., Phelps, D. S., Umstead, T. M., Floros, J. (2000). Human SP-A proteins variants derived from one or both genes stimulate TNF-α production in the THP-1 cell line. *American Journal of Physiology Lung Cellular and Molecular Physiology*. **278:** 946-954.

Wang, H., Head, J., Kosma, P., Brade, H., Muller-Loennies, S., Sheikh, S., McDonald, B., Smith, K., Cafarella, T., Seaton, B., Crouch, E. (2008) Recognition of Heptoses and the Inner Core of Bacterial Lipopolysaccharides by Surfactant Protein D. *Biochemistry*. **47**:710-720.

Wang, J. Y., Kishore, U., Lim, B. L., Strong, P., Reid, K. B. M. (1996). Interaction of human lungs and surfactant proteins A and D with mite (*Dermatophagoides pteronyssinus)* allergens. *Clinical and Experimental Immunology.* **106:** 367-373.

Wang, J-Y., Shieh, C-C., You, P-F., Lei, H-Y., Reid, K. B. M. (1998). Inhibitory Effect of Pulmonary Surfactant Protein A and D on Allergen-induced Lymphocyte Proliferation and Histamine Release in Children with Asthma. *American Journal of Respiratory Critical Care Medicine.* **158:** 510-518.

Wang, X., Quinn, P. J., Yan, A. (2015). Kdo_2-lipid A: structural diversity and impact on immunopharmacology. *Biological Reviews.* **90:** 408-427.

Wang, Y., Voelker, D. R., Lugogo, N. L., Wang, G., Floros, J., Ingram, J. L., Chu, H. W., Church, T. D., Kandasamy, P., Fertel, D., Wright, J. R., Kraft, M. (2011). Surfactant protein A is defective in abrogating inflammation in asthma. *American Journal of Physiology*. **301:** 598-606.

Weichenberger, C. X., Afonine, P. V., Kantardjieff, K., Rupp, B. (2015). The solvent component of macromolecular crystals. *Acta Crystallographica Section D.* **71:** 1023-1038.

Weis, W. I. & Drickamer, K. (1994). Trimeric structure of C-type mannose binding protein. *Structure.* **2(12):** 1227-1240.

Weis, W. I., Drickamer, K., Hendrickson, W. A. (1992). Structure of a C-type mannose-binding protein complexed with oligosaccharide. *Nature.* **360:** 127-134.

Weis, W. I., Kahn, R., Fourme, R., Drickamer, K., Hendrickson, W. A. (1991). Structure of the Calcium-Dependent Lectin Domain from Rat Mannose-Binding Protein Determined by MAD PHASING. *American Association for the Advancement of Science.* **254(5038):** 1608-1615.

Weis, W. I., Taylor, M. E., Drickamer, K. (1998). The C-type lectin superfamily in the immune system. *Immunological Reviews.* **163:** 19-34.

Weiser, J. N. & Pan, N. (1998). Adaption of *Haemophilus influenza* to acquired and innate humoral immunity based on phase variation of lipopolysaccharide. *Molecular Microbiology.* **30(4):** 767-775.

Weiser, M. M. & Rothefield, L. (1967). The Reassociation of Lipopolysaccharide, Phospholipid, and Transferase Enzymes of the Bacterial Cell envelop. *The Journal of Biological Chemistry.* **243(6):** 1320-1328.

Wernicki, M. D. A. & Urban-Chmiel, A. P. R. (2011). Conglutinin is not specific to cattle. *Veterinarni Medicina*. **56(10):** 510-519.

Wert, S. E., Yoshida, M., LeVine, A. M., Ikegami, M., Jones, T., Ross, G. F., Fisher, J. H., Korfhagen, T. R., Whitsett, J. A. (2000). Increased metalloproteinase activity, oxidant production, and emphysema in surfactant protein D gene-inactivated mice. *PNAS.* **97(11):** 5972-5977.

Whitsett, J. A. & Alenghat, T. (2015). Respiratory epithelial cells orchestrate pulmonary innate immunity. *Nature Immunology*. **16(1):** 27-35.

Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B. et al., (2011). Overview of the CCP4 and current developments. *Acta Crystallographica Section D.* **67:** 235-242.

Wlodawer, A., Minor, W., Dauter, Z., Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. *The FEBS Journal.* **275:** 1-21.

Wright, J. R. (2005). Immunoregulatory Functions of Surfactant Proteins. *Nature Reviews.* **5:** 58-68

Wu, E. L., Engstrom, O., Jo, S., Stuhlsatz, D., Yeom, M. S., Klauda, J. B., Wildmalm, G., Im, W. (2013). Molecular Dynamics and NMR Spectroscopy Studies of *E. coli* Lipopolysaccharides Structure and Dynamics. *Biophysical Journal.* **105:** 1444-1455.

Wu, H., Kuzmenko, A., Wan, S., Schaffer, L., Weiss, A., Fisher, J. H., Kim, K. S., McCormack, F. X. (2003). Surfactant proteins A and D inhibit the growth of Gram-negative bacteria by increasing membrane permeability. *Journal of Clinical Investigation.* **111(10):** 1589- 1602.

Xu, H. & Dill, K. A. (2005). Water's Hydrogen Bonds in the Hydrophobic Effect: A simple model. *The Journal of Physical Chemistry.* **109(49):** 23611-23617.

Xuong, N-H, Kraut, J., Seely, O., Freer, S. T., Wright, C. S. (1968). Rapid measurement of large numbers of reflection intensities for proteins. *Acta Crystallographica Section B.* **24:** 289-290.

Yamazoe, M., Nishitami, C., Takahashi, M., Katoh, T., Ariki, S., Shimizu, T., Mitsuzawa, H., Sawada, K. et al., (2008). Pulmonary Surfactant Protein D Inhibits Lipopolysaccharide (LPS)-induced Inflammatory Cell Responses by Altering LPS Binding to its receptors. *The Journal of Biological Chemistry.* **283:** 35878-35888.

Yethon, J. A., Heinrichs, D. E., Monteiro, M. A., Perry, M. B., Whitfield, C. (1998). Involvement of *way, waaQ* and *waaP* in the Modification of *Escherichia coli* Lipopolysaccharide and Their Role in the Formation of a Stable Outer Membrane. *The Journal of Biological Chemistry.* **273(41):** 26310-26316.

Yong, S-J., Vuk-Pavlovic, Z., Standing, J. E., Courch, E. C., Limper, A. H. (2003). Surfactant Protein D-Mediated Aggregation of *Pneumocystis carinii* Impairs Phagocytosis by Alveolar Macrophages. *Infection and Immunity.* **71(4):** 1662-1671.

Yoshida, T., Tsuruta, Y., Iwasaki, M., Yamane, S., Ochi, T., Suzuki, R. (2003). SRCL/CL-P1 Recognises GalNAc and Carcinoma-Associated Antigen, Tn Antigen. *Journal of Biochemistry*. **133:** 271-277.

Zelensky, A. N. & Gready, J. E. (2003). Comparative Analysis of Structural Properties of the C-Type-Lectin-like Domain (CTLD). *Proteins: Structure, Function and Genetics*. **52:** 466-477.

Zelensky, A. N. & Gready, J. E. (2005). The C-type lectin-like domain superfamily. The FEBS Journal. **272:** 6179-6217

Zhang, L., Ikegami, M., Crouch, E. C., Korfhagen, T. R. Whitsett, J. A. (2001). Activity of Pulmonary Surfactant Protein D (SP-D) *in Vivo* Is dependent on Oligomeric Structure. *The Journal of Biological Chemistry*. **276(22):** 19214-19219.

Zhang, P., McAlinde, A., Li, S., Schumacher, T., Wang, H., Hu, S., Sandell, L., Crouch, E. (2001a). The Amino-terminal Heptad repeats of the Coiled-coil Neck domain of Pulmonary Surfactant Protein D Are Necessary for assembly of Trimeric Subunits and Dodecamers. *The Journal of Biological Chemistry.* **276(23):** 19862-19870.

Zhang, X. & Gobas, F. A. P. C. (1995). A Thermodynamic Analysis of the Relationships Between Molecular Size, Hydrophobicity, Aqueous Solubility and Octanol-water Partitioning of Organic Chemicals. *Chemosphere.* **31(6):** 3501-3521.